A Hybrid Intelligent System Based on Feature Selection and Ensemble Learning for Detecting Parkinson's Disease

Seyyed Ahmad Hashemi[1], Faraein Aeini[2*], Homayun Motameni[3], Behnam Barzegar[4]

[1,2,3] Department of Computer Engineering, Islamic Azad University, Sari Branch, Sari, Iran

[4] Department of Computer Engineering, Islamic Azad University, Babol Branch, Babol, Iran

Corresponding Author Tel: 09113511764

E-mail addresses: ahmadhashemi5555@gmail.com (Seyyed Ahmad Hashemi)

aeini@iausari.ac.ir (Faraein Aeini)

motameni@iausari.ac.ir (Homayun Motameni)

barzegar@iauns.ac.ir (Behnam Barzegar)

Abstract

In recent years, Parkinson's disease (PD) has become a global health problem. Early diagnosis of the disease has a high impact on the quality of treatment. Various machine learning methods and classification algorithms have been proposed to enhance the accuracy in PD detection. Accordingly, this paper proposes a hybrid intelligent system, which involves preprocessing using normalization, feature selection using an Improved Binary Whale Optimization Algorithm (IBWOA), and classification using a New Ensemble Learning Strategy (NELS). In this paper, IBWOA was used to choose the optimal subset of features for prediction, while NELS was employed to handle the learning process. The PD dataset required for the purposes of this research was extracted from the UCI machine learning database. The experimental results showed that the combination of preprocessing, feature selection, and ensemble learning gave a classification accuracy of 96.9231% for the PD dataset. The results also showed that both the main phases of feature selection and ensemble learning are effective in improving system performance. The detection accuracy of the proposed system improved by 0.9231% compared to the best model in the current literature.

Keywords: Binary Whale Optimization Algorithm, Classification, Ensemble Learning, Feature Selection, Parkinson's Disease.

1. Introduction

Parkinson's disease (PD) is one of the nervous system diseases that have great impacts on people’s health worldwide. Early detection of PD can significantly reduce its symptoms. Studies show that about 90% of patients with PD have vocal impairments [1]. Therefore, the measurement of vocal impairments is a reliable approach to diagnosing and monitoring PD [2–4]. The problem of PD can be considered as a binary classification problem [5–7]. Each sample represents the value obtained from the vocal signals. Class 0 represents a healthy sample and class 1 represents a diseased sample. The performance of classification significantly depends on three steps, i.e., preprocessing, feature selection, and learning process, which are explained in the following.

A real-world dataset may have problems such as missing values, noisy data, and inconsistent data [8], which negatively affect the classification performance. Data preprocessing is an essential step to improve the performance of classification, which includes the main operations of quantification, discretization, normalization, and removing noisy data.

On the other hand, feature selection is one of the important techniques in the field of dimensionality reduction, which has a great impact on the performance of classification algorithms. Feature selection methods are divided into two categories: filter-based and wrapper-based [9]. Filter-based feature selection methods attempt to find the optimal subset of features using statistical analysis, while wrapper-based feature selection methods use learning algorithms for the same purpose. Many meta-heuristic algorithms have already been developed to solve the problem of feature selection, which include Genetic Algorithm (GA) [10], Binary Particle Swarm Optimization (BPSO) [11], Binary Grey Wolf Optimization (BGWO) [12], Binary Butterfly Optimization Algorithm (BBOA) [13], Binary Ant Colony Optimization (BACO) [14], Binary Grasshopper Optimization Algorithm (BGOA) [15], Binary Artificial Bee Colony (BABC) [16], and Binary Whale Optimization Algorithm (BWOA) [17]. In general, meta-heuristic algorithms have been found effective in solving the feature selection problem.

Wrapper-based feature selection methods exhibit superiority over filter-based methods in diagnosing PD due to their ability to directly evaluate feature subsets based on the performance of the predictive model. Unlike filter-based methods that assess features individually and independently of the learning algorithm, wrapper-based methods consider the interactions and dependencies between features within the context of the classifier. This leads to a more accurate identification of the most relevant and effective features, resulting in enhanced performance and accuracy of the diagnostic system. Additionally, wrapper-based methods allow for better customization and flexibility in incorporating domain-specific knowledge and constraints during the feature selection process, ensuring that the selected features align closely with the underlying problem and objectives of the prediction model, ultimately contributing to more reliable and precise PD detection.

