IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 2 June 2023, accepted 16 July 2023, date of publication 26 July 2023, date of current version 3 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3298955

RESEARCH ARTICLE

Innovative Feature Selection Method Based on Hybrid Sine Cosine and Dipper Throated Optimization Algorithms

ABDELAZIZ A. ABDELHAMID[1,2], EL-SAYED M. EL-KENAWY[3], (Senior Member, IEEE), ABDELHAMEED IBRAHIM[4], (Member, IEEE), MARWA METWALLY EID[5], (Member, IEEE), DOAA SAMI KHAFAGA[6], AMEL ALI ALHUSSAN[6], SEYEDALI MIRJALILI[7,8], (Senior Member, IEEE), NIMA KHODADADI[9], (Member, IEEE), WEI HONG LIM[10], (Senior Member, IEEE), AND MAHMOUD Y. SHAMS[11]

[1]Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt
[2]Department of Computer Science, College of Computing and Information Technology, Shaqra University, Shaqra 11961, Saudi Arabia
[3]Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt
[4]Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt
[5]Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 35712, Egypt
[6]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[7]Centre for Artificial Intelligence Research and Optimization, Torrens University Australia, Fortitude Valley, QLD 4006, Australia
[8]Yonsei Frontier Laboratory, Yonsei University, Seoul 03722, South Korea
[9]Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL 33146, USA
[10]Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur 56000, Malaysia
[11]Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh 33516, Egypt

Corresponding authors: Doaa Sami Khafaga (dskhafga@pnu.edu.sa) and El-Sayed M. El-kenawy (skenawy@ieee.org)

This work was supported by Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, through Researchers Supporting under Project PNURSP2023R 308.

ABSTRACT *Introduction*: In pattern recognition and data mining, feature selection is one of the most crucial tasks. To increase the efficacy of classification algorithms, it is necessary to identify the most relevant subset of features in a given domain. This means that the feature selection challenge can be seen as an optimization problem, and thus meta-heuristic techniques can be utilized to find a solution. *Methodology*: In this work, we propose a novel hybrid binary meta-heuristic algorithm to solve the feature selection problem by combining two algorithms: Dipper Throated Optimization (DTO) and Sine Cosine (SC) algorithm. The new algorithm is referred to as bSCWDTO. We employed the sine cosine algorithm to improve the exploration process and ensure the optimization algorithm converges quickly and accurately. Thirty datasets from the University of California Irvine (UCI) machine learning repository are used to evaluate the robustness and stability of the proposed bSCWDTO algorithm. In addition, the K-Nearest Neighbor (KNN) classifier is used to measure the selected features' effectiveness in classification problems. *Results*: The achieved results demonstrate the algorithm's superiority over ten state-of-the-art optimization methods, including the original DTO and SC, Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), Grey Wolf Optimization (GWO), Multiverse Optimization (MVO), Satin Bowerbird Optimizer (SBO), Genetic Algorithm (GA), the hybrid of GWO and GA, and Firefly Algorithm (FA). Moreover, Wilcoxon's rank-sum test was performed at the 0.05 significance level to study the statistical difference between the proposed method and the alternative feature selection methods. *Conclusion*: These results emphasized the proposed feature selection method's significance, superiority, and statistical difference.

INDEX TERMS Feature selection, dipper throated optimization algorithm, Sine cosine optimization algorithm, meta-heuristic optimization.

The associate editor coordinating the review of this manuscript and approving it for publication was Claudia Raibulet.

I. INTRODUCTION

Feature Selection is the process of selecting relevant features for the machine learning model based on the type of problem being solved. This process is also defined as isolating the most consistent and non-redundant features for use in machine learning tasks [1]. Therefore, the primary goal of feature selection is to improve predictive model performance while lowering computational modeling costs [2]. The advantages of feature selection can be briefly described as follows (1) It helps simplify the machine learning models by reducing the input features. (2) The training time of machine learning models based on feature selection can be significantly reduced as the more precise collection of features reduces the time required to train a model. (3) With the help of feature selection, the precision of machine learning models can be increased. (4) The curse of high dimensionality can be avoided by reducing the input features [3]. The dimensionality of a problem is a significant problem that might degrade classification efficiency. Large datasets with a high number of features are essential for the success of many applications. As a result, classification performance is hampered since many of these features are redundant, unnecessary, or noisy [4]. Thus, to properly prepare data for machine learning algorithms, feature selection is a necessary step [5], [6], [7]. Figure 1 depicts the typical feature selection process. The dataset is usually preprocessed to handle missing values and outliers. Then, the selection of the best features is applied through an iterative process. A subset of features (solution subsets) is selected and evaluated using criteria to decide whether to keep or remove. This iterative process's final output is the best feature set that can be assessed using machine learning classifiers [8], [9].

On the other hand, multiple fields of study use optimization techniques, including computing, agriculture, medicine, engineering, and feature selection. Optimization aims to identify and pick the best solution to a problem from among those that satisfy the problem statement requirements. Further, in optimization methods, the goal is typically to reduce or maximize criteria, depending on the nature of the situation at hand [10]. The main feature selection approaches include wrapper, filter, and hybrid-based methods [11]. Wrapper techniques are accurate but time-consuming since they need to incorporate learning methods into the selection function, reducing the search space for choosing features. The speed and scalability of the filter-based feature selection techniques or the conventional feature selection approaches are an advantage. In computer science, genetic algorithms are based on the randomness of the natural selection process, which is the basis for all biological evolution and can be used in numerous areas, such as machine learning problems, optimization, and feature selection [12]. Evolutionary computing methods are being examined as an option to get the best solution. Nature, biological behavior, and creatures' social behavior, such as birds, whales, bats, grasshoppers, fireflies, salp, fish, wolves, etc., inspire swarm-based algorithms [13]. Researchers in several fields have turned to optimization techniques to find solutions to various problems.

As the feature selection method significantly impacts the performance of machine learning classification models, it is crucial to provide an effective feature selection method to realize this target. This represents the primary motivation of this work. Therefore, in this paper, we proposed a novel hybrid algorithm that effectively selects the most significant set of features to improve classification performance. The proposed algorithm is based on the original DTO and SC algorithm. The effectiveness of the proposed approach was validated in terms of 30 datasets from the UCI machine learning repository, including seven datasets with more than 1000 features. In addition, statistical analysis is performed to study the superiority and stability of the proposed hybrid approach. The achieved results confirmed the findings of this approach. Moreover, the proposed method is compared with other popular metaheuristic algorithms, including the original bDTO and bSC, bPSO, bWOA, bGWO, bMVO, bSBO, bGWO-GA, bFA, and bGA for this research. To sum up the main findings of this work:

- A novel feature selection algorithm is proposed based on hybrid SC and DTO algorithms.
- Evaluation of the proposed bSCWDTO in terms of thirty UCI benchmark datasets.
- A comparison between the proposed approach and other state-of-the-art feature selection methods.
- A statistical analysis of the proposed approach is performed to prove its significance and statistical significance.
- Evaluating the continuous version of the proposed algorithm in terms of the CEC2017 benchmark functions.
- Performing sensitive analysis of the proposed algorithm's convergence time and convergence fitness.

The structure of this paper is presented in six sections. The literature review of the studies on feature selection is provided in Section II. The preliminaries that form the basis of the proposed method are introduced in Section III. The proposed methodology is then explained in Section IV. The achieved results are presented and discussed in Section V. Finally, the conclusion and future perspectives are presented in Section VI.

II. LITERATURE REVIEW

Eberhart and Kennedy first presented an evolutionary algorithm based on swarm intelligence in 1995 as Particle Swarm Optimization (PSO) algorithm [14]. The PSO approach, developed to tackle the feature selection problem, has been widely used in multiple research since its inception. The approach was motivated by the cooperative nature of birds and fish. The PSO technique has several benefits, such as its ease of use and quick convergence rate. However, this strategy has a few problems, including local optimums and a lack of population variety. As a result, some pieces have discussed combining PSO with other algorithms to boost its speed and apply it to feature selection problems. A PSO-based hybrid feature selection technique using a local search strategy is proposed, for instance, by the authors of [15]. The

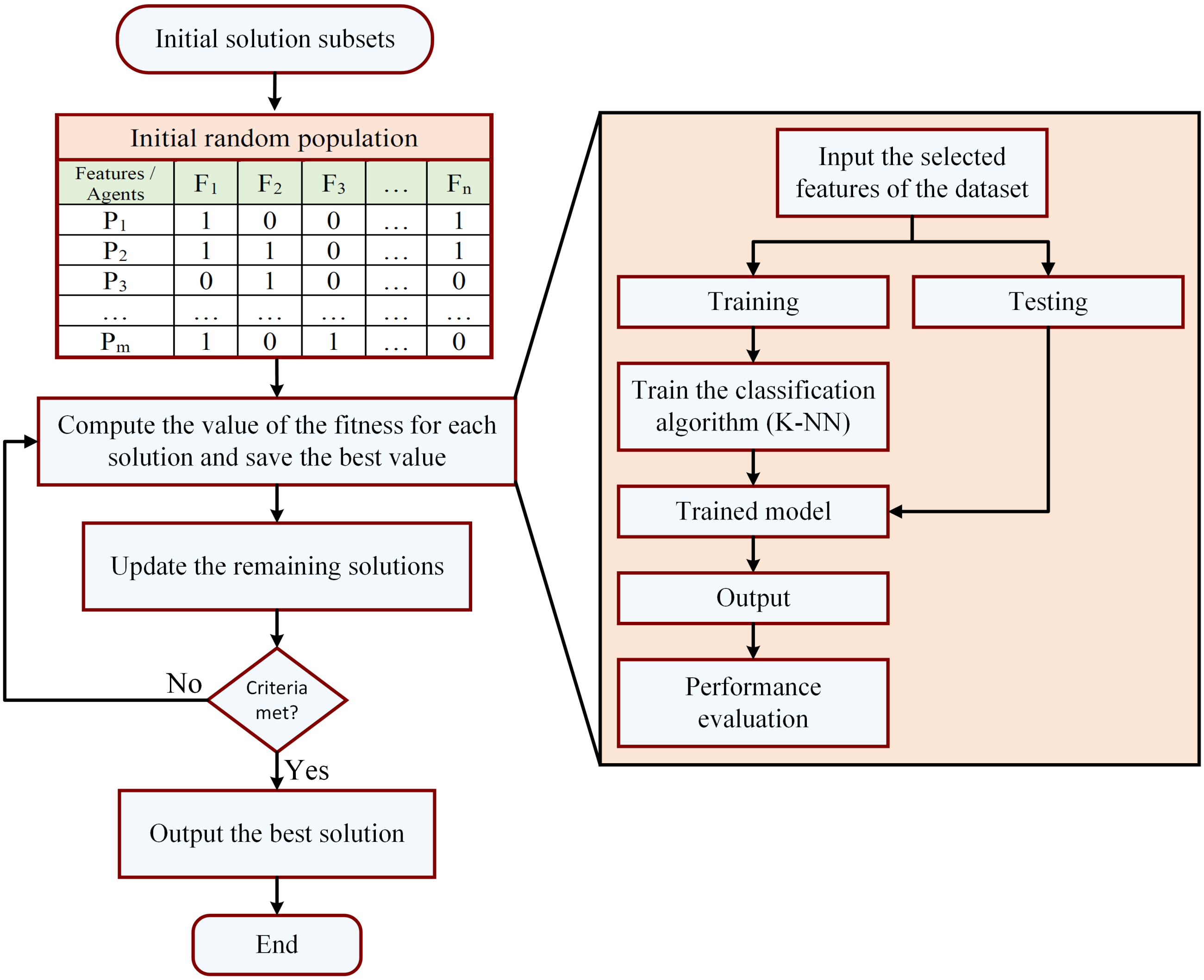


FIGURE 1. The typical process of feature selection.

proposed method uses a local search approach embedded in particle swarm optimization to pick the least correlated and most crucial subset of features. The local search strategy aims to help the particle swarm optimization search process pick relevant features by leveraging data on existing correlations. It was tested on 13 distinct standard classification datasets and compared to 5 popular feature selection techniques. Other writers have employed PSO in their work as well [16], [17]. For complex facial expression recognition problems, the authors presented a genetic algorithm (GA) integrated with PSO for feature selection. Using Gaussian mutation in the equation for updating the particle's velocity, a micro-GA was incorporated into the original PSO technique to delay the onset of convergence. Successful global and local search also requires a system for updating velocity, which depends on the average user's experience [18], [19], [20].

The authors of developed wrapper-based techniques for discovering the most important and optimal features citezR49 using a PSO and a spiral-shaped mechanism (PSO-SSM). The proposed technique is more effective than traditional Gas and PSO-based feature selection algorithms in determining the emotional state of a face. The PSO-SSM made three enhancements. To begin, finding a wide selection of products was aided by logistical maps. Secondly, the position quality of the subsequent generation was significantly enhanced by adding two additional parameters to the original position update method. After locating the ideal solution region, a spiral-shaped mechanism was used as a local search operator. Twenty well-known benchmark datasets were used to analyze the performance of the proposed PSO-SSM using a kNN classifier and to draw comparisons to the wrapper and filter-based techniques [21], [22], [23].

The Grasshopper Optimization System (GOA) is an innovative swarm intelligence algorithm that takes cues from real-life grasshoppers' foraging and hive-mind tactics. Researchers in [24] found that a new binary hybrid algorithm may be developed by combining the GOA algorithm and the mutation operator of GA. The authors could convert continuous GOA to binary by applying transfer functions. Moreover, a mutation operator with a moderate mutation rate was used to provide a wide range of possible answers. Select features were examined using k-NN for 25 reference datasets. The about 92% accuracy in categorization was a substantial improvement above standard benchmarking approaches. The authors of [25] improved the version of GOA with additional evolutionary-based operators for creating a productive wrapper feature selection method. On twenty-two UCI datasets, the proposed approaches were evaluated. The effectiveness of the GOA was shown to be significantly affected by the EPD. By including the selection mechanism, the proposed method became more effective than competing optimizers at locating optimal solutions and displaying superior convergence trends.

The Salp Swarm Algorithm (SSA) is a cutting-edge meta-heuristic algorithm that mimics the actions of salps in the ocean's depths and is invented by the authors of [26] and [27]. In some feature selection approaches, SSA has been used as a search technique [28], [29]. Similar improvements in opportunistic search behavior were also noticed by the authors of [30], who addressed problems with the SSA method. Researchers improved SSA's exploitability using the local search (LS) method. The research also employed a chaotic map and an original equation variable to determine the most effective way of providing followers with their current position. For the feature-selection problem, 20 standard-setting classification datasets and three datasets were used to test the effectiveness of the proposed approach. A dynamic SSA was shown to be superior to other possible alternatives.

To solve global optimization problems, the authors of [31] presented an algorithm called the Sine Cosine Algorithm (SC) that takes advantage of the features of sine and cosine functions. In [32], the authors develop a system combining SSA and SC while introducing a new population diversification mechanism called the "disruption operator." When combining SSA with SC to generate a pool of candidates for a solution, more variety was included to prevent a decline in solution quality. When applied to feature selection problems for datasets with feature sizes between thirteen and eleven thousand, the outcomes were encouraging. The authors of [33] developed a novel hybridization strategy using SC. They transformed traditional PSO into binary variations using massive datasets and added SC to enhance exploration. The clustering problem for seven high- and low-dimensional datasets, including nine to over eleven thousand features, was solved using the k-means approach after preliminary testing with ten standard benchmark test functions. The proposed technique incorporated the SC's location update equation into the PSO's velocity equation [34]. In addition, the PSO's weighting factor was modified, with the value shifting with each iteration. A select group of iterations was selected to inject the maximum inertia weight to improve the capability of searching distant locations. The study claimed a significant improvement in clustering accuracy compared to several previous natural-inspired optimization approaches using statistical t-tests. It was proposed by the authors of [35] that SC and Antlion Optimization (ALO) may be combined to form a novel hybrid optimization method. The first group was updated using SC, and the second was updated using ALO. The position update equations also included numerous random variables to boost the population's diversity. Using V- and S-shaped transport functions in feature selection, the authors implemented a binary variant of the proposed method [35], [36].

Authors in [37] present a Grey Wolf Optimizer algorithm coupled with a Two-phase Mutation to address feature selection for classification problems based on wrapper approaches. To fit the binary form of the feature selection problem, the sigmoid function is utilized to convert the continuous search space to the binary one. The two-phase mutation improves the algorithm's exploitation capacity. The initial mutation phase aims to minimize the number of chosen features while maintaining good classification accuracy. The second mutation phase seeks to incorporate more informative features that improve classification accuracy. Because the mutation phase can be time-demanding, the two-phase mutation is less likely to succeed [38], [39]. Because wrapper approaches can produce high-quality results, authors employed one of the most well-known wrapper methods, the k-Nearest Neighbor (k-NN) classifier. To find the k-NN, they use Euclidean distance is calculated. Each dataset is divided into training and testing data using K-fold cross-validation to avoid overfitting. Several comparisons were made between the flower method, particle swarm optimization algorithm, multi-verse optimizer algorithm, whale optimization algorithm, and bat algorithm. Thirty-five datasets are used in the studies. Statistical analyses are performed to demonstrate the efficacy and outperformance of the proposed method.

Grey Wolf Optimization (GWO) is a meta-heuristic method based on a mathematical grey wolf leadership and hunting model. Grey wolves often dwell in groups of 5-12 individuals and have a rigid social dominating structure. They are classified into four categories based on their dominance: alpha (α), beta (β), delta (δ), and omega (ω). This study presents a feature selection technique for picture steganalysis based on a novel levy flight-based grey wolf optimizer (LFGWO). The method is verified using the BOSS-base ver. 1.01 image dataset, which contains cover and stego images. The feature extraction techniques, such as AlexNet, were used to extract 686 and 1000 features, respectively. The proposed LFGWO-based feature selection strategy is compared to PSO and GWO-based feature selection approaches [40]. Regarding mean fitness, standard deviation values, and convergence behavior, LFGWO surpasses the meta-heuristic algorithms GWO, PSO, and GSA. The proposed LFGWO outperforms previous meta-heuristic algorithms according to practical and statistical results.

The authors of [41] present a hybrid approach to solving function optimization and feature selection issues by combining the Grey Wolf Optimizer (GWO) and the Crowd Search Algorithm. The suggested hybrid algorithm can successfully explore the search space since it combines GWO with other approaches to overcome shortcomings. The suggested approach accelerates the optimization process's early stages by making full use of both algorithms' capabilities through an adaptive balancing probability. While promising ideas may be used early on in the optimization process, it is more probable that they will be utilized later on. In practice, optimization problems having more than three objectives are not unusual. In evolutionary computing, many-objective optimization issues pose significant obstacles. However, algorithm performance analysis and comparison have received comparatively less focus than the rapid development of algorithm design. Many-objective optimization uses several test problem sets initially created for multi-objective optimization [42], [43], [44]. This contest presents a set of test issues that accurately depict various real-world conditions by selecting and designing 15 test problems with distinct qualities to further evolutionary many-objective optimization research.

III. PRELIMINARIES

In this section, the feature selection problem formulation is presented. In addition, the basics of the algorithms employed in developing the proposed hybrid algorithm are demonstrated.

A. FEATURE SELECTION PROBLEM FORMULATION

In this part, we describe the mathematical modeling of feature selection. The typical dimensions of a dataset for classification (i.e., supervised learning) are $N_S \times N_F$, where N_S is the total number of samples, and N_F is the number of features. To accomplish its task, the feature selection algorithm first divides the entire set of features N_F into smaller subsets (S) whose combined dimensions are smaller than N_F. To get to that subset of features, you may use the following objective function:

$$Fit = \lambda \times \gamma_S + (1-\lambda) \times \left(\frac{|S|}{N_F}\right) \tag{1}$$

where λ is selected from the range [0-1], and it is used to balance between $\left(\frac{|S|}{N_F}\right)$ and γ_S. The selected features are denoted by $|S|$, and γ_S is the classification error [45].

B. DIPPER THROATED OPTIMIZATION ALGORITHM

There are three methods used in the explorer stage. Birds' cooperative nature inspired a novel metaheuristic method called DTO. Here, we offer a narrow mathematical mechanism and give a detailed account of its discovery and use. The DTO method uses three different techniques to improve exploration: (1) flying to a new site, (2) switching to another bird, and (3) flying efficiently over a known region. The exploitation process involves watching the birds and trying to out-hunt one another for food [46].

When applying the DTO algorithm, a flock of birds swims through the space to search for food. The positions and speeds of the birds can be represented by the following two matrices, referred to as P and V, respectively. Using these metrics, DTO may probe the search space for the optimal answer. The following matrices explain the DTO algorithm's computations in further detail.

$$P = \begin{bmatrix} P_{1,1} & P_{1,2} & P_{1,3} & \dots & P_{1,d} \\ P_{2,1} & P_{2,2} & P_{2,3} & \dots & P_{2,d} \\ P_{3,1} & P_{3,2} & P_{3,3} & \dots & P_{3,d} \\ \dots & \dots & \dots & \dots & \dots \\ P_{m,1} & P_{m,2} & P_{m,3} & \dots & P_{m,d} \end{bmatrix} \tag{2}$$

$$V = \begin{bmatrix} V_{1,1} & V_{1,2} & V_{1,3} & \dots & V_{1,d} \\ V_{2,1} & V_{2,2} & V_{2,3} & \dots & V_{2,d} \\ V_{3,1} & V_{3,2} & V_{3,3} & \dots & V_{3,d} \\ \dots & \dots & \dots & \dots & \dots \\ V_{m,1} & V_{m,2} & V_{m,3} & \dots & V_{m,d} \end{bmatrix} \tag{3}$$

For the indexes $i \in 1, 2, 3, \dots, m$ and $j \in 1, 2, 3, \dots, d$ and in the j^{th} dimension, the bird i^{th} is referred to as by $P_{i,j}$, and the speed of the bird is denoted by $V_{i,j}$. The following array determines the bird's fitness $f = f_1, f_2, f_3, \dots, f_n$.

$$f = \begin{bmatrix} f_1(P_{1,1}, P_{1,2}, P_{1,3}, \dots, P_{1,d}) \\ f_2(P_{2,1}, P_{2,2}, P_{2,3}, \dots, P_{2,d}) \\ f_3(P_{3,1}, P_{3,2}, P_{3,3}, \dots, P_{3,d}) \\ \dots \\ f_m(P_{m,1}, P_{m,2}, P_{m,3}, \dots, P_{m,d}) \end{bmatrix} \tag{4}$$

Mother birds have the highest fitness among birds because they provide the most offspring with the skills to find food and survive. The best position denoted P_{best}, is updated during the search process. The P_{nd}, which refers to the regular birds, serves as followers to the mother birds. P_{Gbest} refers to the most optimal solution available during the search process. The optimizer uses the DTO strategy to follow the swimming bird using the following equations to account for movement within the population and time.

$$X = P_{best}(i) - K_1.|K_2.P_{best}(i) - P(i)| \tag{5}$$

$$Y = P(i) + V(i+1) \tag{6}$$

$$P(i+1) = \begin{cases} X & \text{if } r_3 < 0.5 \\ Y & \text{otherwise}, \end{cases} \tag{7}$$

$$V(i+1) = K_3 V(i) + K_4 r_1 (P_{best}(i) - P(i)) + K_5 r_2 (P_{Gbest} - P(i)) \tag{8}$$

where the best birds' position is denoted $P_{best}(i)$, the average position of the birds for iteration i is referred to as $P(i)$, and $V(i+1)$ is the speed of the birds on iteration $i+1$. To clarify, K_4 and K_5 are constants with values 1.7 and 1.8, respectively, and K_1, K_2, and K_3 are weight values selected dynamically from the range of [0-2] during the optimization process. Random numbers in the range [0, 1] make up the values of r_1, r_2, and r_3.

C. SINE COSINE OPTIMIZATION ALGORITHM

The Sine Cosine (SC) algorithm was initially introduced in [47]. This algorithm's sines (and cosines) oscillate functions are essential in determining the optimal solution positions. To express SC operations, the following random variables are employed [48].

- The location of the movements.
- The direction of the motion.
- The swapping among the components of sines and cosines.
- Emphasizing/de-emphasizing the destination effect.

The update process of the candidate solutions is performed using the following equation.

$$P(t+1) = \begin{cases} P(t) + r_5.sin(r_6).|r_7 S^*(t) - S(t)| & r_4 < 0.5 \\ P(t) + r_5.cos(r_6).|r_7 S^*(t) - S(t)| & r_4 \geq 0.5 \end{cases} \tag{9}$$

where the number of search iterations is represented by t. The algorithm tracks two important solutions: the current solution, denoted as S, and the best solution, denoted as S^*. Random variables r_4, r_6, and r_7 are allocated values in the range of [0, 1]. These random variables play a crucial role in the algorithm as they influence the positions of the solutions. Specifically, the equation utilized in the algorithm indicates that the location of the best solution obtained thus far impacts the current solution's position. This influence facilitates the exploration of the search space and increases the likelihood of converging to an optimal solution. During the running iterations of the SC algorithm, the value of r_4 is dynamically updated according to the following equation, further enhancing the search process.

$$r_4 = a - \frac{a \times t}{t_{max}} \tag{10}$$

where a is a constant, t and t_{max} represent the current and maximum iterations, respectively.

The SC algorithm is a resilient metaheuristic approach compared to many existing algorithms. Its ability to utilize a single optimal solution to guide the other solutions sets it apart. This approach contributes to a notable reduction in convergence time and memory usage, distinguishing it from alternative algorithms [48]. However, it is important to acknowledge that the efficiency of the SC algorithm can be compromised when confronted with an increasing number of local optima. Stagnant local optima pose a challenge, which can hinder the algorithm's progress. We have integrated the SC optimizer and the Dynamic Throated Optimization (DTO) algorithm into our novel approach to address this issue. By incorporating the fast convergence rates and memory efficiency of the SC optimizer and DTO algorithm, we aim to strike a healthy balance between exploration and exploitation tasks throughout the optimization process, ensuring enhanced performance and overcoming the limitations associated with growing local optima.

D. K-NEAREST NEIGHBOR

In this study, the k-Nearest Neighbor (KNN) classifier, a supervised learning method, is employed as the basis for a wrapper approach to feature selection [30]. KNN does not rely on the construction of models but instead uses training examples to determine the class of the unknown instance. The KNN is employed in the conducted experiments to evaluate the efficacy of the traits. Each sample is assigned to a category based on the majority vote of its nearest K neighbors. Finding the K nearest neighbors to a sample is done by computing the Euclidean distance, Euc_D, between features from the training data and features from the testing data, which is calculated using the following equation.

$$Euc_D = \sqrt{\sum_{i=1}^{k} (|Train\_F_i| - |Test\_F_i|)^2} \tag{11}$$

where $Train\_F_i$ and $Test\_F_i$ are the sets of features in the training and testing sets, respectively, and k refers to the number of features.

IV. THE PROPOSED METHODOLOGY

The proposed feature selection algorithm is based on two optimization algorithms, namely,s DTO and SC algorithms, and is denoted by binary sine-cosine weighted dipper-throated optimization (bSCWDTO). The proposed algorithm exploits the advantages of both algorithms to improve the exploration of the search space and better exploitation of the intermediate solutions to find the best set of features. The steps of the proposed algorithm are presented in Algorithm 1.

A. BINARY OPTIMIZATION

By selecting the best set of features for improving the classification accuracy, the continuous output of bSCWDTO is converted into binary (0 or 1) using the sigmoid function represented by the following equation.

$$P_b^{(t+1)} = \begin{cases} 1 & \text{if } Sigmoid(P_{Best}) \geq 0.5 \\ 0 & otherwise, \end{cases}$$
$$Sigmoid(P_{Best}) = \frac{1}{1 + e^{-10(P_{Best} - 0.5)}} \tag{12}$$

In the algorithm context, the symbol P_{Best} represents the best position achieved thus far in the optimization process. The iteration number is denoted by t, indicating the current stage of the algorithm. A fitness function is employed to assess the quality of candidate solutions during the feature selection process. This function serves as a measure of the suitability or effectiveness of a particular set of selected features. The following equation mathematically represents the fitness function:

$$F_n = w_1 Error(P) + w_2 \frac{Number\ of\ selected\ features}{Total\ number\ of\ features} \tag{13}$$

where P is a solution, $w_1 \in [0, 1]$, and $w_2 = 1 - w_1$, which are used to control the importance of the number of the selected feature for the population of size n and to maintain the classification error rate.

Algorithm 1 The Proposed Binary bSCWDTO Algorithm

Initialize birds' positions $P_i (i = 1, 2, \ldots, n)$ for n birds, birds' velocity $V_i (i = 1, 2, \ldots, n)$, objective function f_n, iterations t, T_{max}, parameters of $r_1, r_2, r_3, r_4, r_5, r_6, r_7, K_1, K_2, K_3, K_4, K_5$
Calculate fitness of f_n for each bird P_i
Find best bird position P_{best}
Convert best solution to binary [0, 1]
Set $t = 1$
while $t \leq Tmax$ **do**
  for $(i = 1 : i < n + 1)$ **do**
    if $(t\%2 == 0)$ **then**
      if $(r_3 < 0.5)$ **then**
        Update the current swimming bird's position as:
        $P(i+1) = P_{best}(i) - K_1.|K_2.P_{best}(i) - P(i)|$
      else
        Update the current flying bird's velocity as:
        $V(i+1) = K_3V(i) + K_4r_1(P_{best}(i) - P(i)) + K_5r_2(P_{Gbest} - P(i))$
        Update the current flying bird's position as:
        $P(i+1) = P(i) + V(i+1)$
      end if
    else
      Update current agents' positions as
      if $(r7 < 0.5)$ **then**
        Update agent position by:
        $P(t+1) = P(t) + r_4 sin(r_5)|r_6 P_{Gbest}(t)P(t)|$
      else
        Update agent position by:
        $P(t+1) = P(t) + r_4 cos(r_5)|r_6 P_{Gbest}(t)P(t)|$
      end if
    end if
  end for
  Update $r_1, r_2, r_3, r_4, r_5, r_6, r_7, K_1, K_2, K_3$
  Convert to binary the updated solution by Equation (12).
  Calculate objective function f_n for each bird P_i
  Find the best position P_{best}
end while
Return the best solution P_{Gbest}
Set $t = t + 1$

B. COMPLEXITY ANALYSIS

The complexity of the proposed bSCWDTO is calculated in the following steps, where T_{max} refers to the maximum iterations and n number of agents.

- Initialize parameters of the bSCWDTO algorithm: $O(1)$.
- Calculate F_n for each bird P_i: $O(n)$.
- Update positions of swimming birds: $O(T_{max} \times n)$.
- Update positions of flying birds: $O(T_{max} \times n)$.
- Update velocities of flying birds: $O(T_{max} \times n)$.
- Update positions Investigating area around best solution: $O(T_{max} \times n)$.
- Calculate updated best solution: $O(T_{max} \times n)$.
- Calculate F_n for each agent S_i: $O(T_{max})$.
- Update bSCWDTO parameters: $O(T_{max})$.
- Convert solution to binary: $O(T_{max})$.
- Obtain best bird P_{best}: $O(T_{max})$.
- Obtain the global best bird P_{Gbest}: $O(1)$

Based on the above steps of complexity analysis, the complexity of the proposed bSCWDTO is measured as $O(T_{max} \times n)$. So our algorithm has $O(n)$ time complexity.

TABLE 1. Configuration parameters of the competing algorithms.

Algorithm	Parameter	Value
DTO [46]	Iterations	500
	Number of runs	30
	Exploration percentage	70
PSO [49]	Acceleration constants	[2, 12]
	Inertia W_{max}, W_{min}	[0.6, 0.9]
	Number of particles	10
	Number of iterations	80
WOA [50]	r	[0, 1]
	Number of iterations	80
	Number of whales	10
	a	2 to 0
GWO [51]	a	2 to 0
	Number of iterations	80
	Number of wolves	10
SBO [52]	Step size	0.94
	Mutation probability	0.05
	Lowe and upper limit difference	0.02
GA [53]	Cross over	0.9
	Mutation ratio	0.1
	Selection mechanism	Roulette wheel
	Number of iterations	80
	Number of agents	10
FA [54]	Number of fireflies	10
MVO [55]	Wormhole existence probability	[0.2, 1]
SC [56]	Inertia factor	[0, 1]

V. EXPERIMENTAL RESULTS

The experimental evaluations were performed on a Windows 11 PC with an Intel(R) Core(TM) i7 CPU operating at 2.40 GHz and 16GB of RAM. To implement the proposed approach, Python 3.9 was utilized. To assess the effectiveness of the proposed bSCWDTO method, thirty datasets from the UCI machine learning repository were selected. The datasets were divided into training, validation, and testing subsets, all assigned identical random sizes. The KNN classifier was trained using the training subset during the learning phase. The performance of the resulting model was evaluated using the testing subset, while the validation subset was employed for calculating the fitness function of a given solution. The experimental setup for the proposed approach and the competing methods are presented in Table 1 and Table 2, respectively. Each optimizer utilized ten search agents and executed 80 iterations over 20 independent runs. A k-fold cross-validation with a value of 10 was applied, using a k-nearest neighbors approach with a neighborhood size of 5 and employing the KNN classifier.

TABLE 2. Configuration parameters of the proposed bSCWDTO algorithm.

Parameter	Value
Iterations	100
Agents	10
Repetitions	20
Dimension	Number of features
Inertia factor of SC	0.1
Domain of Search	[0, 1]
w_1 in Eq. (13)	0.99
w_2 in Eq. (13)	0.01

TABLE 3. Evaluation metrics used in assessing the proposed feature selection method.

Metric	Value
Mean	$\frac{1}{M}\sum_{i=1}^{M} S_i^*$
Best Fitness	$min_{i=1}^{M} S_i^*$
Worst Fitness	$max_{i=1}^{M} S_i^*$
Average fitness size	$\frac{1}{M}\sum_{i=1}^{M} size(S_i^*)$
Average Error	$\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\sum_{i=1}^{N} mse(\hat{V}_i - V_i)$
Standard deviation	$\sqrt{\frac{1}{M-1}\sum_{i=1}^{M}\left(S_i^* - Mean\right)^2}$

A. EVALUATION METRICS

The achieved results were evaluated based on the criteria outlined in Table 3. These criteria were employed to assess the performance of the proposed feature selection method, as indicated by [57], [58], [59], [60], and [61]. Moreover, Table 3 provides information on the number of runs (M) performed by both the proposed algorithm and the competing optimizers. Within this table, the best solution obtained during the jth run is denoted as S_j^*, and its length is represented by $size(S_j^*)$. The variable N corresponds to the number of data points in the test set. Additionally, $\hat{V}_n$ and V_n refer to the predicted and actual values relevant to the evaluation process.

B. THE DATASETS

Experiments were run on thirty datasets in the UCI repository [62] to assess the efficacy of the proposed algorithm. A low-dimensional data set with low and high dimensions and following and small and small in this document has been accepted for evaluating the performance of the method provided herein regarding features and samples. Regarding compression, the KNN classification (with K = 5) employed in the packing approach is superior. Each dataset is then evaluated using a cross-comparison technique in the proposed method. K−1 is used for drilling and verification throughout the inspection process, while the rest of the stocks are put through folding tests. The sample size used for testing is identical to the training data size. Table 4 displays information about the utilized data set, including the feature and sample counts.

To ensure the accuracy and reliability of the feature selection algorithm, it is crucial to preprocess the dataset, considering the potential presence of missing values. When missing values are encountered in the dataset features, a preprocessing step is performed to handle them effectively. The missing values are imputed by averaging the previous and next non-missing values, providing a reasonable estimate. Additionally, scaling and normalization of the dataset values play a vital role in ensuring equitable consideration of all features by the machine learning model.

This study adopts the min-max scaler as a fundamental data scaling technique. This approach transforms and constrains the data features to a standardized range between 0 and 1. Using the min-max scaler, the data values are rescaled proportionally to their original range, preserving the distribution characteristics of the dataset. The min-max scaler equation utilized in this article is as follows:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

where X' represents the scaled value, X denotes the original value, $X_{\min}$ represents the minimum value in the dataset, and $X_{\max}$ represents the maximum value in the dataset. Applying the min-max scaler transforms the dataset to a consistent scale, enabling effective feature selection and subsequent analysis.

C. THE ACHIEVED RESULTS

The effectiveness of the proposed bSCWDTO optimizer is measured in terms of the evaluation criteria, including standard deviation, mean error, worst fitness, best fitness, mean fitness (Mean), and mean select size, and based on the thirty UCI machine learning datasets. Several optimization methods, both single and hybrid methods, are tested along with the proposed bSCWDTO algorithm to find which of them is performing best. The single methods are the binary variants of bDTO [46], bSC [56], bPSO [49], bWOA [50], bGWO [51], bMVO [55], bSBO [52], bGA [53], and bFA [54], where b denotes the binary output of the optimization method. To further elucidate the efficacy of the proposed algorithm, a hybrid algorithm, bGWO_GA [63], is included in the conducted experiments.

Table 5 displays the average error, Table 6 presents the average select size, and Table 7 presents the average fitness achieved by each optimization method. These tables show that the proposed bSCWDTO method achieves the best results for all the evaluation criteria and when tested on all UCI datasets. The proposed algorithm uses the proposed hybrid approach to the best solution, which includes the optimal subset of features that minimizes error. In these tables, the average of the chosen features serves as evidence of the efficiency of the proposed method. The selection of fewer features indicates that the optimizer is engaging in feature selection; nonetheless, keeping the error rate as low as possible is crucial.

TABLE 4. List of UCI datasets employed in this work.

No.	Dataset	No. Attr.	No. Inst.	No.	Dataset	No. Attr.	No. Inst.
1	Zoo	17	101	16	IonosphereEW	34	351
2	Breast cancer tissue	9	106	17	Fri_c0_500_10	10	500
3	Breast cancer Coimbra	9	116	18	Kc2	21	522
4	Lymphography	18	148	19	Climate	20	540
5	Hepatitis	10	155	20	WDBC	30	569
6	WineEW	13	178	21	Australian	14	690
7	Parkinsons	22	195	22	Breast_Cancer	8	700
8	SonarEW	60	208	23	Blood	4	744
9	Seeds	7	210	24	Segment	19	2310
10	Glass	9	214	25	Space-ga	6	3207
11	Lung cancer	21	226	26	WaveformEW	21	5000
12	SpectEW	22	267	27	Diabetic	19	1151
13	HeartEW	13	270	28	Mofn	10	1324
14	Vertebral	6	310	29	HAR	561	10299
15	Ionosphere	34	351	30	ISOLET	617	7797

TABLE 5. The average error measurements based on the selected features using various algorithms.

Dataset	bSCWDTO	bDTO	bSC	bPSO	bWAO	bGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Zoo	0.364	0.377	0.371	0.368	0.365	0.374	0.365	0.382	0.386	0.372	0.365
Breast cancer tissue	0.301	0.314	0.315	0.342	0.327	0.303	0.336	0.322	0.362	0.333	0.322
Breast cancer Coimbra	0.373	0.386	0.384	0.391	0.383	0.375	0.399	0.406	0.392	0.398	0.387
Lymphography	0.454	0.464	0.417	0.447	0.459	0.424	0.446	0.425	0.425	0.457	0.447
Hepatitis	0.302	0.320	0.321	0.318	0.309	0.318	0.321	0.296	0.330	0.308	0.314
WineEW	0.317	0.332	0.320	0.324	0.326	0.348	0.327	0.331	0.324	0.334	0.322
Parkinsons	0.407	0.416	0.421	0.423	0.416	0.416	0.415	0.416	0.405	0.410	0.422
SonarEW	0.209	0.215	0.216	0.211	0.210	0.216	0.211	0.214	0.216	0.213	0.213
Seeds	0.438	0.435	0.423	0.442	0.442	0.442	0.438	0.426	0.425	0.441	0.450
Glass	3.432	3.595	3.482	3.664	3.765	3.423	3.929	3.723	3.585	3.658	3.358
Lung cancer	0.301	0.313	0.305	0.314	0.334	0.313	0.312	0.324	0.336	0.324	0.319
SpectEW	0.324	0.326	0.330	0.332	0.330	0.327	0.327	0.332	0.334	0.330	0.332
HeartEW	0.389	0.412	0.398	0.399	0.396	0.397	0.391	0.381	0.388	0.398	0.393
Vertebral	0.199	0.216	0.210	0.233	0.201	0.200	0.213	0.219	0.236	0.224	0.236
Ionosphere	0.559	0.570	0.595	0.593	0.563	0.603	0.583	0.581	0.603	0.590	0.610
IonosphereEW	0.424	0.439	0.441	0.436	0.423	0.429	0.427	0.467	0.456	0.442	0.440
Fri_c0_500_10	0.432	0.445	0.439	0.436	0.433	0.442	0.433	0.450	0.454	0.440	0.433
Kc2	0.369	0.382	0.383	0.410	0.395	0.371	0.404	0.390	0.430	0.401	0.390
Climate	0.441	0.454	0.452	0.459	0.451	0.443	0.467	0.474	0.460	0.466	0.455
WDBC	0.385	0.400	0.388	0.392	0.394	0.416	0.395	0.399	0.392	0.402	0.390
Australian	0.277	0.283	0.284	0.279	0.278	0.284	0.279	0.282	0.284	0.281	0.281
Breast$_{C}ancer$	0.369	0.381	0.373	0.382	0.402	0.381	0.380	0.392	0.404	0.392	0.387
Blood	0.392	0.394	0.398	0.400	0.398	0.395	0.395	0.400	0.402	0.398	0.400
Segment	0.267	0.284	0.278	0.301	0.269	0.268	0.281	0.287	0.304	0.292	0.304
Space-ga	0.627	0.638	0.663	0.661	0.631	0.671	0.651	0.649	0.671	0.658	0.678
WaveformEW	0.506	0.503	0.491	0.510	0.510	0.510	0.506	0.494	0.493	0.509	0.518
Diabetes	3.500	3.663	3.550	3.732	3.833	3.491	3.997	3.791	3.653	3.726	3.426
Mofn	0.296	0.351	0.372	0.368	0.361	0.332	0.353	0.343	0.367	0.370	0.376
HAR	0.648	1.222	1.122	1.054	1.809	0.900	1.980	2.079	1.012	1.559	1.123
ISOLET	0.914	1.120	0.924	1.147	1.235	0.989	1.203	1.231	1.051	1.200	0.924

Table 8 shows the best fitness values, Table 9 displays the worst fitness values, and Table 10 presents the standard deviation values obtained from several distinct optimization strategies. From these tables, the proposed bSCWDTO algorithm is clearly shown to be the most stable and resilient of the algorithms tested, as evidenced by its low standard deviation compared to other algorithms. Based on the data, the proposed bSCWDTO algorithm can consistently outperform other optimization methods in fitness. All the datasets show a superior performance from the proposed bSCWDTO algorithm, demonstrating the proposed approach's capability to discover the best subset of features better than the other strategies.