Ensemble learning is a special case of approaches where several learning models are combined to provide an integrated model for solving the same problem. The purpose of designing an ensemble learning strategy is to build a global model that improves classification criteria such as accuracy, error rate, sensitivity, and specificity. To combine basic classifiers in an ensemble learning strategy, the voting system, the statistical techniques, and the belief functions can be used [18]. Studies show that an ensemble learning strategy has better performance compared to basic classifiers [19]. In addition, studies have shown that the variety of basic classifiers in the ensemble learning strategy can affect the performance of the aggregated model [20]. In general, basic classifiers such as Decision Tree (DT) [21], Random Forest (RF) [22], K-Nearest Neighbor (KNN) [23], Support Vector Machine (SVM) [24], Artificial Neural Network (ANN) [25], and Bayesian Network (BN) [26] can be used in the ensemble learning strategy.

The Improved Binary Whale Optimization Algorithm (IBWOA) incorporates several significant improvements that elevate its performance and efficiency. Three notable enhancements include:

1. Random Initial Population Generation with Chaotic Maps: The algorithm now utilizes chaotic maps to generate a more diverse initial population. This innovative approach enables a broader exploration of the solution space, reducing the likelihood of getting trapped in local optima and ultimately leading to more effective convergence towards the global optimum.
2. Random Algorithm Parameter Generation with Chaotic Maps: The introduction of chaotic maps for generating random algorithm parameters ensures a higher degree of randomness. This modification contributes to a more robust optimization process by mitigating the risk of premature convergence and fostering improved search capabilities.
3. Random Decision Strategy on the Best Population Solution in Each Iteration: To address convergence speed concerns, a random decision strategy has been implemented, focusing on the best population solution in each iteration. This adaptive mechanism reduces the convergence speed, allowing for a more thorough exploration of the problem space and enhancing the algorithm's ability to identify the most optimal solution.

These enhancements collectively elevate the binary whale optimization algorithm's performance, enabling it to tackle complex problems with greater precision and effectiveness while maintaining the robustness of its core search mechanism.

This paper proposes a hybrid intelligent system for the diagnosis of PD, which involves the preprocessing step using normalization, feature selection step using an IBWOA, and learning process step using a New Ensemble Learning Strategy (NELS). The following are key contributions of the present research:

- Developing a new hybrid intelligent system for PD detection, combining preprocessing, feature selection, and ensemble learning.
- Proposing a new feature selection method, an IBWOA, for selecting the optimal subset of features for prediction.
- Developing a New Ensemble Learning Strategy (NELS) based on SVM, KNN, and RF to handle the learning process in PD detection.
- Comprehensively evaluating the proposed system on a benchmark dataset, demonstrating its high accuracy in classifying PD cases.
- Offering insights into the effectiveness of combining preprocessing, feature selection, and ensemble learning for the purpose of PD detection.

The rest of this paper is organized as follows. The related works are presented in section 2. The used algorithms are described in section 3. Section 4 is devoted to designing an intelligent PD system based on an IBWOA and ensemble classifier. Implementation results are detailed in section 5, and finally, the conclusion and future works are explained in section 6.

2. Related works

The authors in [27] proposed a parallel neural network method to predict PD. In this method, a parallel neural network structure is defined in order to reduce the possibility of wrong decisions. A simple voting system is additionally presented to decide between possible outputs. [28] presented a computer-based system for the diagnosis of PD based on feature selection and ensemble learning. In this method, a linear SVM is used to select the optimal subset of features, and a rotation forest ensemble classifier is performed for the learning process. In [29], the authors presented a computer-based system for the diagnosis of PD similar to the previous system. In this method, a linear SVM selects the optimal subset of features, and a KNN ensemble classifier is performed for the learning process.

[30] studied the effect of reducing the dataset dimensions on the diagnosis of PD. They used GA and BPSO to determine the optimal subset of features and also trained 11 basic classifiers for the learning process. The hybrid system of GA and Ada Boost had the best performance in the PD diagnosis. [31] proposed a hybrid method based on the wrapper-based feature selection and a tree-based boosting algorithm. The main focus of this work was on finding the best speech processing algorithm. In [32], the authors presented a method for diagnosing PD based on GA and SVM, where GA was used to solve the feature selection problem. The accuracy of the SVM was considered as a fitness function of GA.