Statistical analysis is performed to profoundly investigate the performance of the proposed optimization algorithm based on the achieved results. Table 11 presents the analysis

TABLE 6. Average select size of the selected features.

Dataset	bSCWDTO	bDTO	bSC	bPSO	bWAO	bGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Zoo	0.4735	0.5375	0.5325	0.6675	0.7625	0.5837	0.6375	0.6525	0.6125	0.6725	0.6325
Breast cancer tissue	0.2689	0.3992	0.4234	0.6234	0.4643	0.3866	0.6037	0.5628	0.5567	0.6204	0.5234
Breast cancer Coimbra	0.5391	0.6325	0.6325	0.6408	0.6408	0.6212	0.6492	0.5655	0.8325	0.6408	0.6408
Lymphography	0.6611	0.6396	0.6745	0.8325	0.8182	0.6937	0.6611	0.7325	0.6465	0.7182	0.6825
Hepatitis	0.3887	0.5439	0.5234	0.6007	0.6166	0.4415	0.6689	0.6225	0.6325	0.6030	0.5802
WineEW	0.4046	0.5611	0.5896	0.6646	0.8111	0.5039	0.6468	0.7039	0.6611	0.6575	0.6396
Parkinsons	0.7750	0.8325	0.7325	0.7825	0.9075	0.6761	0.8700	0.7825	0.7825	0.8950	0.9075
SonarEW	0.6090	0.6575	0.6325	0.7263	0.7700	0.6539	0.7263	0.7825	0.7825	0.7700	0.7138
Seeds	0.4758	0.6450	0.6575	0.7388	0.7700	0.5575	0.7200	0.6575	0.6575	0.7075	0.6950
Glass	0.6525	0.5075	0.4103	0.6158	0.6408	0.6536	0.6103	0.6658	0.6658	0.5908	0.5631
Lung cancer	0.3916	0.4939	0.5143	0.6257	0.6325	0.5234	0.6075	0.6234	0.6143	0.6234	0.5916
SpectEW	0.4325	0.4525	0.4725	0.4925	0.4575	0.4536	0.4675	0.4725	0.4525	0.4875	0.4675
HeartEW	0.9325	0.9325	0.9992	0.9492	0.9825	0.9636	1.0158	0.9465	0.9325	1.0158	0.9825
Vertebral	0.7725	0.9850	0.9825	0.8225	1.1100	0.8625	0.9825	0.9225	0.8025	0.8800	1.0025
Ionosphere	0.7809	0.8643	0.8500	0.9167	1.2309	0.9357	0.9762	0.9357	1.0214	0.9428	0.9881
IonosphereEW	0.4594	0.6146	0.5941	0.6714	0.6873	0.5122	0.7396	0.6932	0.7032	0.6737	0.6509
Fri_c0_500_10	0.4753	0.6318	0.6603	0.7353	0.8818	0.5746	0.7175	0.7746	0.7318	0.7282	0.7103
Kc2	0.8457	0.9032	0.8032	0.8532	0.9782	0.7468	0.9407	0.8532	0.8532	0.9657	0.9782
Climate	0.6797	0.7282	0.7032	0.7970	0.8407	0.7246	0.7970	0.8532	0.8532	0.8407	0.7845
WDBC	0.5465	0.7157	0.7282	0.8095	0.8407	0.6282	0.7907	0.7282	0.7282	0.7782	0.7657
Australian	0.7232	0.5782	0.4810	0.6865	0.7115	0.7243	0.6810	0.7365	0.7365	0.6615	0.6338
Breast$_{C}$*ancer*	0.4623	0.5646	0.5850	0.6964	0.7032	0.5941	0.6782	0.6941	0.6850	0.6941	0.6623
Blood	0.5032	0.5232	0.5432	0.5632	0.5282	0.5243	0.5382	0.5432	0.5232	0.5582	0.5382
Segment	1.0032	1.0032	1.0699	1.0199	1.0532	1.0343	1.0865	1.0172	1.0032	1.0865	1.0532
Space-ga	0.8432	1.0557	1.0532	0.8932	1.1807	0.9332	1.0532	0.9932	0.8732	0.9507	1.0732
WaveformEW	0.6484	0.7318	0.7175	0.7842	1.0984	0.8032	0.8437	0.8032	0.8889	0.8103	0.8556
Diabetes	0.6643	0.7365	0.6778	0.8143	0.9532	0.6766	0.8143	0.6735	0.7929	0.8365	0.8143
Mofn	0.3849	0.8032	0.4008	0.8682	1.0632	0.4102	0.8482	0.6111	0.6356	0.8782	0.8882
HAR	0.7703	1.0877	0.9784	1.0826	1.1067	0.9019	1.1197	1.1582	0.9153	1.1593	0.9746
ISOLET	0.9019	1.0803	0.9817	1.0173	1.1683	0.9745	1.1388	1.1846	0.9897	1.1364	0.9984

TABLE 7. The average fitness of the selected features.

Dataset	bSCWDTO	bDTO	bSC	bPSO	bWAO	bGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Zoo	0.3816	0.4175	0.3998	0.4039	0.4049	0.3833	0.4049	0.4118	0.3758	0.4117	0.4049
Breast cancer tissue	0.3033	0.3382	0.3048	0.3657	0.3509	0.3349	0.3598	0.3118	0.3508	0.3564	0.3458
Breast cancer Coimbra	0.4994	0.5474	0.3738	0.5526	0.5445	0.3798	0.5603	0.4948	0.4818	0.5589	0.5483
Lymphography	0.5211	0.5510	0.6108	0.5347	0.5460	0.5445	0.5333	0.5848	0.6048	0.5439	0.5340
Hepatitis	0.3272	0.3196	0.3098	0.3364	0.3272	0.3281	0.3387	0.3458	0.3198	0.3265	0.3318
WineEW	0.4698	0.4847	0.3098	0.4765	0.4788	0.3385	0.4801	0.3211	0.3142	0.4870	0.4748
Parkinsons	1.0087	1.0285	0.4114	1.0357	1.0289	0.4062	1.0279	0.4058	0.3954	1.0206	1.0345
SonarEW	0.4802	0.4961	0.5066	0.4922	0.4912	0.5036	0.4918	0.7040	0.5057	0.4939	0.4944
Seeds	0.7335	0.7425	0.7431	0.7496	0.7491	0.7426	0.7452	0.7502	0.7427	0.7485	0.7570
Glass	3.2304	3.5789	4.5446	3.6466	3.7466	3.6466	3.9092	4.8956	5.4058	3.6405	3.3435
Lung cancer	0.3189	0.3310	0.3353	0.3318	0.3516	0.3230	0.3303	0.3337	0.3260	0.3425	0.3371
SpectEW	1.5460	1.5479	0.3203	1.5535	1.5520	0.2971	1.5492	0.3225	0.3241	1.5517	1.5538
HeartEW	2.8234	2.8466	2.8741	2.8331	2.8306	2.8760	2.8256	3.1041	2.8781	2.8327	2.8273
Vertebral	1.3326	1.4391	1.5996	1.4560	1.4244	1.5051	1.4359	1.7092	1.8264	1.4467	1.4586
Ionosphere	1.3393	1.3498	1.5854	1.3726	1.3426	1.4804	1.3624	1.5716	0.5933	1.3695	1.3891
IonosphereEW	0.7668	0.7821	0.9468	0.7791	0.7662	0.8006	0.7698	0.7771	0.8465	0.7852	0.7827
Fri_c0_500_10	0.4698	0.4847	0.3098	0.4765	0.4788	0.3385	0.4801	0.3211	0.3142	0.4870	0.4748
Kc2	0.8519	0.8717	0.2546	0.8789	0.8721	0.2494	0.8711	0.2490	0.2386	0.8638	0.8777
Climate	0.5249	0.5408	0.5513	0.5369	0.5359	0.5483	0.5365	0.7487	0.5504	0.5386	0.5391
WDBC	0.7782	0.7872	0.7878	0.7943	0.7938	0.7873	0.7899	0.7949	0.7874	0.7932	0.8017
Australian	3.2751	3.6236	4.5893	3.6913	3.7913	3.6913	3.9539	4.9403	5.4505	3.6852	3.3882
Breast$_{C}$*ancer*	0.3636	0.3757	0.3800	0.3765	0.3963	0.3677	0.3750	0.3784	0.3707	0.3872	0.3818
Blood	1.5907	1.5926	0.3650	1.5982	1.5967	0.3418	1.5939	0.3672	0.3688	1.5964	1.5985
Segment	2.8681	2.8913	2.9188	2.8778	2.8753	2.9207	2.8703	3.1488	2.9228	2.8774	2.8720
Space-ga	1.3773	1.4838	1.6443	1.5007	1.4691	1.5498	1.4806	1.7539	1.8711	1.4914	1.5033
WaveformEW	1.3840	1.3945	1.6301	1.4173	1.3873	1.5251	1.4071	1.6163	0.6380	1.4142	1.4338
Diabetes	0.8115	0.8268	0.9915	0.8238	0.8109	0.8453	0.8145	0.8218	0.8912	0.8299	0.8274
Mofn	0.7051	0.7587	0.7194	0.7751	0.7685	0.7229	0.7604	0.7809	0.7339	0.7774	0.7837
HAR	0.7458	1.0582	0.9906	1.0571	1.1150	0.9669	1.1363	1.1760	0.9909	1.1240	0.9756
ISOLET	0.8786	1.0458	0.9771	1.0827	1.1471	0.9780	1.1327	1.1871	1.0016	1.1163	1.0027

results. This table shows that the proposed algorithm is stable as the values 10%, 25%, 75%, and 90% corresponding to the proposed approach are better than those achieved by the other methods. In addition, the values of mean, median, standard deviation, lower, upper, Skewness, and Kurtosis are all the best for the proposed algorithm if compared to other methods.

TABLE 8. The best fitness of the selected features.

Dataset	bSCWDTO	bDTO	bSC	bPSO	bWAO	bGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Zoo	0.2766	0.2960	0.3736	0.2960	0.3348	0.3242	0.3542	0.3542	0.3542	0.3348	0.2960
Breast cancer tissue	0.2416	0.2585	0.2839	0.2924	0.2670	0.2917	0.2501	0.2754	0.3262	0.2416	0.2754
Breast cancer Coimbra	0.4548	0.4740	0.5221	0.4644	0.4644	0.4632	0.4836	0.5221	0.4932	0.4836	0.4644
Lymphography	0.2469	0.3318	0.3883	0.3035	0.3035	0.3136	0.3742	0.3742	0.4166	0.3883	0.3318
Hepatitis	0.2074	0.2221	0.2526	0.2678	0.2374	0.3135	0.2374	0.2374	0.2983	0.2221	0.2374
WineEW	0.4231	0.4360	0.4446	0.4317	0.4274	0.4403	0.4446	0.4532	0.4532	0.4403	0.4360
Parkinsons	0.9787	0.9787	0.9826	0.9866	0.9787	1.0036	0.9866	0.9847	0.9747	0.9747	0.9747
SonarEW	0.4556	0.4594	0.4764	0.4636	0.4552	0.4619	0.4594	0.4721	0.4764	0.4509	0.4636
Seeds	0.6758	0.6951	0.7105	0.7028	0.6835	0.7011	0.7028	0.6951	0.6989	0.6912	0.6912
Glass	2.1143	2.0537	3.0032	2.0537	1.6294	1.5688	1.8112	3.9730	4.0337	2.1143	1.8920
Lung cancer	0.2374	0.2374	0.2678	0.2374	0.2526	0.2811	0.2374	0.2526	0.2983	0.2374	0.2526
SpectEW	1.5161	1.5173	1.5318	1.5310	1.5289	1.5318	1.5273	1.5313	1.5378	1.5173	1.5326
HeartEW	2.7908	2.7908	2.7827	2.7908	2.7908	2.7835	2.7840	2.7927	2.7827	2.7908	2.7908
Vertebral	1.3401	1.4177	1.4141	1.4201	1.4085	1.3587	1.4109	1.4274	1.4282	1.4254	1.4306
Ionosphere	1.2640	1.3014	1.3484	1.2955	1.2747	1.3548	1.3145	1.3109	1.3359	1.2877	1.3204
IonosphereEW	0.7969	0.7969	0.8155	0.8000	0.8124	0.8221	0.7969	0.8186	0.8155	0.8062	0.8093
Fri_c0_500_10	0.5096	0.5225	0.5311	0.5182	0.5139	0.5268	0.5311	0.5397	0.5397	0.5268	0.5225
Kc2	1.0652	1.0652	1.0691	1.0731	1.0652	1.0901	1.0731	1.0712	1.0612	1.0612	1.0612
Climate	0.5421	0.5459	0.5629	0.5501	0.5417	0.5484	0.5459	0.5586	0.5629	0.5374	0.5501
WDBC	0.7623	0.7816	0.7970	0.7893	0.7700	0.7876	0.7893	0.7816	0.7854	0.7777	0.7777
Australian	2.2008	2.1402	3.0897	2.1402	1.7159	1.6553	1.8977	4.0595	4.1202	2.2008	1.9785
Breast_Cancer	0.3239	0.3239	0.3543	0.3239	0.3391	0.3676	0.3239	0.3391	0.3848	0.3239	0.3391
Blood	1.6026	1.6038	1.6183	1.6175	1.6154	1.6183	1.6138	1.6178	1.6243	1.6038	1.6191
Segment	2.8773	2.8773	2.8692	2.8773	2.8773	2.8700	2.8705	2.8792	2.8692	2.8773	2.8773
Space-ga	1.4266	1.5042	1.5006	1.5066	1.4950	1.4452	1.4974	1.5139	1.5147	1.5119	1.5171
WaveformEW	1.3505	1.3879	1.4349	1.3820	1.3612	1.4413	1.4010	1.3974	1.4224	1.3742	1.4069
Diabetes	0.7969	0.7969	0.8155	0.8000	0.8124	0.8221	0.7969	0.8186	0.8155	0.8062	0.8093
Mofn	0.6844	0.7181	0.7113	0.7315	0.7562	0.7013	0.7181	0.6956	0.7966	0.7719	0.7472
HAR	0.7335	1.0666	0.9957	1.0477	1.1438	0.9400	1.1141	1.1341	0.9389	1.1449	0.9953
ISOLET	0.8437	1.0379	0.9445	1.0790	1.1457	0.9668	1.1461	1.0330	1.0113	1.1227	1.0229

TABLE 9. The worst fitness of the selected features.

Dataset	bSCWDTO	bDTO	bSC	bPSO	bWAO	bGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Zoo	0.5116	0.5306	0.4918	0.5500	0.5500	0.5306	0.5112	0.5112	0.5112	0.5695	0.5112
Breast cancer tissue	0.4334	0.4429	0.3921	0.4514	0.4683	0.4514	0.4344	0.4175	0.4768	0.4852	0.4937
Breast cancer Coimbra	0.6371	0.6203	0.5819	0.6491	0.6491	0.6587	0.6876	0.6876	0.6299	0.7933	0.6876
Lymphography	0.6974	0.8390	0.6269	0.7966	0.7966	0.6069	0.7259	0.6269	0.6269	0.7400	0.8107
Hepatitis	0.4286	0.4150	0.4302	0.4607	0.4455	0.4455	0.4455	0.4302	0.4607	0.4302	0.4302
WineEW	0.5343	0.5842	0.5282	0.5670	0.5497	0.6229	0.5325	0.5239	0.5239	0.6832	0.5411
Parkinsons	1.1226	1.1226	1.1067	1.1226	1.1027	1.1067	1.1226	1.1067	1.1067	1.0828	1.1226
SonarEW	0.5476	0.5637	0.5467	0.5424	0.5382	0.5552	0.5424	0.5424	0.5339	0.5509	0.5552
Seeds	0.8154	0.8439	0.7666	0.8091	0.8284	0.8130	0.8323	0.7743	0.7820	0.8207	0.8942
Glass	5.4700	5.4279	5.2258	6.1148	5.6501	5.8148	5.6703	6.8422	7.4887	5.2865	5.3067
Lung cancer	0.4163	0.4607	0.4150	0.4759	0.5064	0.4555	0.4302	0.4455	0.4659	0.4302	0.4759
SpectEW	1.5892	1.5912	1.5980	1.6024	1.5984	1.5754	1.5956	1.6004	1.5932	1.6036	1.5996
HeartEW	2.6732	3.2959	2.8651	2.9205	2.9677	2.8394	2.9056	2.8745	2.8772	3.0488	2.9191
Vertebral	1.3532	1.4860	1.4703	1.5277	1.4752	1.4832	1.4832	1.4800	1.5004	1.4912	1.5277
Ionosphere	1.3761	1.4780	1.4376	1.4561	1.3984	1.4323	1.4430	1.4471	1.4323	1.4644	1.4792
IonosphereEW	0.8465	0.8813	0.8720	0.9371	0.8471	0.8751	0.8689	0.8937	0.8782	0.8689	0.8565
Fri_c0_500_10	1.2396	1.2396	1.2237	1.2396	1.2197	1.2237	1.2396	1.2237	1.2237	1.1998	1.2396
Kc2	0.6646	0.6807	0.6637	0.6595	0.6552	0.6722	0.6595	0.6595	0.6510	0.6680	0.6722
Climate	0.9324	0.9609	0.8836	0.9261	0.9455	0.9300	0.9493	0.8913	0.8990	0.9377	1.0112
WDBC	5.5870	5.5449	5.3429	6.2318	5.7672	5.9318	5.7874	6.9592	7.6057	5.4035	5.4237
Australian	0.5334	0.5777	0.5320	0.5929	0.6234	0.5725	0.5473	0.5625	0.5829	0.5473	0.5929
Breast_Cancer	1.7062	1.7082	1.7150	1.7194	1.7154	1.6924	1.7126	1.7174	1.7102	1.7206	1.7166
Blood	2.7902	3.4129	2.9821	3.0375	3.0847	2.9564	3.0226	2.9915	2.9942	3.1658	3.0361
Segment	1.4702	1.6030	1.5873	1.6448	1.5922	1.6002	1.6002	1.5970	1.6175	1.6082	1.6448
Space-ga	1.4931	1.5951	1.5547	1.5731	1.5154	1.5493	1.5600	1.5642	1.5493	1.5814	1.5963
WaveformEW	0.9635	0.9983	0.9890	1.0542	0.9642	0.9921	0.9859	1.0107	0.9952	0.9859	0.9735
Diabetes	0.6513	0.7012	0.6452	0.6840	0.6668	0.7399	0.6495	0.6409	0.6409	0.8002	0.6582
Mofn	0.8847	0.9328	0.8767	0.9440	0.9104	0.9052	0.9328	0.8879	0.8857	0.9059	0.9328
HAR	0.8721	1.1709	1.1723	1.1821	1.2922	1.0926	1.2500	1.1501	1.1049	1.2369	1.1145
ISOLET	0.9844	1.1656	1.1178	1.1944	1.2690	1.1191	1.2589	1.4285	1.1468	1.2416	1.1378

These results emphasize the superiority of the proposed optimization algorithm for solving the feature selection problem.

On the other hand, the one-way analysis-of-variance (ANOVA) test is conducted to study whether there is any

TABLE 10. The standard deviation fitness of the selected features.

Dataset	bSCWDTO	bDTO	bSC	bPSO	bWAO	bGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Zoo	0.1785	0.1975	0.1811	0.2052	0.1918	0.1880	0.1854	0.1931	0.1899	0.1962	0.1959
Breast cancer tissue	0.1786	0.1881	0.1839	0.1789	0.1879	0.1833	0.1853	0.1832	0.1921	0.1936	0.1931
Breast cancer Coimbra	0.1716	0.1744	0.1742	0.1856	0.1833	0.1832	0.1860	0.1967	0.1812	0.2096	0.1906
Lymphography	0.2386	0.2618	0.2271	0.2494	0.2528	0.2407	0.2250	0.2450	0.2315	0.2287	0.2513
Hepatitis	0.1775	0.1869	0.1973	0.1885	0.1981	0.1810	0.1984	0.2032	0.1933	0.1900	0.1831
WineEW	0.1610	0.1776	0.1645	0.1675	0.1692	0.1797	0.1610	0.1623	0.1649	0.1911	0.1635
Parkinsons	0.1659	0.1707	0.1794	0.1741	0.1704	0.1682	0.1706	0.1859	0.1817	0.1665	0.1770
SonarEW	0.1583	0.1598	0.1582	0.1549	0.1567	0.1563	0.1571	0.1572	0.1520	0.1588	0.1589
Seeds	0.1614	0.1758	0.1631	0.1626	0.1681	0.1772	0.1723	0.1621	0.1646	0.1711	0.1807
Glass	0.9864	1.0470	1.0260	1.2820	0.9928	1.9081	1.1177	1.2778	1.5125	1.0470	1.0836
Lung cancer	0.1863	0.1910	0.1900	0.1951	0.2014	0.1902	0.1921	0.2088	0.2016	0.1910	0.1932
SpectEW	0.1523	0.1540	0.1555	0.1534	0.1550	0.1523	0.1534	0.1564	0.1515	0.1573	0.1519
HeartEW	0.1604	0.2453	0.1610	0.1726	0.1777	0.1904	0.1649	0.1643	0.1685	0.1917	0.1671
Vertebral	0.1516	0.1543	0.1521	0.1619	0.1522	0.1563	0.1539	0.1540	0.1582	0.1531	0.1603
Ionosphere	0.1668	0.1772	0.1662	0.1750	0.1711	0.1770	0.1769	0.1881	0.1686	0.1790	0.1793
IonosphereEW	0.1660	0.1801	0.1852	0.1693	0.1750	0.1677	0.1660	0.1890	0.1875	0.1752	0.1755
Fri_c0_500_10	0.2510	0.2676	0.2545	0.2575	0.2592	0.2697	0.2510	0.2523	0.2549	0.2811	0.2535
Kc2	0.2559	0.2607	0.2694	0.2641	0.2604	0.2582	0.2606	0.2759	0.2717	0.2565	0.2670
Climate	0.2483	0.2498	0.2482	0.2449	0.2467	0.2463	0.2471	0.2472	0.2420	0.2488	0.2489
WDBC	0.2514	0.2658	0.2531	0.2526	0.2581	0.2672	0.2623	0.2521	0.2546	0.2611	0.2707
Australian	1.0764	1.1370	1.1160	1.3720	1.0828	1.9981	1.2077	1.3678	1.6025	1.1370	1.1736
Breast_Cancer	0.2763	0.2810	0.2800	0.2851	0.2914	0.2802	0.2821	0.2988	0.2916	0.2810	0.2832
Blood	0.2423	0.2440	0.2455	0.2434	0.2450	0.2423	0.2434	0.2464	0.2415	0.2473	0.2419
Segment	0.2504	0.3353	0.2510	0.2626	0.2677	0.2804	0.2549	0.2543	0.2585	0.2817	0.2571
Space-ga	0.2416	0.2443	0.2421	0.2519	0.2422	0.2463	0.2439	0.2440	0.2482	0.2431	0.2503
WaveformEW	0.2568	0.2672	0.2562	0.2650	0.2611	0.2670	0.2669	0.2781	0.2586	0.2690	0.2693
Diabetes	0.2560	0.2701	0.2752	0.2593	0.2650	0.2577	0.2560	0.2790	0.2775	0.2652	0.2655
Mofn	0.2655	0.2740	0.2764	0.2818	0.2691	0.3047	0.2567	0.2847	0.2424	0.2752	0.2522
HAR	0.2525	0.2690	0.2655	0.2668	0.2900	0.2578	0.2812	0.2857	0.5367	0.2801	0.2657
ISOLET	0.2589	0.2770	0.2734	0.2755	0.2979	0.2657	0.2868	0.2879	0.5596	0.2880	0.2658

TABLE 11. The statistical analysis of the achieved results using the proposed and other competing methods.

	bSCWDTO	bDTO	bSC	bPSO	bWAO	bGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Number of values	12	12	12	12	12	12	12	12	12	12	12
Minimum	0.6072	0.6451	0.6256	0.6489	0.6782	0.6167	0.7054	0.6769	0.6427	0.6742	0.6237
25% Percentile	0.6072	0.6551	0.6356	0.6589	0.6835	0.6267	0.7054	0.6969	0.6527	0.6774	0.6337
Median	0.6072	0.6551	0.6356	0.6589	0.6882	0.6267	0.7054	0.6969	0.6527	0.6774	0.6337
75% Percentile	0.6072	0.6551	0.6356	0.6589	0.6882	0.6267	0.7054	0.6969	0.6527	0.6849	0.6337
Maximum	0.6072	0.6651	0.6561	0.6789	0.6882	0.6467	0.7254	0.6969	0.6727	0.6974	0.6687
Range	0	0.02	0.03049	0.03	0.01	0.03	0.02	0.02	0.03	0.02318	0.04501
10% Percentile	0.6072	0.6481	0.6286	0.6519	0.6782	0.6197	0.7054	0.6799	0.6457	0.6752	0.6267
90% Percentile	0.6072	0.6621	0.6499	0.6759	0.6882	0.6437	0.7224	0.6969	0.6667	0.6974	0.6612
Actual confidence level	96.14%	96.14%	96.14%	96.14%	96.14%	96.14%	96.14%	96.14%	96.14%	96.14%	96.14%
Lower confidence limit	0.6072	0.6551	0.6356	0.6589	0.6819	0.6267	0.7054	0.6969	0.6527	0.6774	0.6337
Upper confidence limit	0.6072	0.6551	0.6356	0.6589	0.6882	0.6267	0.7054	0.6969	0.6527	0.6874	0.6337
Mean	0.6072	0.6551	0.6365	0.6605	0.686	0.6284	0.7079	0.6944	0.6535	0.6813	0.6366
Std. Deviation	0	0.0042	0.0068	0.0071	0.0040	0.0071	0.0062	0.0062	0.0066	0.0081	0.0109
Std. Error of Mean	0	0.0012	0.0019	0.0020	0.0011	0.0020	0.0017	0.0017	0.0019	0.0023	0.0031
Lower 95% CI of mean	0.6072	0.6524	0.6322	0.656	0.6834	0.6238	0.704	0.6904	0.6493	0.6762	0.6296
Upper 95% CI of mean	0.6072	0.6578	0.6408	0.6651	0.6886	0.633	0.7119	0.6983	0.6578	0.6865	0.6436
Coefficient of variation	0.000%	0.650%	1.070%	1.087%	0.593%	1.142%	0.878%	0.895%	1.023%	1.194%	1.723%
Geometric mean	0.6072	0.6551	0.6364	0.6605	0.686	0.6284	0.7079	0.6944	0.6535	0.6813	0.6365
Geometric SD factor	1	1.007	1.011	1.011	1.006	1.011	1.009	1.009	1.01	1.012	1.017
Lower 95% CI of geo. mean	0.6072	0.6524	0.6322	0.656	0.6834	0.6239	0.704	0.6904	0.6493	0.6762	0.6297
Upper 95% CI of geo. mean	0.6072	0.6578	0.6408	0.6651	0.6886	0.6329	0.7118	0.6984	0.6577	0.6864	0.6434
Harmonic mean	0.6072	0.6551	0.6364	0.6605	0.686	0.6283	0.7079	0.6943	0.6535	0.6812	0.6364
Lower 95% CI of harm. mean	0.6072	0.6524	0.6322	0.656	0.6834	0.6239	0.704	0.6903	0.6493	0.6762	0.6298
Upper 95% CI of harm. mean	0.6072	0.6578	0.6407	0.665	0.6886	0.6329	0.7118	0.6984	0.6577	0.6864	0.6432
Quadratic mean	0.6072	0.6551	0.6365	0.6606	0.686	0.6284	0.7079	0.6944	0.6536	0.6814	0.6367
Lower 95% CI of quad. mean	0.6072	0.6524	0.6321	0.656	0.6834	0.6238	0.7039	0.6905	0.6493	0.6761	0.6296
Upper 95% CI of quad. mean	0.6072	0.6578	0.6409	0.6652	0.6886	0.633	0.7119	0.6983	0.6579	0.6866	0.6438
Skewness		0	2.157	1.508	-1.501	1.508	2.555	-2.555	2.104	1.57	2.565
Kurtosis		5.5	7.813	4.065	0.5225	4.065	6.242	6.242	7.698	1.014	7.85
Sum	7.286	7.861	7.638	7.927	8.232	7.541	8.495	8.333	7.843	8.176	7.639

statistical difference between the means of optimization methods included in the conducted experiments. Table 12 depicts the test results. From this table, it can be noted the p-value is less than 0.05, and F is 255.9. Therefore, there is a statistically significant difference between the means of the optimization methods. Moreover, there is a statistically

TABLE 12. ANOVA test results when the proposed feature selection method and the other methods are applied to the adopted datasets.

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment	0.1146	10	0.01146	F (10, 121) = 255.9	P<0.0001
Residual	0.005419	121	0.00004478		
Total	0.12	131			

TABLE 13. p-values of Wilcoxon's rank-sum using the proposed approach compared to other methods ($p > 0.05$ are underlined).

Dataset	bDTO	bSC	bPSO	bWAO	bGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Zoo	6.69E-05	7.58E-05	7.58E-05	1.80E-01	6.69E-05	6.26E-02	7.58E-05	6.59E-05	6.69E-05	7.89E-02
Breast cancer tissue	6.69E-05	7.58E-05	7.58E-05	6.69E-05	7.62E-02	6.69E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05
Breast cancer Coimbra	6.69E-05	7.58E-05	7.58E-05	6.69E-05	7.82E-02	6.69E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05
Lymphography	6.69E-05	7.58E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05
Hepatitis	6.69E-05	7.58E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05
WineEW	6.69E-05	7.58E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.69E-05	6.69E-05	6.59E-05
Parkinsons	3.67E-04	7.58E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	7.85E-02	6.69E-05	6.69E-05
SonarEW	3.67E-04	7.58E-05	7.58E-05	6.34E-02	6.69E-05	6.69E-05	7.58E-05	6.69E-05	6.69E-05	6.59E-05
Seeds	3.67E-04	7.58E-05	7.58E-05	6.69E-05	6.69E-05	6.62E-02	7.58E-05	6.59E-05	6.69E-05	6.69E-05
Glass	3.67E-04	7.58E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.06E-02
Lung cancer	3.67E-04	7.58E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
SpectEW	3.67E-04	7.58E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
HeartEW	3.67E-04	7.58E-05	8.90E-02	6.69E-05	6.69E-05	6.69E-05	7.58E-05	9.92E-02	6.69E-05	6.69E-05
Vertebral	3.67E-04	7.58E-05	6.69E-05	6.69E-05	5.88E-02	6.69E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05
Ionosphere	3.67E-04	7.58E-05	8.72E-02	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05
IonosphereEW	3.67E-04	7.58E-05	6.69E-05	8.86E-02	6.69E-05	6.69E-05	7.58E-05	6.69E-05	6.69E-05	6.69E-05
Fri_c0_500_10	3.67E-04	7.58E-05	6.69E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
Kc2	3.67E-04	7.58E-05	6.59E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
Climate	9.16E-02	3.25E-01	6.69E-05	6.69E-05	6.59E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	7.33E-02
WDBC	3.67E-04	7.58E-05	6.59E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
Australian	3.67E-04	7.58E-05	6.59E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
Breast_Cancer	3.67E-04	7.58E-05	6.59E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
Blood	3.67E-04	7.58E-05	6.59E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
Segment	3.67E-04	7.58E-05	6.59E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
Space-ga	3.67E-04	7.58E-05	6.59E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
WaveformEW	3.67E-04	7.58E-05	7.58E-05	7.58E-05	7.58E-05	7.58E-05	7.58E-05	7.58E-05	7.58E-05	6.69E-05
Diabetes	3.67E-04	6.59E-05	6.59E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
Mofn	3.67E-04	6.59E-05	6.59E-05	9.78E-02	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
HAR	3.67E-04	6.59E-05	6.59E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.59E-05	6.69E-05	6.69E-05
ISOLET	3.67E-04	6.59E-05	6.59E-05	6.69E-05	6.69E-05	6.69E-05	7.58E-05	6.68E-02	6.69E-05	6.69E-05

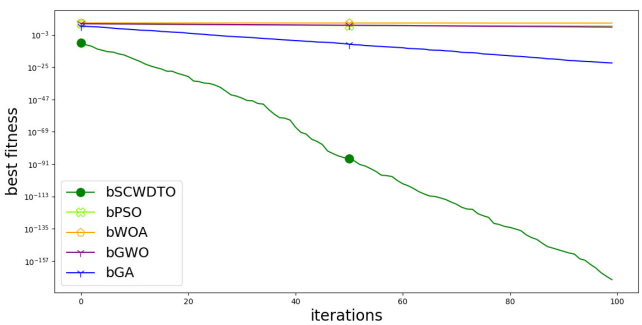


FIGURE 2. Convergence time of the proposed approach with comparison to other approaches.

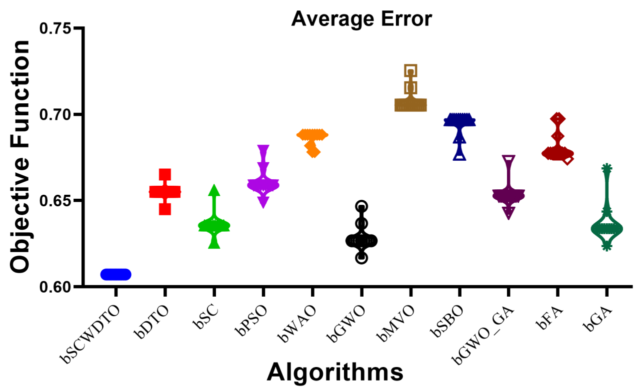


FIGURE 3. Average error of the results achieved by the proposed and other optimization algorithms.

significant difference between the means of the optimization methods.

In addition, the Wilcoxon rank-sum test is performed to determine the significance levels of the proposed bSCWDTO algorithm concerning existing meta-heuristic algorithms. When comparing the proposed algorithm's output to other algorithms, this test can assist in revealing whether or not the results differ significantly. The proposed algorithm's findings differ substantially from the compared methods if the p-value is less than 0.05. In contrast, insignificant results are indicated by a p-value greater than 0.05. The worst p-values in Table 13 are those larger than 0.05, which is the significance level used. When comparing the proposed technique to others, the table shows that the p-values obtained using this test are less

TABLE 14. The time (seconds) consumed by the feature selection algorithms applied to the UCI benchmark datasets.

Dataset	bSCWDTO	bDTO	bSC	bPSO	bWAO	bGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Zoo	3.91	4.55	4.14	4.23	5.63	4.04	4.32	4.49	4.96	4.68	5.26
Breast cancer tissue	4.32	6.26	5.39	5.83	5.58	5.16	4.86	5.77	5.79	6.06	5.84
Breast cancer Coimbra	3.58	4.77	5.01	4.85	4.94	3.59	4.27	4.54	4.73	5.05	5.10
Lymphography	3.58	4.24	4.37	5.09	4.60	3.68	3.0.2	4.86	4.32	4.22	4.18
Hepatitis	3.58	4.52	3.97	4.30	4.81	4.26	4.31	4.43	5.42	4.58	4.51
WineEW	5.65	6.91	7.02	6.83	8.16	7.00	5.97	6.81	7.55	7.07	7.02
Parkinsons	5.66	5.38	5.96	3.86	6.48	5.77	6.07	6.91	5.73	6.81	7.22
SonarEW	5.46	6.77	5.96	7.08	6.40	6.45	6.28	7.12	7.14	6.84	6.66
Seeds	8.32	9.33	9.95	8.82	10.64	8.95	9.32	9.30	10.05	9.83	8.89
Glass	5.62	5.94	6.30	6.63	7.03	7.50	6.50	7.83	7.52	6.61	8.48
Lung cancer	6.10	6.74	6.33	6.42	7.82	6.23	6.51	6.68	7.15	6.87	7.45
SpectEW	6.51	8.44	7.58	8.02	7.77	7.35	7.05	7.96	7.98	8.25	8.03
HeartEW	5.77	6.96	7.20	7.04	7.13	5.78	6.46	6.73	6.92	7.24	7.29
Vertebral	5.77	6.43	6.56	7.28	6.79	5.87	3.0.2	7.05	6.51	6.41	6.37
Ionosphere	5.77	6.71	6.16	6.49	7.00	6.44	6.50	6.62	7.61	6.77	6.70
IonosphereEW	7.83	9.10	9.21	9.02	10.35	9.19	8.16	9.00	9.74	9.26	9.21
$Fri_{0\_500\_10}$	7.84	7.57	8.15	6.05	8.67	7.96	8.26	9.10	7.92	9.00	9.41
Kc2	7.65	8.96	8.15	9.27	8.59	8.64	8.47	9.31	9.33	9.03	8.85
Climate	10.51	11.52	12.14	11.00	12.83	11.14	11.51	11.49	12.24	12.02	11.08
WDBC	7.81	8.13	8.49	8.82	9.22	9.69	8.69	10.02	9.71	8.80	10.67
Australian	10.51	11.52	12.14	11.00	12.83	11.14	11.51	11.49	12.24	12.02	11.08
$Breast_{C}ancer$	6.43	6.97	7.35	8.12	9.15	6.63	7.27	7.15	7.68	7.93	7.70
Blood	10.56	11.81	12.01	10.61	13.11	11.71	12.34	11.06	13.34	12.35	11.06
Segment	54.57	108.47	118.93	83.41	138.81	76.18	62.58	83.52	91.05	86.15	84.39
Space-ga	14.51	19.28	17.31	17.84	22.26	18.70	19.06	20.75	96.82	34.83	20.31
WaveformEW	102.78	135.22	139.79	144.71	153.24	730.11	107.33	165.38	153.20	143.16	172.33
Diabetes	35.86	55.51	66.09	61.92	87.35	71.81	74.06	57.30	62.24	53.44	59.52
Mofn	12.92	14.41	14.58	14.40	13.45	13.90	15.15	14.16	15.61	15.12	14.43
HAR Using Smartphones	326.69	463.26	443.07	474.04	606.82	631.05	423.45	620.34	614.91	466.79	607.34
ISOLET	433.01	496.45	484.36	462.76	621.26	737.21	453.07	643.23	673.81	496.16	699.33
Average Time	37.64	48.74	48.46	47.52	60.96	81.44	46.76	59.35	62.98	49.11	61.19

than 0.05. The results prove the statistical significance and superiority of the bSCWDTO method.

To prove the efficiency of the proposed algorithm, additional experiment is conducted to measure the convergence time of the feature selection process consumed by the proposed approach and the other feature selection methods. Figure 2. In this figure, it can be noted that the proposed algorithm can reach the optimal set of features in less number of iterations when compared to the other methods. In addition, the measurement of the run time consumed by each algorithm when applied to each set of the 30 benchmark datasets is presented in Table 14. This table clearly shows that the proposed algorithm achieves the smallest time required to find the best set of features for each dataset. The average time in the last row of this table emphasizes the speed of the proposed approach when compared to the other feature selection methods.