[33] developed an intelligent system for diagnosing PD based on deep learning, which was a learning process using convolutional neural networks. The goal of the learning process is to find the severity of the disease. In another study [34], an efficient system was proposed for diagnosing PD based on deep learning. This system used a deep belief network comprising two piled Restricted Boltzmann Machines. In [35], the authors proposed a new meta-heuristic algorithm, Coronavirus Herd Immunity Optimizer (CHIO). A greedy crossover strategy was designed in this algorithm to better explore the search space.

The researchers in [36] suggested an ensemble approach to diagnosing PD, where feature selection was integrated with Deep Neural Network (DNN) to select the optimal subset of features. [37] proposed a method for the early diagnosis of PD based on machine learning, where four basic classifiers, i.e., SVM, KNN, RF, and logistic regression, were used for the learning process. They also adopted the Principal Component Analysis (PCA) technique to solve the feature selection problem. [38] proposed a method for diagnosing PD by the RF classifier and feature selection technique. This method employed Recursive Feature Elimination (RFE) to select the most appropriate feature subset. Table 1 compares the models already proposed in the relevant literature for diagnosing PD.

These research gaps highlight important areas for further investigation and improvement in the development of PD diagnosis models using machine learning and artificial intelligence techniques:

1. Feature selection: Reducing the dimensionality of the data by selecting the most relevant features can improve model performance and reduce computational complexity. However, identifying the optimal feature subset remains a challenge. Further research is needed to develop advanced feature selection techniques tailored to PD diagnosis.
2. Enhancing the learning process with ensemble learning strategies: Most existing studies focus on individual learning algorithms. Incorporating ensemble learning strategies could potentially improve model accuracy and robustness by leveraging the collective knowledge of multiple models. Further research should explore the benefits and challenges of ensemble learning strategies in the context of PD diagnosis.
3. Optimization of meta-heuristic algorithms for feature selection: Meta-heuristic algorithms have shown promise in addressing the feature selection problem. However, there is still room for improvement in terms of their efficiency, accuracy, and convergence speed. Future research could focus on refining existing algorithms or developing new ones specifically designed for feature selection in PD diagnosis models.

Addressing these research gaps could lead to more accurate, efficient, and reliable PD diagnosis models, ultimately supporting early detection and effective treatment of the disease.

3. Preliminaries

This section describes the three fundamental concepts used in the present article, i.e., BWOA, basic classifiers, and chaotic maps.

3.1. Binary Whale Optimization Algorithm

Whale Optimization Algorithm (WOA) was first developed by Mirjalili and Lewis [39]. Each solution in this algorithm is treated as a whale in N-dimensional space, and each whale has a position vector. WOA is a population-based algorithm inspired by the behavior of humpback whales. This optimization algorithm involves eight main steps each of which is described as follows:

- Step 1 (Initializing)

In this step, the parameters of WOA can be adjusted. Table 2 shows the characteristics of the parameters of WOA. Among all, the three parameters of P, $\bar{r}$, and L must be generated randomly; therefore, improving the simulation process of generating random numbers can lead to improved algorithm performance.

- Step 2 (Generating a random initial population)

A population is a group of whales that interact with each other to solve an optimization problem. The population cases are stored as a two-dimensional matrix of size $ps \times N$, according to Equation (1). The initial population in the algorithm should be generated randomly.

$$pop = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ . & . & & . \\ . & . & \cdots & . \\ . & . & & . \\ x_{ps1} & x_{ps2} & \cdots & x_{psN} \end{bmatrix} \tag{1}$$

- Step 3 (Applying fitness function)

The fitness function is defined as a maximization or minimization function based on the nature of the optimization problem. With respect to objective functionality, the fitness function is formulated as shown in Equation (2), where $\bar{X}$ denotes the whale in the population:

$$fit_i = Min / Max\,(f\,(\bar{X}_l)), \quad \bar{X}_l \in (LB_i, UB_i) \tag{2}$$

- Step 4 (Selecting the best whale of the population)

The best whale in the population is found based on fitness function. If the optimization problem is maximization, Equation (3) is used to select the best whale, whereas if the problem is minimization, Equation (4) is used to select the best whale. In both equations, $\bar{X}_{best}$ denotes the best whale found in the population:

$$\bar{X}_{best} = Max\,(fit_1, fit_2, fit_3, ..., fit_{ps}) \tag{3}$$

$$\bar{X}_{best} = Min(fit_1, fit_2, fit_3, ..., fit_{ps}) \tag{4}$$

- Step 5 (Developing the first search model)

In this search model, the position vector of whales in the population is updated based on the best whale found that this behavior can be formulated based on Equations (5). To realize his search model, Equations (6-8) should be calculated.