D. VISUAL REPRESENTATION OF THE RESULTS

The visual representation of the results achieved by the proposed method in comparison to the other methods is depicted in Figure 5, Figure 4, and Figure 3. The comparison depicted in these figures emphasizes the proposed approach's effectiveness and superiority compared to the other methods.

The proposed algorithm's stability is depicted here compared to existing methods. Figure 6 shows the averaged error, average size, average mean, best fitness, worst fitness, and standard deviation fitness across all the thirty datasets investigated using the various optimization methods. It is clear from these results that the proposed bSCWDTO algorithm outperforms all of the competing optimization methods.

E. DISCUSSION

The primary objective of our research was to evaluate the performance of the proposed feature selection algorithm, bSCWDTO, compared to other existing algorithms, including the original SC and DTO algorithms. Our experiments were conducted using thirty datasets, and the results indicate that the proposed algorithm consistently outperformed the compared algorithms in terms of fitness value for the selected features. Notably, the bSCWDTO algorithm demonstrated exceptional performance by achieving the lowest fitness value among all tested algorithms. This achievement highlights the algorithm's ability to identify the most optimal subset of features across diverse datasets. Furthermore, the proposed algorithm exhibited the lowest standard deviation compared to the other algorithms, which emphasizes its stability and robustness. The quantitative evaluation metrics further support the superiority of the proposed feature selection algorithm. On average, it achieved an error rate of 0.607174, a select size of 0.615652, and a fitness value of 1.09021. Additionally, it achieved the best fitness value of 0.991305 and the worst fitness value of 1.34082. These results surpass state-of-the-art feature selection meth-

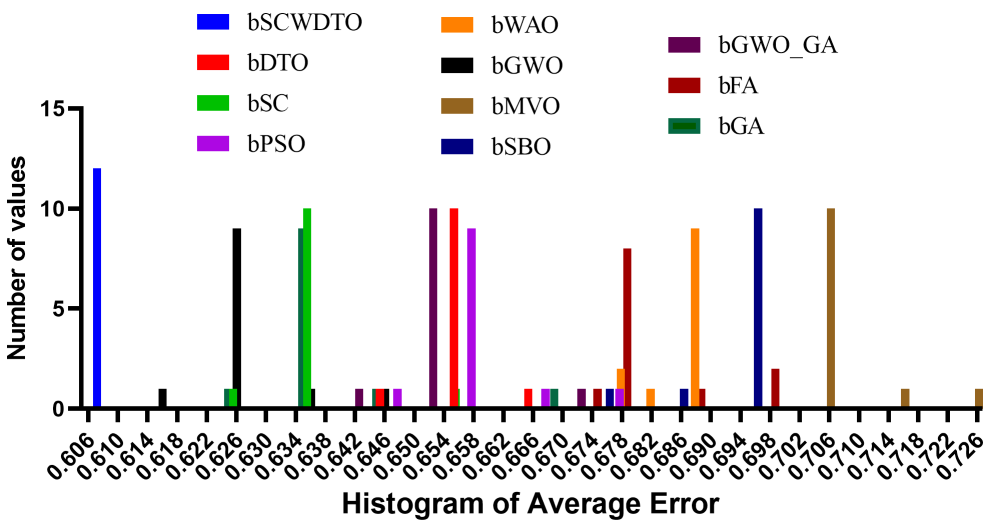


FIGURE 4. Histogram of average error for the results achieved by the optimization algorithms.

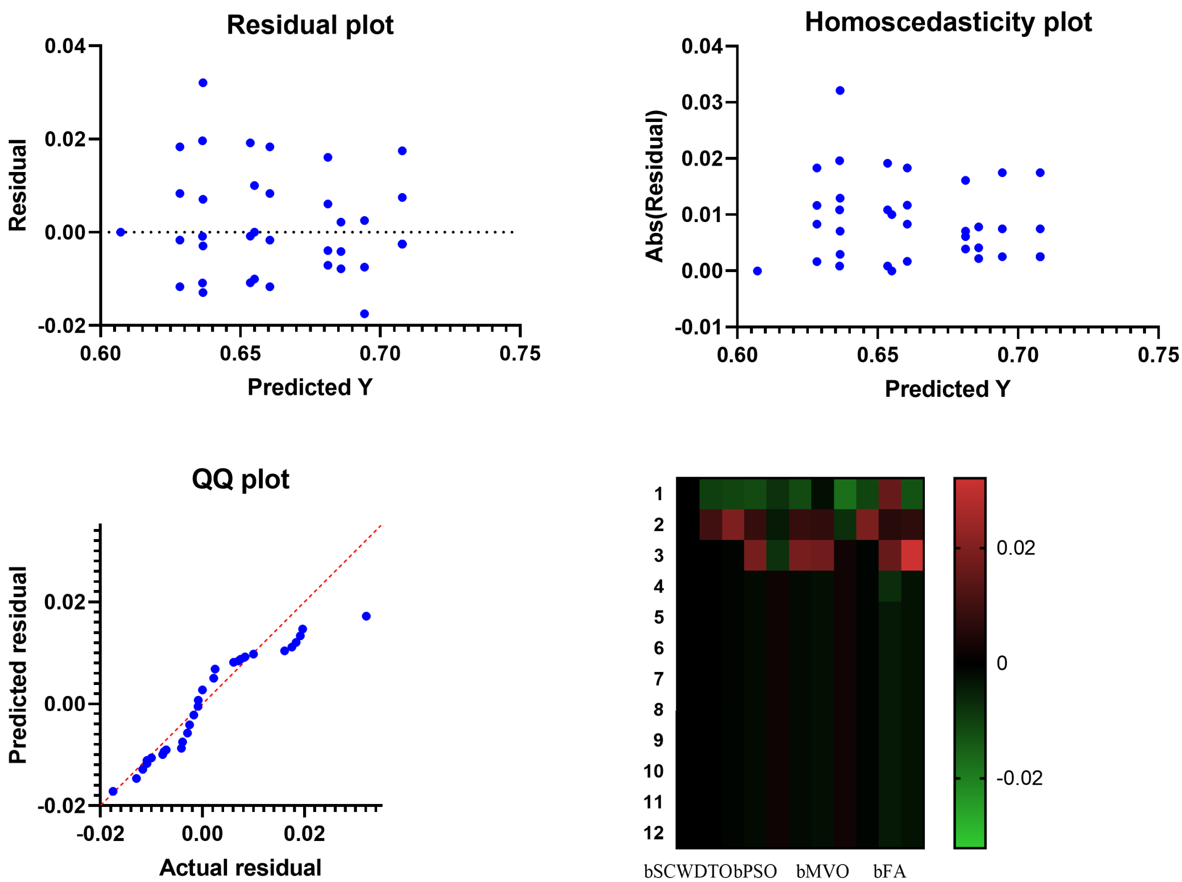
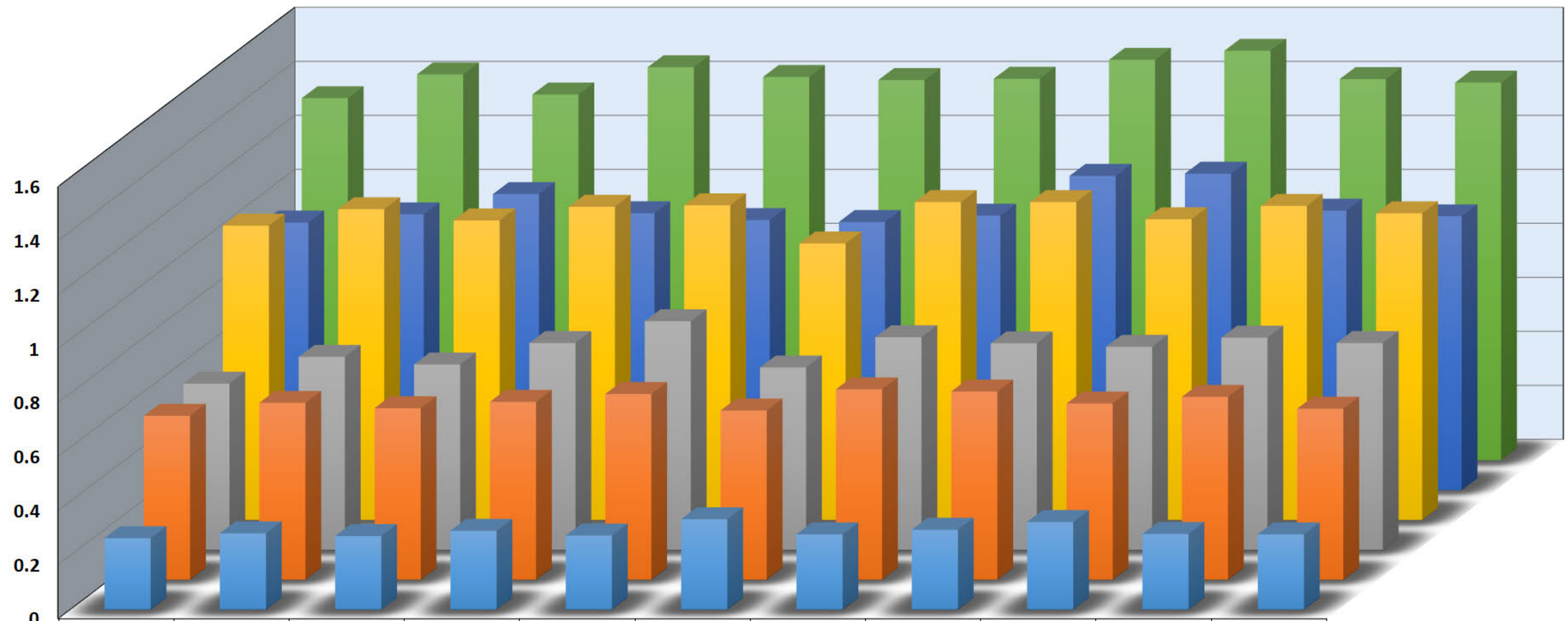


FIGURE 5. The Residual, Homoscedasticity, QQ, and Heatmap plots of the ANOVA test results.

ods when applied to the thirty benchmark datasets. The outcomes of our research confirm that the proposed approach significantly improves the quality of the selected features. Based on these compelling results, we highly recommend using the bSCWDTO algorithm for various feature selection tasks. Its superior performance, stability, and robustness make it a valuable tool in improving classification accuracy and enhancing the overall efficiency of machine learning models.



	bSCWDTO	bDTO	bSC	bPSO	bWAO	bGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Standard deviation	0.264821333	0.282805667	0.27237209	0.291943667	0.274660333	0.334706311	0.278878333	0.296040861	0.324661717	0.280494	0.278983333
Average error	0.607174	0.655076333	0.635609409	0.658878	0.688187667	0.626743153	0.705410333	0.696890729	0.65271566	0.677424333	0.633687
Average select size	0.615652667	0.714738667	0.686824409	0.766381667	0.847022	0.675790196	0.787678	0.764866399	0.752215198	0.786119	0.766398333
Average fitness	1.090218667	1.151195333	1.109514962	1.160226	1.165785	1.024106748	1.177550333	1.177612686	1.114041174	1.163264667	1.135911333
Best fitness	0.991305667	1.021280667	1.09640979	1.025780667	1.000050667	0.993150146	1.016463333	1.16380892	1.172355896	1.034906	1.014096333
Worst fitness	1.3408241	1.429163	1.354691005	1.455555333	1.419765333	1.408121618	1.412886333	1.48294592	1.516495496	1.411731333	1.398706333

FIGURE 6. Average standard deviation fitness, error, select size, mean fitness, best, and worst fitness achieved by the proposed method and other methods over the 30 UCI datasets.

TABLE 15. The measured best fitness, worst fitness, median fitness, mean fitness, and standard deviation fitness resulted from the application of the proposed SCWDTO to CEC2017 in 10k.

	Best	Worst	Median	Mean	Std
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	5.514270671	4.215750798	6.398642505	10.26361419
8	0	12.32450769	13.22356022	13.23418264	13.39175701
9	0	0	0	0	0
10	0.116954116	178.5530942	0.295919969	8.544988267	0
11	0	0	0	0	0
12	0.09749966	210.9195507	0	14.12308169	39.30065309
13	0	9.321175672	10.21412352	11.38184232	0
14	0	0	0	0	0
15	6.14685E-05	0	0	0	0
16	0.098012892	110.2678942	0	0	2.713094934
17	0.954609002	18.29334262	5.904828528	4.161463056	7.468256503
18	0.002058369	8.462249376	0.333393784	0	5.614939499
19	0	0	0	0	0
20	0	7.080770399	4.236071263	-2.278109116	6.056449327
21	98.91012039	181.9922845	178.539683	173.7738568	2.129324934
22	0	80.476085	80.476085	73.35120359	9.938441953
23	97.31951488	281.6734604	276.2583036	276.544148	13.32914834
24	171.54	307.6833607	302.6321954	299.8577142	3.308961838
25	293.233049	407.714545	364.4786519	367.0407102	4.557444475
26	219.19	316.0094203	271.076085	271.9571308	8.531989557
27	180.1222668	359.4048268	356.3867297	356.4001372	14.07429484
28	190.6	568.2422681	338.9184334	364.7679322	95.1929238
29	110.0336364	250.4418952	209.5127665	212.5760269	5.616896406

TABLE 16. Experimental results of GWO, PSO, SFS and SCWDTO over 51 independent runs on 29 test functions of 10 variables with 100,000 FES.

	GWO	PSO	SFS	SCWDTO
1	1.50E+08 ± 6.50E+07	2.30E+03 ± 3.05E+03	5.60E+03 ± 3.26E+03	0.00E+00 ± 0.00E+00
2	5.28E+02 ± 7.98E+02	0.00E+00 ± 0.00E+00	2.05E02 ± 1.17E02	0.00E+00 ± 0.00E+00
3	1.84E+01 ± 1.62E+01	2.82E+00 ± 1.21E+00	9.37E01 ± 6.74E01	0.00E+00 ± 0.00E+00
4	3.01E+01 ± 4.85E+00	1.59E+01 ± 7.08E+00	8.68E+00 ± 3.30E+00	0.00E+00 ± 0.00E+00
5	7.81E+00 ± 1.09E+00	8.27E02 ± 3.36E01	5.35E03 ± 1.59E03	0.00E+00 ± 0.00E+00
6	4.35E+01 ± 5.14E+00	1.72E+01 ± 4.46E+00	2.33E+01 ± 3.60E+00	0.00E+00 ± 0.00E+00
7	2.38E+01 ± 4.44E+00	1.23E+01 ± 5.33E+00	7.42E+00 ± 2.76E+00	3.91E+01 ± 2.98E+00
8	1.10E+01 ± 3.79E+00	0.00E+00 ± 0.00E+00	6.68E06 ± 4.34E06	2.34E+01 ± 2.53E+00
9	8.53E+02 ± 2.50E+02	6.30E+02 ± 2.64E+02	3.63E+02 ± 1.91E+02	0.00E+00 ± 0.00E+00
10	4.03E+01 ± 9.91E+00	1.10E+01 ± 7.27E+00	4.61E+00 ± 1.25E+00	0.00E+00 ± 0.00E+00
11	2.58E+06 ± 3.10E+06	1.33E+04 ± 1.24E+04	5.04E+03 ± 2.11E+03	0.00E+00 ± 0.00E+00
12	1.16E+04 ± 8.11E+03	6.45E+03 ± 5.72E+03	4.55E+01 ± 9.83E+00	7.66E+00 ± 1.81E+00
13	5.91E+02 ± 1.21E+03	4.57E+01 ± 1.93E+01	2.20E+01 ± 3.82E+00	0.00E+00 ± 0.00E+00
14	8.04E+02 ± 1.12E+03	5.62E+01 ± 5.89E+01	1.00E+01 ± 2.14E+00	0.00E+00 ± 0.00E+00
15	8.52E+01 ± 9.41E+01	2.05E+02 ± 1.19E+02	4.17E+00 ± 3.16E+00	3.12E+00 ± 3.65E+00
16	5.42E+01 ± 8.45E+00	4.56E+01 ± 2.30E+01	2.34E+01 ± 5.66E+00	1.11E+01 ± 6.08E+00
17	3.68E+04 ± 2.11E+04	5.10E+03 ± 5.76E+03	5.22E+01 ± 1.05E+01	3.14E01 ± 1.9E-01
18	1.75E+03 ± 3.81E+03	9.57E+01 ± 2.95E+02	5.84E+00 ± 8.94E01	3.38E+00 ± 3.877E+00
19	7.76E+01 ± 3.83E+01	5.51E+01 ± 5.42E+01	1.21E+01 ± 3.31E+00	0.00E+00 ± 0.00E+00
20	2.03E+02 ± 4.93E+01	1.79E+02 ± 5.65E+01	1.00E+02 ± 5.06E02	2.33E+02 ± 5.88E+01
21	1.25E+02 ± 6.03E+00	9.38E+01 ± 2.55E+01	9.24E+01 ± 3.00E+01	1.13E+02 ± 5.27E01
22	3.33E+02 ± 3.86E+00	3.28E+02 ± 1.24E+01	3.03E+02 ± 4.35E+01	4.123E+02 ± 3.62E+00
23	3.63E+02 ± 4.56E+00	3.24E+02 ± 8.33E+01	2.18E+02 ± 1.18E+02	2.14E+02 ± 1.07E+01
24	4.42E+02 ± 1.56E+01	4.25E+02 ± 2.29E+01	4.21E+02 ± 2.30E+01	7.22E+02 ± 2.15E+01
25	4.09E+02 ± 1.47E+02	2.74E+02 ± 7.63E+01	2.92E+02 ± 4.40E+01	3.62E+02 ± 0.11E+00
26	3.96E+02 ± 1.15E+00	4.03E+02 ± 1.97E+01	3.92E+02 ± 1.85E+00	3.72E+02 ± 1.111E01
27	5.39E+02 ± 9.99E+01	4.54E+02 ± 1.57E+02	3.06E+02 ± 3.97E+01	3.88E+02 ± 4.93E+01
28	2.94E+02 ± 3.04E+01	3.05E+02 ± 4.50E+01	2.59E+02 ± 1.18E+01	2.94E+02 ± 3.99E+00
29	4.84E+05 ± 7.31E+05	2.00E+05 ± 3.79E+05	2.03E+03 ± 1.63E+03	4.62E+02 ± 3.17E+01

VI. CONCLUSION

Applying feature selection to the data set before the learning phase is crucial to increase the effectiveness of the classification process. To extract the most relevant features from a dataset, a feature selection method first explores all potential subsets of features. It then picks the best one based on an

TABLE 17. The convergence of time resulting from different values of SCWDTO's parameters.