$$\bar{X}(t+1) = \bar{X}_{best}(t) - \bar{A} \cdot \bar{D}_1 \tag{5}$$

$$\bar{A} = 2a \cdot \bar{r} - a \tag{6}$$

$$\bar{D}_1 = \left| \bar{C} \cdot \bar{X}_{best}(t) - \bar{X}(t) \right| \tag{7}$$

$$\bar{C} = 2 \cdot \bar{r} \tag{8}$$

- Step 6 (Developing the second search model)

Similar to the previous search model, in this model, the position vector of whales in the population is updated based on the best whale found. This behavior can be formulated based on Equation (9). To realize his search model, Equation (10) should be calculated.

$$\bar{X}(t+1) = \bar{D}_2 \cdot e^{BL} \cdot \cos(2\pi L) + \bar{X}_{best}(t) \tag{9}$$

$$\bar{D}_2 = \left| \bar{X}_{best}(t) - \bar{X}(t) \right| \tag{10}$$

- Step 7 (Developing the third search model)

In this search model, the position vector of whales in the population is update based on the randomly whale. This behavior can be formulated based on Equation (11) where $\overline{X_{rand}}$ denotes the randomly whale. To realize his search model, Equation (12) should be calculated.

$$\bar{X}(t+1) = \bar{X}_{rand}(t) - \bar{A} \cdot \bar{D}_3 \tag{11}$$

$$\bar{D}_3 = \left| \bar{C} \cdot \bar{X}_{rand}(t) - \bar{X}(t) \right| \tag{12}$$

- Step 8 (Applying the control mechanism)

The control mechanism is formulated by Equation (13). For each whale in the population and in each round of iteration, a decision is made to select the updated model based on the control parameters P and $|\bar{A}|$.

$$\bar{X}(t+1) = \begin{cases} Use\_Equation9 & if \quad P >= 0.5 \\ Use\_Equation5 & if \quad P < 0.5 \quad and \quad |\bar{A}| < 1 \\ Use\_Equation11 & if \quad P < 0.5 \quad and \quad |\bar{A}| >= 1 \end{cases} \tag{13}$$

WOA is categorized as one of the swarm-based algorithms where the initial population is produced randomly. The quality of each whale is evaluated by a fitness function, and the best individual is selected from the initial population. Afterward, an iterative process starts until reaching the pre-defined stopping criterion. Whales update their positions at each iteration based on Equations (5), (9), and (11). By reaching the stopping condition, the best whale of the final population is selected as the final answer**.**

WOA is used to solve continuous optimization problems so that each whale can be moved anywhere in space. Although, many computer problems, including feature selection, have limitations in coding solutions. To solve discrete optimization problems with binary coding, another version of WOA, named BWOA, should be used. Transfer functions are used in meta-heuristic algorithms to convert continuous space into discrete space. The S-shaped transfer function [40] and the V-shaped transfer function [41] are the two main functions in converting a continuous space of solutions to a discrete one. Fig. 1 shows the flowchart of BWOA.

- S-shaped transfer function

Let $\bar{X}(t)$ represent the whale in the population at iteration *t* before the update process and $\bar{Z}(t)$ represent the same whale in the population at iteration *t* after the update process. The S-shaped transfer function is modeled based on Equations (14) and (15):

$$S(\bar{Z}(t)) = \frac{1}{1+e^{-\bar{Z}(t)}} \tag{14}$$

$$\bar{X}(t+1) = \begin{cases} 0 & rand < S(\bar{Z}(t)) \\ 1 & rand >= S(\bar{Z}(t)) \end{cases} \tag{15}$$

- V-shaped transfer function

Assume that $\bar{X}(t)$ represents the whale in the population at iteration *t* before the update process and $\bar{Z}(t)$ represents the same whale in the population at iteration *t* after the update process. The V-shaped transfer function is modeled according to Equations (16) and (17):

$$V(\bar{Z}(t)) = \left| erf\left(\frac{\sqrt{\pi}}{2} \bar{Z}(t)\right) \right| \tag{16}$$

$$\bar{X}(t+1)=\begin{cases}\square\,\bar{X}(t) & rand < V(\bar{Z}(t))\\ \bar{X}(t) & rand >= V(\bar{Z}(t))\end{cases} \tag{17}$$

3.2. Basic classifiers

This study uses three classifiers, SVM, KNN, and RF, which are explained in the following.

SVM is an efficient classification algorithm based on statistical learning theory. This algorithm is a learning framework used to solve two types of problem: two-class classification and multi-class classification. SVM solves the classification problem with datasets that are linearly separable by finding the hyperplane between the samples. The algorithm selects a hyperplane which has the largest distance from the closest sample in each class. SVM uses mapping functions known as kernel functions for datasets that are not linearly separable. The radius-based kernel function is one of the most well-known mapping functions, which is used in SVM classifiers [42]. This is a popular kernel function that is used in SVM and defined as follows: where *gamma* is a parameter that controls the width of the kernel function, and $\|x-y\|$ is the Euclidean distance between x and y.

$$F(x,y)=\exp(-gamma\times\|x-y\|^2) \tag{18}$$

KNN is also an efficient classification algorithm that works based on the majority voting rule. This non-parametric algorithm determines the class of the test sample based on the class of the neighboring training samples. It works as follows: first, the Euclidean distance between the test sample and all the training samples is calculated, and then the K of the nearest training samples is selected as neighbors. Finally, the class of the test sample is estimated based on the class of the neighbors [43].

FR is an ensemble learning method that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. RF has several advantages or efficiencies that make it a popular choice in machine learning such as Handle high-dimensional data and Handle missing data. RF is likewise a classification algorithm, but it works based on the combination of multiple decision trees. In this algorithm, a large number of decision trees are developed, each decision has the ability to make the decision on its own. The final decision on the RF algorithm is made based on the averaging of decision trees. The number of trees is an important factor that affects the performance of the RF algorithm [44].

3.3. Chaotic maps

Chaotic maps have key properties such as periodicity, quasi-random behavior, high sensitivity to initial conditions, and parameter control, which make them very suitable for generating pseudo-random numbers. In the following, the logistic chaotic map and Chebishev chaotic map are described.

- Logistic chaotic map

The mathematical model of the logistic chaotic map is shown in Equation (19) [45] where γ is the control parameter and RN_0 is the initial state variable, which is in the interval [0,1]:

$$RN_{n+1}=\gamma\times RN_n\times(1-RN_n),\qquad 0<\gamma\le 4 \tag{19}$$

- Chebishev chaotic map

The mathematical model of the Chebishev chaotic map is shown in Equation (20) [46] where μ is the control parameter and RN_0 is the initial state variable, which is in the interval [-1,1]:

$$RN_{n+1}=\cos(\mu\cos^{-1}RN_n),\qquad \mu>0 \tag{20}$$

4. Hybrid intelligent system

This section describes the proposed hybrid intelligent system for PD diagnosis. The overall framework of the proposed hybrid intelligent system is illustrated in Fig. 2. The proposed system consists of four steps, namely preprocessing, dataset splitting, learning process, and feature selection, which are described in detail as follows.

4.1. Preprocessing

The PD dataset selected in this study does not have the missing data, but it has an inherent sparsity of data. The inherent sparsity of data can lead to incorrect analysis of the dataset. Feature normalization is one of the pre-processing stages, which normalizes the features to a specific range of values without modifying the essence of data. In addition, feature normalization affects the efficiency of basic classifiers that work based on distance calculation, such as KNN. Feature normalization is modeled according to Equation (21). The Min/Max normalization technique normalizes the values of each feature in the range of [0,1].

$$F_{ij} = \left(\frac{F_{ij} - (\min\_value\_of\_F_j)}{(\max\_value\_of\_F_j) - (\min\_value\_of\_F_j)}\right) \tag{21}$$

The min-max normalization technique offers several advantages when applied to PD data sets:

- Data standardization: Min-max normalization helps standardize the data, ensuring that all input features are on the same scale. This is particularly important when dealing with diverse data sources in PD diagnosis, such as clinical measurements, imaging data, and sensor data.
- Improved algorithm performance: Normalized data can improve the performance of machine learning algorithms by minimizing the impact of outliers and ensuring that all features contribute equally to the model. This leads to more accurate and reliable predictions.
- Faster training and convergence: With min-max normalization, the data is transformed to a specific range, typically between 0 and 1. This can help machine learning models converge faster and reduce the overall training time.
- Preserving data distribution: The min-max normalization technique preserves the original distribution and shape of the data, which is crucial for maintaining the integrity of the PD data and ensuring that the results accurately represent the underlying data structure.