r1 Values	Time	r2 Values	Time	r3 Values	Time	r4 Values	Time	r5 Values	Time	r6 Values	Time	r7 Values	Time	K1 Values	Time	K2 Values	Time
0.05	7.744	0.05	7.604	0.05	7.498	0.05	7.542	0.05	7.377	0.1	7.471	0.1	7.360	0.1	7.411	0.1	7.420
0.1	7.742	0.1	7.380	0.1	7.742	0.1	7.377	0.1	7.865	0.2	7.440	0.2	7.536	0.2	7.651	0.2	7.746
0.15	7.837	0.15	7.709	0.15	7.616	0.15	7.366	0.15	7.859	0.3	7.342	0.3	7.869	0.3	7.373	0.3	7.450
0.2	7.496	0.2	7.738	0.2	7.597	0.2	7.811	0.2	7.504	0.4	7.376	0.4	7.799	0.4	7.773	0.4	7.847
0.25	7.465	0.25	7.636	0.25	7.685	0.25	7.626	0.25	7.550	0.5	7.466	0.5	7.369	0.5	7.528	0.5	7.851
0.3	7.595	0.3	7.441	0.3	7.349	0.3	7.704	0.3	7.622	0.6	7.765	0.6	7.678	0.6	7.781	0.6	7.597
0.35	7.810	0.35	7.353	0.35	7.866	0.35	7.608	0.35	7.534	0.7	7.837	0.7	7.400	0.7	7.717	0.7	7.388
0.4	7.614	0.4	7.835	0.4	7.373	0.4	7.472	0.4	7.689	0.8	7.366	0.8	7.747	0.8	7.683	0.8	7.814
0.45	7.891	0.45	7.431	0.45	7.651	0.45	7.887	0.45	7.853	0.9	7.816	0.9	7.535	0.9	7.761	0.9	7.513
0.5	7.656	0.5	7.806	0.5	7.364	0.5	7.439	0.5	7.660	1	7.404	1	7.567	1	7.433	1	7.811
0.55	7.381	0.55	7.533	0.55	7.836	0.55	7.897	0.55	7.412	1.1	7.396	1.1	7.720	1.1	7.774	1.1	7.669
0.6	7.388	0.6	7.491	0.6	7.650	0.6	7.641	0.6	7.548	1.2	7.380	1.2	7.676	1.2	7.377	1.2	7.784
0.65	7.738	0.65	7.485	0.65	7.369	0.65	7.634	0.65	7.594	1.3	7.544	1.3	7.424	1.3	7.423	1.3	7.756
0.7	7.701	0.7	7.464	0.7	7.457	0.7	7.888	0.7	7.857	1.4	7.677	1.4	7.508	1.4	7.584	1.4	7.722
0.75	7.363	0.75	7.459	0.75	7.343	0.75	7.763	0.75	7.453	1.5	7.765	1.5	7.513	1.5	7.832	1.5	7.796
0.8	7.737	0.8	7.407	0.8	7.703	0.8	7.778	0.8	7.817	1.6	7.892	1.6	7.697	1.6	7.408	1.6	7.871
0.85	7.457	0.85	7.404	0.85	7.368	0.85	7.593	0.85	7.378	1.7	7.828	1.7	7.463	1.7	7.630	1.7	7.638
0.9	7.506	0.9	7.553	0.9	7.405	0.9	7.713	0.9	7.864	1.8	7.579	1.8	7.688	1.8	7.361	1.8	7.695
0.95	7.469	0.95	7.840	0.95	7.574	0.95	7.484	0.95	7.730	1.9	7.746	1.9	7.792	1.9	7.881	1.9	7.386
1	7.579	1	7.766	1	7.475	1	7.709	1	7.792	2	7.410	2	7.561	2	7.639	2	7.869

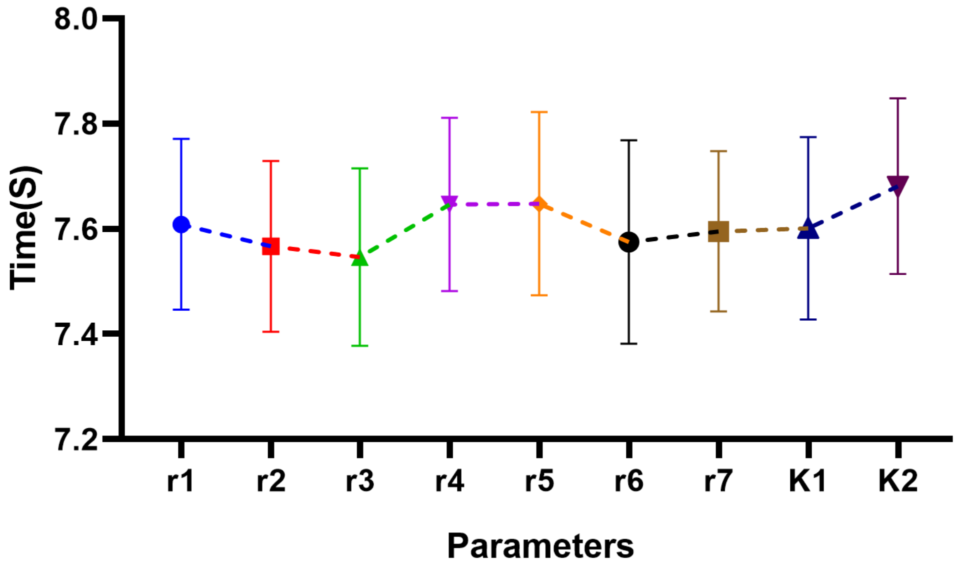


FIGURE 7. The sensitivity analysis of the convergence time.

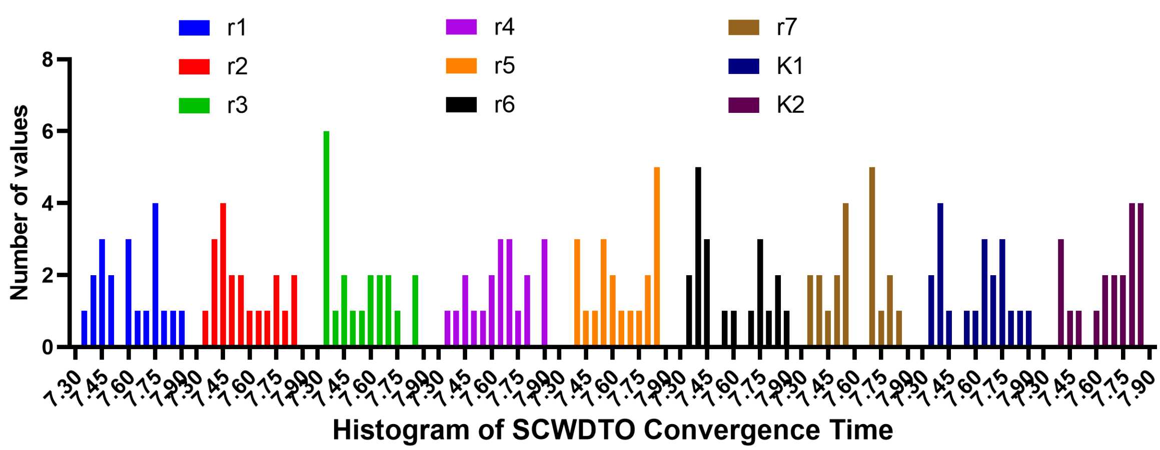


FIGURE 8. Histogram of the convergence time.

evaluation metric. To pick the best subset of features for various problems, this work proposed a hybrid approach called bSCWDTO, a binary DTO algorithm based on the SC algorithm and used in conjunction with the KNN classifier. In the proposed algorithm, the original DTO is used to expand the search space and the SC to expand the diversity of the

TABLE 18. The convergence of fitness resulting from different values of SCWDTO's parameters.

r1		r2		r3		r4		r5		r6		r7		K1		K2	
Values	Fitness	Values	Fitness	Values	Fitness	Values	Fitness	Values	Fitness	Values	Fitness	Values	Fitness	Values	Fitness	Values	Fitness
0.05	74.503	0.05	73.616	0.05	74.207	0.05	73.885	0.05	74.036	0.1	73.468	0.1	74.101	0.1	73.830	0.1	74.338
0.1	74.415	0.1	74.546	0.1	73.319	0.1	73.838	0.1	73.942	0.2	74.213	0.2	73.830	0.2	74.217	0.2	73.171
0.15	73.457	0.15	73.200	0.15	74.029	0.15	74.475	0.15	74.424	0.3	74.546	0.3	74.165	0.3	73.579	0.3	73.364
0.2	73.683	0.2	74.544	0.2	73.253	0.2	73.792	0.2	74.517	0.4	73.519	0.4	74.551	0.4	73.527	0.4	74.186
0.25	73.303	0.25	74.568	0.25	74.147	0.25	74.029	0.25	74.135	0.5	73.319	0.5	73.886	0.5	74.600	0.5	73.193
0.3	73.476	0.3	74.193	0.3	74.450	0.3	73.647	0.3	74.522	0.6	73.219	0.6	73.972	0.6	74.069	0.6	73.884
0.35	73.879	0.35	74.542	0.35	74.056	0.35	73.641	0.35	73.698	0.7	73.935	0.7	73.845	0.7	74.135	0.7	73.544
0.4	73.295	0.4	73.889	0.4	74.059	0.4	74.250	0.4	74.072	0.8	74.193	0.8	73.971	0.8	73.651	0.8	74.308
0.45	73.545	0.45	74.504	0.45	73.683	0.45	73.604	0.45	73.391	0.9	74.041	0.9	73.795	0.9	74.286	0.9	73.164
0.5	73.629	0.5	73.262	0.5	73.203	0.5	74.176	0.5	74.200	1	73.380	1	74.042	1	74.336	1	73.773
0.55	74.482	0.55	74.521	0.55	74.424	0.55	73.953	0.55	73.143	1.1	73.184	1.1	74.066	1.1	74.193	1.1	73.340
0.6	73.397	0.6	74.042	0.6	73.680	0.6	74.273	0.6	74.020	1.2	73.431	1.2	74.287	1.2	74.358	1.2	73.444
0.65	73.778	0.65	73.470	0.65	74.216	0.65	74.073	0.65	74.037	1.3	74.134	1.3	73.946	1.3	73.440	1.3	74.298
0.7	73.418	0.7	74.181	0.7	73.944	0.7	74.120	0.7	74.138	1.4	73.229	1.4	73.395	1.4	73.843	1.4	73.748
0.75	73.971	0.75	73.794	0.75	73.859	0.75	74.159	0.75	73.536	1.5	73.496	1.5	73.802	1.5	73.572	1.5	73.561
0.8	74.366	0.8	73.840	0.8	73.942	0.8	74.003	0.8	74.065	1.6	74.505	1.6	74.105	1.6	73.414	1.6	74.193
0.85	74.382	0.85	73.277	0.85	74.128	0.85	73.609	0.85	73.262	1.7	74.406	1.7	73.264	1.7	73.523	1.7	74.093
0.9	73.434	0.9	74.454	0.9	74.358	0.9	73.816	0.9	73.411	1.8	74.402	1.8	73.879	1.8	74.243	1.8	73.945
0.95	73.167	0.95	74.356	0.95	74.322	0.95	73.876	0.95	73.893	1.9	73.420	1.9	74.359	1.9	73.581	1.9	74.041
1	73.228	1	74.022	1	73.598	1	73.319	1	74.496	2	74.123	2	73.433	2	73.825	2	73.891

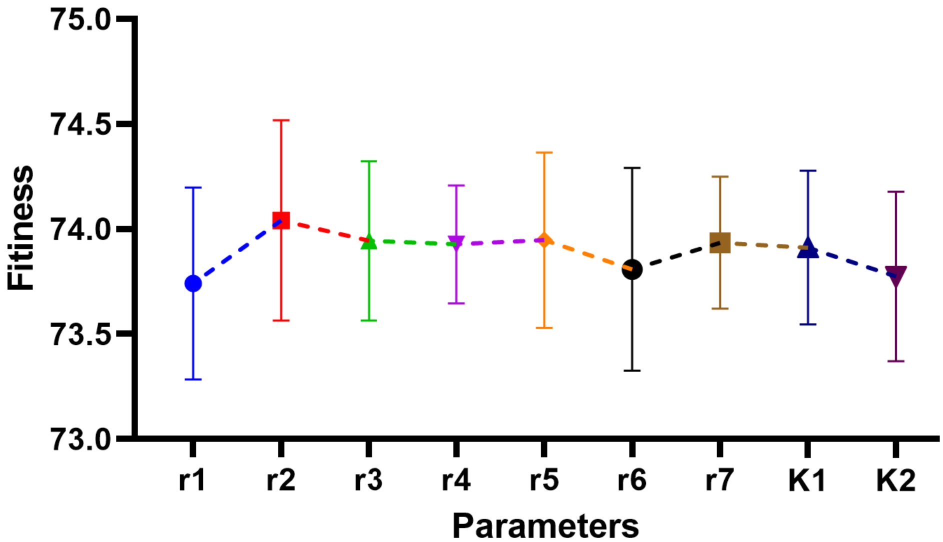


FIGURE 9. The sensitivity analysis of the convergence fitness.

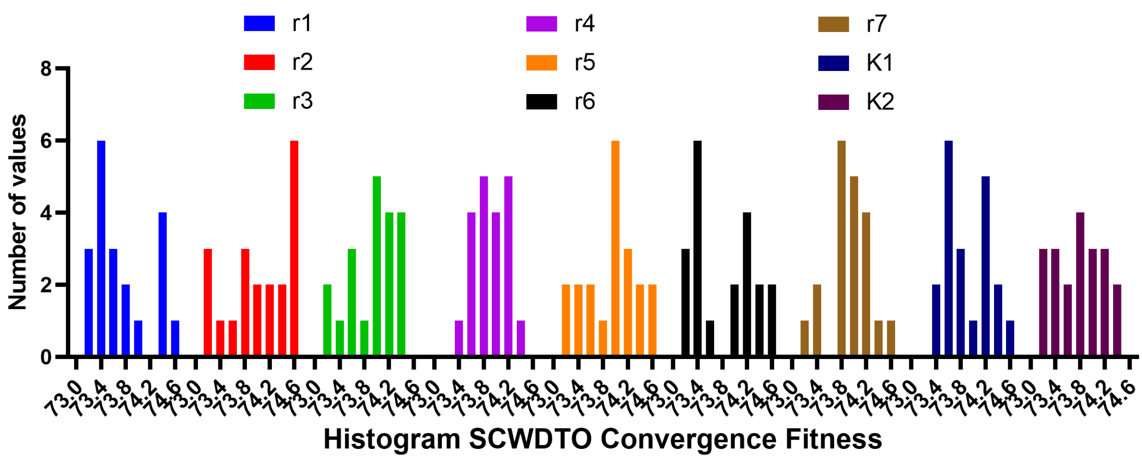


FIGURE 10. Histogram of the convergence fitness.

population. The continuous values were discretized using the sigmoid function to apply the proposed algorithm to the feature selection problem. Experiments were conducted on thirty UCI machine learning repository datasets to examine the algorithm's stability and robustness. The achieved results were compared with those obtained using the bDTO,

TABLE 19. Statistical analysis applied to the results of the sensitivity analysis of the convergence time.

	r1	r2	r3	r4	r5	r6	r7	K1	K2
Number of values	20	20	20	20	20	20	20	20	20
Minimum	7.363	7.353	7.343	7.366	7.377	7.342	7.36	7.361	7.386
25% Percentile	7.466	7.434	7.37	7.499	7.512	7.398	7.474	7.414	7.534
Median	7.605	7.512	7.536	7.638	7.641	7.508	7.564	7.635	7.734
75% Percentile	7.741	7.731	7.677	7.774	7.844	7.765	7.714	7.77	7.813
Maximum	7.891	7.84	7.866	7.897	7.865	7.892	7.869	7.881	7.871
Range	0.528	0.487	0.523	0.531	0.488	0.55	0.509	0.52	0.485
Mean	7.608	7.567	7.546	7.647	7.648	7.575	7.595	7.601	7.681
Std. Deviation	0.1624	0.1626	0.1686	0.1647	0.1745	0.1932	0.1524	0.1736	0.167
Std. Error of Mean	0.0363	0.03635	0.03771	0.03683	0.03903	0.0432	0.03407	0.03883	0.03733
Sum	152.2	151.3	150.9	152.9	153	151.5	151.9	152	153.6

TABLE 20. ANOVA test applied to the results of the sensitivity analysis of the convergence time.

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	0.3055	8	0.03819	F (8, 171) = 1.335	P=0.02290
Residual (within columns)	4.89	171	0.0286		
Total	5.196	179			

TABLE 21. Wilcoxon signed-rank test applied to the results of the sensitivity analysis of the convergence time.

	r1	r2	r3	r4	r5	r6	r7	K1	K2
Theoretical mean	0	0	0	0	0	0	0	0	0
Actual mean	7.608	7.567	7.546	7.647	7.648	7.575	7.595	7.601	7.681
Number of values	20	20	20	20	20	20	20	20	20
One sample t test									
t, df	t=209.6, df=19	t=208.2, df=19	t=200.1, df=19	t=207.6, df=19	t=196.0, df=19	t=175.4, df=19	t=222.9, df=19	t=195.8, df=19	t=205.7, df=19
P value (two tailed)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
P value summary	****	****	****	****	****	****	****	****	****
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
How big is the discrepancy?									
Discrepancy	7.608	7.567	7.546	7.647	7.648	7.575	7.595	7.601	7.681
SD of discrepancy	0.1624	0.1626	0.1686	0.1647	0.1745	0.1932	0.1524	0.1736	0.167
SEM of discrepancy	0.0363	0.03635	0.03771	0.03683	0.03903	0.0432	0.03407	0.03883	0.03733
95% confidence interval	7.532 to 7.684	7.491 to 7.643	7.467 to 7.625	7.570 to 7.724	7.566 to 7.730	7.485 to 7.665	7.524 to 7.666	7.520 to 7.682	7.603 to 7.759
R squared (partial eta squared)	0.9996	0.9996	0.9995	0.9996	0.9995	0.9994	0.9996	0.9995	0.9996

TABLE 22. Statistical analysis applied to the results of the sensitivity analysis of the convergence fitness.

	r1	r2	r3	r4	r5	r6	r7	K1	K2
Number of values	20	20	20	20	20	20	20	20	20
Minimum	73.17	73.2	73.2	73.32	73.14	73.18	73.26	73.41	73.16
25% Percentile	73.4	73.66	73.68	73.68	73.58	73.39	73.81	73.57	73.38
Median	73.59	74.11	74.04	73.92	74.04	73.73	73.96	73.84	73.83
75% Percentile	74.27	74.52	74.21	74.15	74.18	74.21	74.1	74.24	74.16
Maximum	74.5	74.57	74.45	74.48	74.52	74.55	74.55	74.6	74.34
Range	1.336	1.368	1.247	1.156	1.379	1.362	1.287	1.186	1.174
Mean	73.74	74.04	73.94	73.93	73.95	73.81	73.93	73.91	73.77
Std. Deviation	0.4568	0.477	0.3793	0.2804	0.4178	0.4819	0.3145	0.3657	0.4032
Std. Error of Mean	0.1021	0.1067	0.08482	0.0627	0.09343	0.1078	0.07033	0.08178	0.09016
Sum	1475	1481	1479	1479	1479	1476	1479	1478	1475

TABLE 23. ANOVA test applied to the sensitivity analysis results of the convergence fitness.

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	1.505	8	0.1882	F (8, 171) = 1.159	P=0.03265
Residual (within columns)	27.75	171	0.1623		
Total	29.26	179			

bSC, bPSO, bWOA, bGWO, bMVO, bSBO, bGA, and bFA optimization algorithms. As proved by the outcomes, the proposed bSCWDTO algorithm is superior. The proposed approach is planned to be evaluated in further work on continuous and engineering problems with constraints. In addition, the proposed approach will be evaluated in terms of the CEC2019 problems to give additional evidence of its robustness, superiority, and generalization.

APPENDIX

In this appendix, an additional investigation of the effectiveness of the proposed optimization algorithm is performed to prove its efficiency in continuous optimization problems. This appendix consists of two scenarios, the first is the optimization of CEC2017 functions using the continuous version of the proposed optimization algorithm. Whereas the second scenario is the sensitivity analysis of the proposed algorithm.

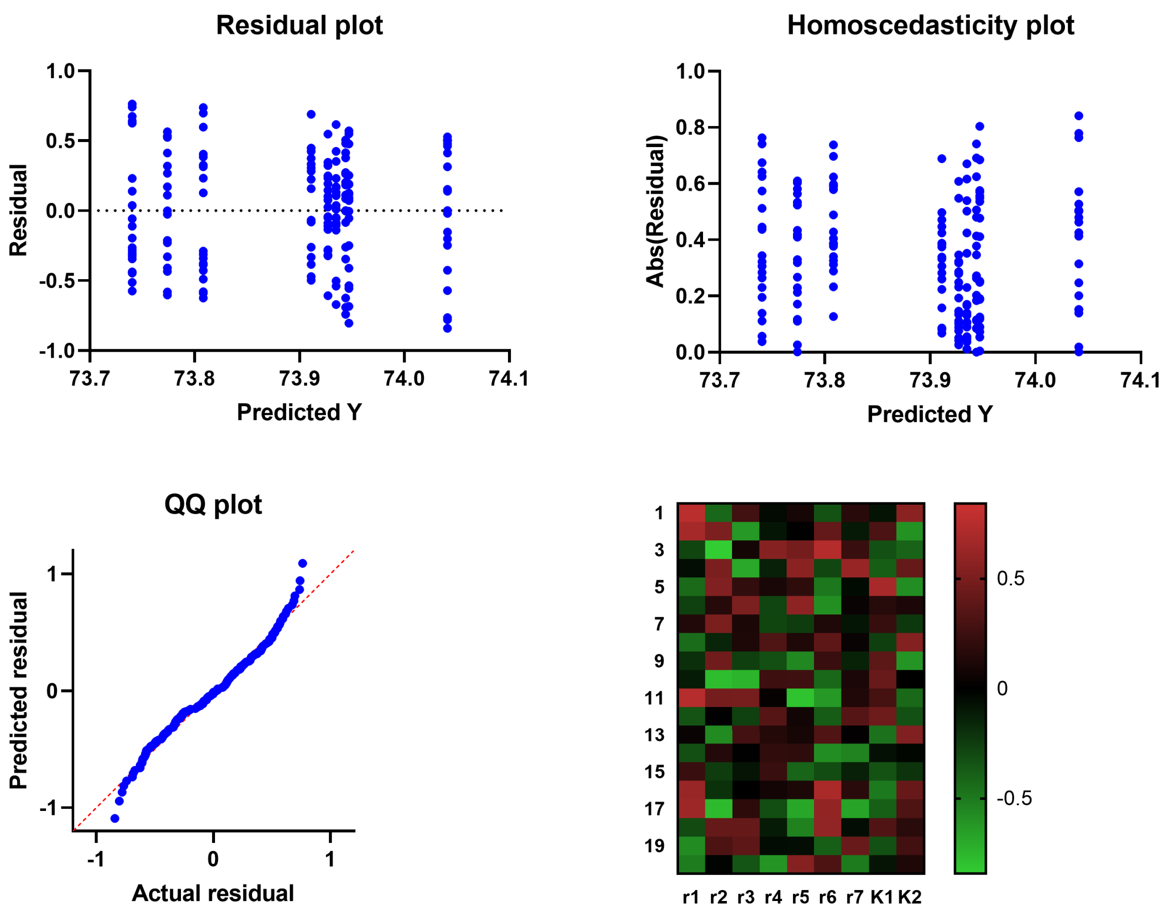


FIGURE 11. The residual, Homoscedasticity, QQ, and heatmap of the ANOVA test applied to the fitness sensitivity.

TABLE 24. Wilcoxon signed-rank test applied to the sensitivity analysis of the convergence fitness results.

	r1	r2	r3	r4	r5	r6	r7	K1	K2
Theoretical mean	0	0	0	0	0	0	0	0	0
Actual mean	73.74	74.04	73.94	73.93	73.95	73.81	73.93	73.91	73.77
Number of values	20	20	20	20	20	20	20	20	20
One sample t test									
t, df	t=722.0, df=19	t=694.1, df=19	t=871.8, df=19	t=1179, df=19	t=791.5, df=19	t=685.0, df=19	t=1051, df=19	t=903.8, df=19	t=818.3, df=19
P value (two tailed)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
P value summary	****	****	****	****	****	****	****	****	****
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
How big is the discrepancy?									
Discrepancy	73.74	74.04	73.94	73.93	73.95	73.81	73.93	73.91	73.77
SD of discrepancy	0.4568	0.477	0.3793	0.2804	0.4178	0.4819	0.3145	0.3657	0.4032
SEM of discrepancy	0.1021	0.1067	0.08482	0.0627	0.09343	0.1078	0.07033	0.08178	0.09016
95% confidence interval	73.53 to 73.95	73.82 to 74.26	73.77 to 74.12	73.80 to 74.06	73.75 to 74.14	73.58 to 74.03	73.79 to 74.08	73.74 to 74.08	73.59 to 73.96
R squared (partial eta squared)	1	1	1	1	1	1	1	1	1

A. CEC2017

For the CEC2017 benchmark problems, the proposed SCWDTO algorithm has proven very useful. The GWO, PSO, and SFS algorithms are three existing methods that this one outperforms. The SCWDTO algorithm has improved exploration and exploitation capabilities using the sine-cosine weighting method and the dipper-throated mechanism. The algorithm dynamically balances exploration and exploitation based on the sine and cosine functions for improved convergence to optimal solutions. However, GWO, PSO, and SFS algorithms often use inefficient methods of finding optimal solutions, such as predetermined weight values or random search strategies. Second, the SCWDTO algorithm excels at solving complex optimization problems. Because of its dipper-throated mechanism and sine-cosine weighting, it can efficiently go across high-dimensional solution spaces, avoiding the curse of dimensionality. In contrast, scalability is a common problem for GWO, PSO, and SFS algorithms, making it hard to explore and exploit solutions in large-scale problems accurately. Third, the SCWDTO algorithm has a robust global exploration potential because of the special combination of the sine-cosine weighting and the dipper-throated processes. Since GWO, PSO, and SFS can become stuck in suboptimal portions of the search space, this one has a leg up on them when breaking free of local optima. In addressing the CEC2017 benchmark problems, the proposed SCWDTO algorithm is superior to the GWO, PSO, and SFS algorithms. Improved exploration and exploitation, effective management of high-dimensional problems, and reliable global exploration are all made possible by the system's novel combination of sine-cosine weighting and dipper-throated methods. These features make the SCWDTO algorithm a potentially helpful tool for solving challenging optimization problems. Table 15 presents the results of the best, worst, median, mean, and standard deviation fitness achieved by the proposed SCWDTO algorithm for the 29 CEC2017 benchmark functions. In addition, the results applying the proposed optimization algorithm and three other competing algorithms, GWO, PSO, and SFS, are presented in Table 16. These results confirm the superiority of the proposed optimization algorithm in solving the CEC2017 benchmark functions.

B. SENSITIVITY ANALYSIS

A sensitivity analysis is conducted to understand further how the proposed SCWDTO method performs and behaves with different values of its parameters. The method's performance can be fine-tuned and optimized for certain problem domains by studying its sensitivity to parameter alterations. The population size is an important input to the SCWDTO algorithm. By conducting a sensitivity analysis, we may learn how changing the population size impacts the algorithm's convergence time and the quality of its solution. It aids in determining the optimum population size that allows for both exploration and exploitation. There is some evidence that larger populations improve exploration at the expense of computing cost, while smaller populations are more prone to premature convergence and less-than-optimal solutions. The maximum number of iterations or generations is another key factor to examine. The convergence behavior of the algorithm and the best stopping criterion can be evaluated by adjusting this value. By performing a sensitivity analysis, one may determine when further iterations will no longer significantly increase the quality of the answer and so save computational time. The sensitivity analysis can also examine how changing the crossover rate or weighting factor affects the algorithm's performance. Sensitivity analysis can be used to determine the best settings for these parameters, which affect the algorithm's exploration-exploitation trade-off. Researchers can learn about the resilience and adaptability of SCWDTO in various problem domains by assessing how the algorithm responds to varying values of its parameters. The effect of the dipper-throated mechanism on the algorithm's efficacy can also be investigated through sensitivity analysis. The dipper-throated parameter controls the depth to which a creature will venture into unknown territory. Understanding the algorithm's sensitivity to changes in exploration and exploitation might help fine-tune that balance for individual optimization tasks. Sensitivity analysis is a must to learn how the SCWDTO algorithm acts and find the sweet spot for its settings. Understanding how the algorithm's parameters affect convergence, solution quality, and the exploration-exploitation trade-off is aided by this tool. Systematic sensitivity analysis allows for further refinement and customization of the SCWDTO algorithm, making it applicable to various practical optimization problems.

1) SENSITIVITY ANALYSIS OF THE CONVERGENCE TIME

The proposed SCWDTO algorithm's convergence time is an essential metric for sensitivity analysis. Table 17 presents the resulting convergence of time for the changes of each parameter used in the proposed SCWDTO algorithm. In addition, Figure 7 and Figure 8 show the sensitivity and histogram of the convergence time. The time it takes for an algorithm to converge to a reasonable solution or meet another stopping requirement is meant here. The algorithm's efficiency and efficacy in addressing optimization issues can be learned by examining the convergence time. During a sensitivity analysis, researchers can examine the time required for convergence by changing factors like population size, the maximum number of iterations, and the control parameters. Researchers can evaluate the algorithm's sensitivity to these parameters and determine which combinations result in faster convergence without compromising solution quality by monitoring the convergence time under different parameter settings. If an algorithm can swiftly converge to optimal or near-optimal solutions, that's a good thing. Convergence time and solution quality are two factors that must be balanced. An algorithm may converge quickly, yet its results may be less than ideal. In this way, sensitivity analysis speeds up the algorithm's convergence and guarantees its continued precision when solving the problem. Furthermore, the convergence time can be investigated for numerous problem domains and benchmark functions via sensitivity analysis. The algorithm's robustness is evaluated across various problem dimensions, sizes, and types. It's helpful since it shows how the algorithm handles different kinds of optimization issues and how well it scales. In addition, sensitivity analysis can show if there are any compromises between convergence speed and computing power. Some choices of parameters may result in faster convergence, albeit at the expense of more CPU time. Using sensitivity analysis, scientists may zero in on the parameter settings that strike the best balance between fast convergence and low computing cost. In general, the sensitivity analysis of the SCWDTO method relies heavily on the convergence time. Optimizing the algorithm's performance, balancing convergence time and solution quality, and ensuring optimal resource use can be achieved by studying the algorithm's convergence behavior under varied parameter settings and problem domains.

The statistical analysis results help to understand how long it will take for the proposed SCWDTO algorithm to converge as presented in Table 19. Descriptive statistics for a sample size of 20 values are provided for the parameters r1, r2, r3, r4, r5, r6, r7, K1, and K2. The fastest observed convergence time for any given parameter ranges from 7.363 to 7.386, where 7.363 is the minimum value. To better understand how convergence times are distributed, we can look at the

25th, median, and 75th percentiles, with the median being the typical number. Variability in convergence times is represented by a range from 0.485 to 0.531, where 0.485 is the minimum, and 0.531 is the maximum. All parameters have a mean convergence time between 7.546 and 7.681, with r1's averaging 7.608 seconds, r2's averaging 7.567 seconds, and so on. Convergence times show reasonably constant behavior across parameters, as measured by the standard deviation, which ranges from 0.1624 to 0.1932. Indicating how confident one can be in the calculated means, the standard error of the mean approximates the sample mean's precision. The total observed convergence time ranges from 150.9 to 153.6, calculated by adding the convergence times for each of the sample's parameters. The statistical evaluation provides a wide-ranging survey of the features affecting SCWDTO's convergence time. Convergence times, along with their range, mean, variability, and consistency, are highlighted so researchers can evaluate the algorithm's effectiveness and make educated decisions about parameter adjustment and optimization.

The ANOVA table details the treatment-specific (between columns), residual-specific (within columns), and total-specific variance components and degrees of freedom, as presented in Table 20. The treatment SS captures the variance that can be traced to the various parameter settings. At the same time, the residual SS accounts for the variation that cannot be assigned to any one setting. The sum of the SS quantifies the complete range of values during the convergence time. By comparing the treatment MS to the residual MS, we can obtain the F-statistic. In this situation, the F value is 1.335 (DFn, DFd), where DFn is the number of degrees of freedom associated with the treatment, and DFd is the number of degrees of freedom associated with the residual. The corresponding p-value of 0.02290 represents the statistical likelihood of observing this outcome due to chance alone. The ANOVA test indicates significant differences in convergence time between the various parameter configurations because the p-value ($P = 0.02290$) is less than the predetermined significance level. To isolate the specific pairwise changes between the configurations that contribute to these substantial variations, however, more post hoc testing would be required. Using the ANOVA test, Researchers can infer the effect of various parameter settings on the SCWDTO algorithm's temporal convergence. There appears to be a statistically significant relationship between the parameter configurations and the convergence time, suggesting that this topic deserves more attention.

Time convergence from a sensitivity study of the proposed SCWDTO method was subjected to the Wilcoxon test, also known as the Wilcoxon signed-rank test, as presented in Table 21. The test yielded the following findings: The SCWDTO algorithm's r1, r2, r3, r4, r5, r6, r7, K1, and K2 were put to the test. All parameters had their theoretical means set to zero, and their respective average convergence times were then determined. Twenty values were explored for every parameter. The t-values and degrees of freedom (df) from one-sample t-tests on each parameter are provided. All parameters have two-tailed p-values smaller than 0.0001, as computed. According to the summary of p-values, there were statistically significant deviations from the theoretical mean in the convergence times for all parameters. The variance between the observed and expected convergence times was calculated for each parameter. The discrepancies were analyzed by computing their standard deviation (SD) and standard error of the mean (SEM), which shed light on their variability and accuracy. A range estimate of the real disparity was produced for each parameter using the 95% confidence intervals. The R-squared (partial eta squared) values were also determined, showing what percentage of the overall convergence time variation may be attributed to parameter variations. R-squared values near 1 indicated high degrees of correlation between the parameters and the convergence time. According to the Wilcoxon test, the SCWDTO algorithm's convergence time is considerably impacted by all of its parameters. The mean's standard deviations and standard errors were minimal for the estimations supplied by the differences between the observed and theoretical mean convergence periods. High R-squared values indicate strong correlations between parameter differences and the observed convergence time. These results highlight the significance of proper parameter selection in achieving the fastest possible SCWDTO convergence time.

2) SENSITIVITY ANALYSIS OF THE CONVERGENCE FITNESS

Results from the proposed SCWDTO optimization method were subjected to a sensitivity analysis of the convergence fitness for different parameters r1, r2, r3, r4, r5, r6, r7, K1, and K2 values. Table 18 presents the resulting convergence of fitness for the changes of each parameter used in the proposed SCWDTO algorithm. In addition, Figure 9 and Figure 10 show the sensitivity and histogram of the convergence fitness. The fitness values for each parameter combination were recorded and evaluated to learn how they affected the algorithm's ability to converge. The fitness values associated with the optimization algorithm's parameter settings are listed in the table below. Parameter values and their corresponding fitness ratings are shown in the columns, while rows represent possible parameter combinations. Several conclusions can be drawn from this table. First, it is clear that the fitness values associated with various parameter settings differ. A fitness value of 74.503 is achieved with $r1 = 0.05$, whereas a value of 73.629 is attained with $r1 = 0.5$. This suggests that the algorithm's fitness for convergence can be influenced by adjusting r1. Other factors, such as r2, r3, r4, r5, r6, r7, K1, and K2, also show fitness value changes. Changing the parameters causes a noticeable shift in the fitness values, indicating that they also affect the convergence fitness. Additional investigation of the parameter combinations is needed for a thorough sensitivity analysis. It is feasible to determine whether values of the parameters lead to more or lesser convergence fitness by comparing the fitness values across different parameter settings. This study is useful for learning how the algorithm responds to changes in its parameters and for determining which values to use for those parameters to achieve the desired level of convergence. In conclusion, the impact of different parameters on the performance of the SCWDTO optimization algorithm is shown by a sensitivity analysis of the convergence fitness based on the results obtained. Insights into the algorithm's sensitivity to different parameter values can be gleaned by comparing the fitness values corresponding to different parameter combinations. We can enhance its convergence properties with this information by fine-tuning the method and choosing suitable parameter settings.

The statistical tests conducted on the proposed SCWDTO algorithm yield useful information about its convergence fitness as presented in Table 22. Each of the parameters r1, r2, r3, r4, r5, r6, r7, K1, and K2 are analyzed with a sample size of 20 values. The results provide the minimum, maximum, average, and consistency of fitness levels, giving researchers a basis for judging the algorithm's efficacy and making decisions about fitness evaluation and optimization. The lowest observed fitness levels attained during convergence have minimum values between 73.14 and 73.41, respectively. As a central tendency measure, the median is a value halfway between the 25th and 75th percentiles, depicting the distribution of fitness values. The range, defined as the difference between the least and maximum values, spans from 1.156 to 1.379, illustrating the spectrum of fitness levels. The average fitness for each parameter ranges from 73.74 to 74.04, with r1 averaging 73.74, r2 averaging 74.04, and so on. Variability in fitness levels around the mean is measured by the standard deviation, which varies from 0.2804 to 0.4819, suggesting a wide range of dispersion. Indicating how confident one can be in the calculated means, the standard error of the mean approximates the sample mean's precision. The overall fitness achieved in the sample was between 1475 and 1481, and it was calculated by adding the fitness values for each parameter. Overall, the statistical analysis tells us much about the SCWDTO algorithm's fitness for convergence.

The ANOVA table details the treatment-specific (between columns), residual-specific (within columns), and total-specific variance components and degrees of freedom, as presented in Table 23 and as shown in the plots of Figure 11. The treatment SS captures the variance that can be traced to the various parameter settings. At the same time, the residual SS accounts for the variation that cannot be assigned to any one setting. The sum of SS can be considered a measure of the fitness convergence's overall variability. By comparing the treatment MS to the residual MS, we can obtain the F-statistic. The degrees of freedom associated with the treatment (DFn) and the residual (DFd) add up to a total of 1.159 for the F statistic. The corresponding p-value of 0.03265 represents the statistical likelihood of observing this outcome due to chance alone. The ANOVA p-value ($P = 0.03265$) is below the set statistical significance threshold, suggesting that different parameter settings produce significantly different fitness convergent distributions. This indicates that the parameter settings significantly affect the SCWDTO algorithm's convergence. However, post hoc testing or additional research is required to identify the specific pairwise differences across the configurations that account for these substantial changes. Researchers can learn how different sets of parameters affect the SCWDTO algorithm's fitness convergence with the help of the ANOVA test. These results show that careful choice of parameter values is necessary to achieve optimal method convergence. Researchers can use these results to fine-tune the algorithm and select appropriate parameters for the desired fitness convergence.

The fitness convergence from the sensitivity analysis of the proposed SCWDTO algorithm was examined using the Wilcoxon test, especially the Wilcoxon signed-rank test, as presented in Table 24. Here are what we found in our tests: The SCWDTO algorithm's parameters K1, K2, K3, K4, K5, K6, and r7 were put to the test. We first fixed the theoretical mean fitness convergence to zero for all parameters to compute the true values for mean convergence. Twenty values were explored for every parameter. T-values and degrees of freedom (df) were calculated using one-sample t-tests on each parameter. All parameters have two-tailed p-values below 0.0001, according to our calculations. According to the summary p-values, there are statistically significant deviations from the predicted mean in fitness convergence for all parameters. The deviations from the theoretical mean fitness convergence were calculated for each variable. To learn more about the variability and accuracy of the discrepancies, we computed their standard deviation (SD) and standard error of the mean (SEM). In addition, 95% confidence intervals were determined to approximate the likely range in which the disparities fall. R-squared values, also called partial eta-squared values, were calculated and determined to equal 1. This means that every variation in the data may be attributed to differences in fitness convergence. The Wilcoxon test shows that the fitness convergence of all SCWDTO algorithm parameters varies significantly from the theoretical mean. Consistently small standard deviations and standard errors of the mean characterize the gaps between observed and predicted mean values of convergence. The high values of R2 indicate a robust correlation between the parameter deviations and the measurable fitness convergence. These results highlight the significance of picking suitable parameter values to maximize the SCWDTO algorithm's fitness convergence. Using the Wilcoxon test, we can see that the parameters we choose significantly affect the algorithm's fitness convergence. These findings can help researchers identify optimal parameter settings for the algorithm during fine-tuning.

ACKNOWLEDGMENT

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R 308), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

REFERENCES

[1] O. O. Akinola, A. E. Ezugwu, J. O. Agushaka, R. A. Zitar, and L. Abualigah, "Multiclass feature selection with metaheuristic optimization algorithms: A review," *Neural Comput. Appl.*, vol. 34, no. 22, pp. 19751–19790, Nov. 2022.

[2] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egyptian Informat. J.*, vol. 19, no. 3, pp. 179–189, Nov. 2018.

[3] A. A. Ewees, M. A. Gaheen, Z. M. Yaseen, and R. M. Ghoniem, "Grasshopper optimization algorithm with crossover operators for feature selection and solving engineering problems," *IEEE Access*, vol. 10, pp. 23304–23320, 2022.

[4] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*. Norwell, MA, USA: Kluwer Academic, 1998.

[5] M. Macedo, H. Siqueira, E. Figueiredo, C. Santana, R. C. Lira, A. Gokhale, and C. Bastos-Filho, "Overview on binary optimization using swarm-inspired algorithms," *IEEE Access*, vol. 9, pp. 149814–149858, 2021.

[6] O. Nave, "Modification of semi-analytical method applied system of ODE," *Mod. Appl. Sci.*, vol. 14, no. 6, p. 75, May 2020.

[7] D. S. Khafaga, E.-S. M. El-Kenawy, F. K. Karim, M. Abotaleb, A. Ibrahim, A. A. Abdelhamid, and D. L. Elsheweikh, "Hybrid dipper throated and grey wolf optimization for feature selection applied to life benchmark datasets," *Comput., Mater. Continua*, vol. 74, no. 2, pp. 4531–4545, 2023.

[8] Y. Li, T. Li, and H. Liu, "Recent advances in feature selection and its applications," *Knowl. Inf. Syst.*, vol. 53, no. 3, pp. 551–577, Dec. 2017.

[9] B. Ji, X. Lu, G. Sun, W. Zhang, J. Li, and Y. Xiao, "Bio-inspired feature selection: An improved binary particle swarm optimization approach," *IEEE Access*, vol. 8, pp. 85989–86002, 2020.

[10] A.-D. Li, B. Xue, and M. Zhang, "Improved binary particle swarm optimization for feature selection with new initialization and search space reduction strategies," *Appl. Soft Comput.*, vol. 106, Jul. 2021, Art. no. 107302.

[11] M. Tubishat, M. A. M. Abushariah, N. Idris, and I. Aljarah, "Improved whale optimization algorithm for feature selection in Arabic sentiment analysis," *Int. J. Speech Technol.*, vol. 49, no. 5, pp. 1688–1707, May 2019.

[12] I. Letteri, A. Di Cecco, A. Dyoub, and G. D. Penna, "Imbalanced dataset optimization with new resampling techniques," in *Intelligent Systems and Applications*, K. Arai, Ed. Cham, Switzerland: Springer, 2022, pp. 199–215.

[13] R. Hans and H. Kaur, "Feature selection using metaheuristic algorithms: Concept, applications and population based comparison," in *Proc. Int. Conf. Comput. Perform. Eval. (ComPE)*, Jul. 2020, pp. 558–563.

[14] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE ICNN*, vol. 4, Nov. 1995, pp. 1942–1948.

[15] P. Moradi and M. Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," *Appl. Soft Comput.*, vol. 43, pp. 117–130, Jun. 2016.

[16] K. Mistry, L. Zhang, S. C. Neoh, C. P. Lim, and B. Fielding, "A micro-GA embedded PSO feature selection approach to intelligent facial emotion recognition," *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1496–1509, Jun. 2017.

[17] M. Y. Shams, "Hybrid neural networks in generic biometric system: A survey," *J. Artif. Intell. Metaheuristics*, vol. 1, no. 1, pp. 20–26, 2022.

[18] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, Aug. 2016.

[19] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proc. IEEE Int. Conf. Syst., Man, Cybern., Comput. Cybern. Simul.*, Oct. 1997, pp. 4104–4108.

[20] C.-Y. Lee, T.-A. Le, and C.-L. Hung, "A feature selection approach based on memory space computation genetic algorithm applied in bearing fault diagnosis model," *IEEE Access*, vol. 11, pp. 51282–51295, 2023.

[21] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, nos. 1–4, pp. 131–156, 1997.

[22] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.

[23] O. Sanghoun and C. W. Ahn, "Evolutionary approach for interpretable feature selection algorithm in manufacturing industry," *IEEE Access*, vol. 11, pp. 46604–46614, 2023.

[24] L. Meenachi and S. Ramakrishnan, "Metaheuristic search based feature selection methods for classification of cancer," *Pattern Recognit.*, vol. 119, Nov. 2021, Art. no. 108079.

[25] K. Chen, B. Xue, M. Zhang, and F. Zhou, "An evolutionary multitasking-based feature selection method for high-dimensional classification," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 7172–7186, Jul. 2022.

[26] M. B. Khan, M. A. Noor, K. I. Noor, and Y.-M. Chu, "Higher-order strongly preinvex fuzzy mappings and fuzzy mixed variational-like inequalities," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 1856–1870, Jun. 2021.

[27] H. A. Alsayadi, N. Khodadadi, and S. Kumar, "Improving the regression of communities and crime using ensemble of machine learning models," *J. Artif. Intell. Metaheuristics*, vol. 1, no. 1, pp. 27–34, 2022.

[28] M. B. Khan, M. A. Noor, K. I. Noor, H. Almusawa, and K. S. Nisar, "Exponentially preinvex fuzzy mappings and fuzzy exponentially mixed variational-like inequalities," *Int. J. Anal. Appl.*, vol. 19, no. 4, pp. 518–541, May 2021.

[29] M. B. Khan, M. A. Noor, P. O. Mohammed, J. L. G. Guirao, and K. I. Noor, "Some integral inequalities for generalized convex fuzzy-interval-valued functions via fuzzy Riemann integrals," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 1–16, Sep. 2021.

[30] M. B. Khan, H. M. Srivastava, P. O. Mohammed, and J. L. G. Guirao, "Fuzzy mixed variational-like and integral inequalities for strongly preinvex fuzzy mappings," *Symmetry*, vol. 13, no. 10, p. 1816, Sep. 2021.

[31] A. Oubelaid, N. Taib, and T. Rekioua, "Novel coordinated power sources switching strategy for transient performance enhancement of hybrid electric vehicles," *COMPEL Int. J. Comput. Math. Electr. Electron. Eng.*, vol. 41, no. 5, pp. 1880–1919, Jan. 2022.

[32] N. Neggaz, A. A. Ewees, M. A. Elaziz, and M. Mafarja, "Boosting Salp swarm algorithm by sine cosine algorithm and disrupt operator for feature selection," *Expert Syst. Appl.*, vol. 145, May 2020, Art. no. 113103.

[33] L. Kumar and K. K. Bharti, "A novel hybrid BPSO–SCA approach for feature selection," *Natural Comput.*, vol. 20, no. 1, pp. 39–61, Mar. 2021.

[34] L. M. A. Al-Saedi, M. T. Gaata, M. Abotaleb, and H. Alkattan, "New approach of estimating sarcasm based on the percentage of happiness of facial expression using fuzzy inference system," *J. Artif. Intell. Metaheuristics*, vol. 1, no. 1, pp. 35–44, 2022.

[35] R. Hans and H. Kaur, "Hybrid binary sine cosine algorithm and ant lion optimization (SCALO) approaches for feature selection problem," *Int. J. Comput. Mater. Sci. Eng.*, vol. 9, no. 1, Mar. 2020, Art. no. 1950021.

[36] M. M. Eid, E.-S.-M. El-Kenawy, N. Khodadadi, S. Mirjalili, E. Khodadadi, M. Abotaleb, A. H. Alharbi, A. A. Abdelhamid, A. Ibrahim, G. M. Amer, A. Kadi, and D. S. Khafaga, "Meta-heuristic optimization of LSTM-based deep network for boosting the prediction of monkeypox cases," *Mathematics*, vol. 10, no. 20, p. 3845, Oct. 2022.

[37] H. Saoud, A. Ghadi, M. Ghailani, and B. A. Abdelhakim, "Using feature selection techniques to improve the accuracy of breast cancer classification," in *Innovations in Smart Cities Applications Edition 2*, M. B. Ahmed, A. A. Boudhir, and A. Younes, Eds. Cham, Switzerland: Springer, 2019, pp. 307–315.

[38] H. N. AlEisa, E.-S. M. El-Kenawy, A. A. Alhussan, M. Saber, A. A. Abdelhamid, and D. S. Khafaga, "Transfer learning for chest X-rays diagnosis using dipper throated lgorithm," *Comput., Mater. Continua*, vol. 73, no. 2, pp. 2371–2387, 2022.

[39] M. Saber, "Removing powerline interference from EEG signal using optimized FIR filters," *J. Artif. Intell. Metaheuristics*, vol. 1, no. 1, pp. 8–19, 2022.

[40] Y. Pathak, K. V. Arya, and S. Tiwari, "Feature selection for image steganalysis using Levy flight-based grey wolf optimization," *Multimedia Tools Appl.*, vol. 78, no. 2, pp. 1473–1494, Jan. 2019.

[41] S. Arora, H. Singh, M. Sharma, S. Sharma, and P. Anand, "A new hybrid algorithm based on grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection," *IEEE Access*, vol. 7, pp. 26343–26361, 2019.

[42] H. Jia, Z. Xing, and W. Song, "A new hybrid seagull optimization algorithm for feature selection," *IEEE Access*, vol. 7, pp. 49614–49631, 2019.

[43] A. Oubelaid, M. Y. Shams, and M. Abotaleb, "Energy efficiency modeling using whale optimization algorithm and ensemble model," *J. Artif. Intell. Metaheuristics*, vol. 2, no. 1, pp. 27–35, 2022.

[44] B. T. A. Al-Nuaimi and T. I. Baker, "Weather forecasting over Iraq using machine learning," *J. Artif. Intell. Metaheuristics*, vol. 2, no. 2, pp. 39–45, 2022.

[45] D. S. Khafaga, A. A. Alhussan, E.-S. M. El-kenawy, A. E. Takieldeen, T. M. Hassan, E. A. Hegazy, E. Abdel F. Eid, A. Ibrahim, and A. A. Abdelhamid, "Meta-heuristics for feature selection and classification in diagnostic breast-cancer," *Comput., Mater. Continua*, vol. 73, no. 1, pp. 749–765, 2022.

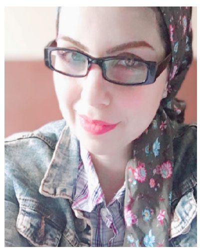
[46] A. E. Takieldeen, E.-S. M. El-kenawy, M. Hadwan, and R. M. Zaki, "Dipper throated optimization algorithm for unconstrained function and feature selection," *Comput., Mater. Continua*, vol. 72, no. 1, pp. 1465–1481, 2022.

[47] K. Kakouche, T. Rekioua, S. Mezani, A. Oubelaid, D. Rekioua, V. Blazek, L. Prokop, S. Misak, M. Bajaj, and S. S. M. Ghoneim, "Model predictive direct torque control and fuzzy logic energy management for multi power source electric vehicles," *Sensors*, vol. 22, no. 15, p. 5669, Jul. 2022.
[48] L. Abualigah and A. J. Dulaimi, "A novel feature selection method for data mining tasks using hybrid sine cosine algorithm and genetic algorithm," *Cluster Comput.*, vol. 24, no. 3, pp. 2161–2176, Sep. 2021.
[49] J. L. Awange, B. Paláncz, R. H. Lewis, and L. Völgyesi, "Particle swarm optimization," in *Mathematical Geosciences: Hybrid Symbolic-Numeric Methods*. Cham, Switzerland: Springer, 2018, pp. 167–184.
[50] V. Kumar and D. Kumar, "Binary whale optimization algorithm and its application to unit commitment problem," *Neural Comput. Appl.*, vol. 32, no. 7, pp. 2095–2123, Apr. 2020.
[51] Y. Chen, J. Xi, H. Wang, and X. Liu, "Grey wolf optimization algorithm based on dynamically adjusting inertial weight and Levy flight strategy," *Evol. Intell.*, vol. 16, no. 3, pp. 917–927, Feb. 2022.
[52] S. H. S. Moosavi and V. K. Bardsiri, "Satin bowerbird optimizer: A new optimization algorithm to optimize ANFIS for software development effort estimation," *Eng. Appl. Artif. Intell.*, vol. 60, pp. 1–15, Apr. 2017.
[53] S. D. Immanuel and U. Kr. Chakraborty, "Genetic algorithm: An approach on optimization," in *Proc. Int. Conf. Commun. Electron. Syst. (ICCES)*, Jul. 2019, pp. 701–708.
[54] X.-S. Yang, "Chapter 9—Firefly algorithms," in *Nature-Inspired Optimization Algorithms*, X.-S. Yang, Ed., 2nd ed. New York, NY, USA: Academic, 2021, pp. 123–139.
[55] L. Abualigah, "Multi-verse optimizer algorithm: A comprehensive survey of its results, variants, and applications," *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12381–12401, Aug. 2020.
[56] A. I. Hafez, H. M. Zawbaa, E. Emary, and A. E. Hassanien, "Sine cosine optimization algorithm for feature selection," in *Proc. Int. Symp. Innov. Intell. Syst. Appl. (INISTA)*, Aug. 2016, pp. 1–5.
[57] E. M. El-Kenawy, S. Mirjalili, F. Alassery, Y.-D. Zhang, M. M. Eid, S. Y. El-Mashad, B. A. Aloyaydi, A. Ibrahim, and A. A. Abdelhamid, "Novel meta-heuristic algorithm for feature selection, unconstrained functions and engineering problems," *IEEE Access*, vol. 10, pp. 40536–40555, 2022.
[58] D. S. Khafaga, A. A. Alhussan, E. M. El-Kenawy, A. Ibrahim, M. M. Eid, and A. A. Abdelhamid, "Solving optimization problems of metamaterial and double T-shape antennas using advanced meta-heuristics algorithms," *IEEE Access*, vol. 10, pp. 74449–74471, 2022.
[59] A. A. Alhussan, D. S. Khafaga, E. M. El-Kenawy, A. Ibrahim, M. M. Eid, and A. A. Abdelhamid, "Pothole and plain road classification using adaptive mutation dipper throated optimization and transfer learning for self driving cars," *IEEE Access*, vol. 10, pp. 84188–84211, 2022.
[60] E.-S.-M. El-Kenawy, S. Mirjalili, A. A. Abdelhamid, A. Ibrahim, N. Khodadadi, and M. M. Eid, "Meta-heuristic optimization and keystroke dynamics for authentication of smartphone users," *Mathematics*, vol. 10, no. 16, p. 2912, Aug. 2022.
[61] E.-S.-M. El-kenawy, F. Albalawi, S. A. Ward, S. S. M. Ghoneim, M. M. Eid, A. A. Abdelhamid, N. Bailek, and A. Ibrahim, "Feature selection and classification of transformer faults based on novel meta-heuristic algorithm," *Mathematics*, vol. 10, no. 17, p. 3144, Sep. 2022.
[62] C. L. Blake and C. J. Merz. (1998). *UCI Repository of Machine Learning Databases*. Accessed: May 1, 2023. [Online]. Available: https://archive.ics.uci.edu/ml
[63] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016.

ABDELAZIZ A. ABDELHAMID received the M.Sc. degree in computer science from the Faculty of Computer and Information Sciences, Ain Shams University, and the Ph.D. degree in computer engineering from the Faculty of Engineering, Auckland University, New Zealand. He is currently an Assistant Professor with the Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University. He is also an Assistant Professor with the Computer Science Department, College of Computing and Information Technology, Shaqra University. His research interests include speech and image processing, and machine learning-based intelligent systems.

EL-SAYED M. EL-KENAWY (Senior Member, IEEE) is currently an Assistant Professor with the Delta Higher Institute for Engineering and Technology (DHIET), Mansoura, Egypt. He is inspiring and motivating students by providing a thorough understanding of a variety of computer concepts. He has pioneered and launched independent research programs. His research interests include computer science and machine learning. He is also adept at explaining sometimes complex concepts in an easy-to-understand manner.

ABDELHAMEED IBRAHIM (Member, IEEE) received the bachelor's and master's degrees in engineering from the Computer Engineering and Systems Department, in 2001 and 2005, respectively, and the Ph.D. degree in engineering from the Faculty of Engineering, Chiba University, Japan, in 2011. He was with the Faculty of Engineering, Mansoura University, Egypt, from 2001 to 2007. He is currently an Associate Professor of computer engineering. He has published more than 110 publications with over 3200 citations, and his H-index is 28. His research interests include machine learning, optimization, swarm intelligence, metaheuristic, and pattern recognition. He serves as a Reviewer for Engineering Applications of Artificial Intelligence, IEEE Access, IEEE Journal of Biomedical and Health Informatics, Expert Systems with Applications, *Biomedical Signal Processing and Control, Sensors, PLOS ONE, Alexandria Engineering Journal, Scientific Reports, Renewable and Sustainable Energy Reviews*, and other respected journals.

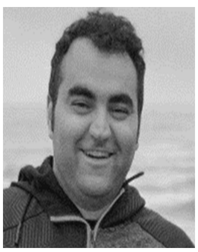


MARWA METWALLY EID (Member, IEEE) received the Ph.D. degree in electronics and communications engineering from the Faculty of Engineering, Mansoura University, Egypt, in 2015. She has been an Assistant Professor with the Delta Higher Institute for Engineering and Technology, since 2011. Her current research interests include image processing, encryption, wireless communication systems, and field programmable gate array (FPGA) applications.

DOAA SAMI KHAFAGA received the B.Sc. degree (Hons.) in computer and information sciences and in computer science and the M.Sc. and Ph.D. degrees in computer science from the Faculty of Computers and Artificial Intelligence, Helwan University, Egypt, in 2003, 2008, and 2013, respectively. She has 18 years academic experience. She was with the Computer Science Department, College of Information Technology and Artificial Intelligence, Misr University for Science and Technology, Egypt, Computer Science Department, Institute of Public Administration, Saudi Arabia, and Computer Science Department, Faculty of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University. She is a fellow of the U.K. Higher Education Academy (FHEA). Currently, she is a reviewer in some journals. Her research interests include data science, artificial intelligence, machine learning, data mining, and software engineering.

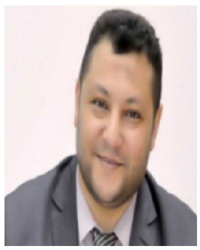
AMEL ALI ALHUSSAN received the B.Sc., M.Sc., and the Ph.D. degrees in computer and information sciences from King Saud University, Saudi Arabia. Her M.Sc. thesis in software engineering and the Ph.D. thesis in artificial intelligent. She is currently an Assistant Professor with the Computer Sciences Department, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University (PNU), Saudi Arabia. She is worked in her collage in various administrative and academic positions. Her research interests include machine leaning, networking, and software engineering.

SEYEDALI MIRJALILI (Senior Member, IEEE) is currently a Full Professor, a top AI Scientist, and the Director of the Centre for Artificial Intelligence Research and Optimization, Torrens University Australia. He is internationally recognized for his advances in swarm intelligence and optimization, including the first set of algorithms from a synthetic intelligence standpoint—a radical departure from how natural systems are typically understood; and a systematic design framework to reliably benchmark, evaluate, and propose computationally cheap robust optimization algorithms. He has published over 500 publications with over 72000 citations and an H-index of 88. As the most cited researcher in robust optimization, he is in the list of 1% highly-cited researchers and named as one of the most influential researchers in the world by the Web of Science. His research interests include robust optimization, engineering optimization, multi-objective optimization, swarm intelligence, evolutionary algorithms, and artificial neural networks. He is an Editor of several journals, including *Neurocomputing*, *Applied Soft Computing*, *Advances in Engineering Software*, *Applied Intelligence*, and *Engineering Applications of Artificial Intelligence*.



NIMA KHODADADI (Member, IEEE) received the bachelor's degree in civil engineering and the master's degree in structural engineering from the University of Tabriz (one of the ten top universities in Iran). He is currently a Researcher with the Iran University of Science and Technology (IUST). His research interests include the field of steel structures, particularly in the experimental and numerical investigation of steel braced frames. In addition, he has been actively involved in the area of engineering optimization, especially in evolutionary algorithms. Using comprehensive finite element analyses, he has also investigated the use of different shaped sections in real steel frames. Recently, he has been engaged in research in the area of engineering optimization, especially in solving large-scale and practical structural design problems.

WEI HONG LIM (Senior Member, IEEE) received the B.Eng. degree (Hons.) in mechatronic engineering and the Ph.D. degree in computational intelligence from Universiti Sains Malaysia, Penang, Malaysia, in 2011 and 2014, respectively. From 2015 to 2017, he was a Postdoctoral Researcher with the Intelligent Control Laboratory, National Taipei University of Technology, Taiwan, where he was a Visiting Researcher, in 2019. He is currently an Assistant Professor with the Faculty of Engineering, Technology and Built Environment, UCSI University. He is also working with three national research grants awarded by the Ministry of Education, Malaysia, and five internal grant projects supported by UCSI University. He has published more than 50 research articles in research areas related to computational intelligence, optimization algorithms, energy management, and digital image processing. He has been actively involved in various professional bodies and registered as the European Engineer (EUR ING) with the Fédération Européenne d'Associations Nationales d'Ingénieurs (FEANI), a Chartered Engineer (C.Eng.) with the Engineering Council, U.K., (ECUK), and a Professional Technologist (P.Tech.) with the Malaysia Board of Technologist (MBOT). He is also an active Reviewer of various reputable journals, such as IEEE ACCESS, *Complexity*, *Mathematical Problem in Engineering*, and *Computational Intelligence and Neuroscience*.



MAHMOUD Y. SHAMS received the bachelor's degree in electronics and communication from the Faculty of Engineering, Mansoura University, in 2004, the master's degree in computer vision and pattern recognition from the Faculty of Computer and Information Sciences, Mansoura University, and the Ph.D. degree from the Computer Science Department, Mansoura University. He is currently an Assistant Professor with the Machine Learning and Information Retrieval Department, Faculty of Artificial Intelligence, Kafr Elsheikh University. He has published over 40 articles in refereed international journals. His research interests include the application of deep learning approaches and computer vision techniques to the field of medical images.