4.2. Dataset splitting

In the proposed system, the k-fold cross validation technique is used to divide the standard dataset into two parts: the training dataset and the test dataset. This system uses the 5-fold cross validation technique to evaluate the proposed classification model. For the purposes of this research, the training dataset accounted for 80% of the total samples, while the test dataset for 20%, in each run of the k-fold cross validation technique.

4.3. Learning process

In this section, a New Ensemble Learning Strategy (NELS) is proposed based on SVM, KNN, and RF for handling the learning process. Since the ability of basic classifiers on the PD dataset is different, designing an ensemble learning strategy can lead to improving the accuracy of PD diagnosis. Fig. 3 depicts the scheme of NELS. The NELS consists of three steps:

evaluation, selection, and aggregation. In the following, each of the steps of NELS is introduced.

- Step 1 (Evaluation)

In this step, the average accuracy of basic classifiers on five independent test datasets is calculated. The pool includes an initial group of 15 different basic classifiers composed of 5 SVM expert systems, 5 KNN expert systems, and 5 RF expert systems.

- Step 2 (Selection)

In the selection step, from among the 15 basic classifiers, some of them are selected to structure the ensemble classifier. The selection step is shown in Algorithm 1.

Algorithm 1: Selection step of NELS

Input:
Initial \\ the Initial Set of Classifiers
Output:
Optimal \\ the Optimal set of classifiers

```
Initial={CL_1,CL_2,…,CL_15}
Optimal={}
CN=15
for i=1:CN
   Mat_1(1,i)=Accuracy (CL_i)
end for
Mat_2=sort(Mat_1)
SUM=0
for i=1:(CN/2)
    SUM=SUM+Mat_2(1,i)
end for
T=SUM/(CN/2)
for i=1:CN
     if Accuracy (CL_i)> T
        Optimal= Optimal+ CL_i
     end if
end for
```

- Step 3 (Aggregation)

After selecting classifiers, a voting system is defined to aggregate the classifiers. In this system, each classifier announces its vote for each test sample. If a classifier categorizes a test sample as healthy, then +1 is produced as output; on the other hand, if the class of the test sample is diagnosed as patient, then -1 is produced as output. The final vote for a sample test is calculated using Equation (22), where O is the number of selected classifiers and $Y_i = \{+1.-1\}$:

$$Final\_decision = \mathrm{sgn}\left(\sum_{i=1}^{O} Y_i\right) \tag{22}$$

In Equation (22), *sgn* is a well-known mathematical function. If the outcome is +1, it means that the ensemble classifier has categorized the test sample as healthy, and in the case of -1, as patient. In addition, if the result of Equation (22) is 0, it clarifies that the ensemble classifier is unable to detect the class of the test sample and the decision-making process has failed. The same process is done on all the test samples, and the accuracy of the ensemble classifier is measured.

4.4. Feature selection

Feature selection is a discrete optimization problem where search agents are limited to binary values 0 and 1. In this method, each solution is represented by a one-dimensional array, and the length of this array is equal to the number of features in the dataset. Each element of this array can consist of two values, 0 and 1. The value of 0 indicates that the corresponding feature has not been selected, whereas the value of 1 indicates that the corresponding feature has been selected. In this paper, an Improved Binary Whale Optimization Algorithm (IBWOA) is proposed for the feature selection task.

- Step 1 (Initialization)

In this step, all parameters of IBWOA are initialized. Moreover, the chaotic maps are used to generate three random parameters P, $\bar{r}$, and L. The chaotic maps are also used to generate the initial random population of IBWOA. Table 3 shows the characteristics of the chaotic maps used in IBWOA. The initial values of chaotic maps for random population generation and random parameter generation in Table 3 were carefully selected after conducting a thorough analysis of various simulations. Our objective was to identify values that would generate more random-like behavior, thus ensuring a higher degree of entropy. By promoting randomness, we aimed to increase the probability of exploring a wider range of solutions, leading to better optimization outcomes. Furthermore, our goal was to enhance the transparency and comprehensibility of our methodology by clearly documenting the selected values and their contribution to achieving higher entropy.

- Step 2 (Iteration process)

This step continues until the stopping condition is reached. In each iteration of IBWOA, the whales update their positions based on one of Equations (5), (9), or (11). Afterward, the S-shaped or V-shaped transfer functions are used to convert the solutions into a binary format. In IBWOA, a strategy is presented to reduce the speed of convergence. Algorithm 2 presents the strategy defined to reduce the speed of convergence in IBWOA. The proposed strategy consists of three main parts. In the first part, the best solution is selected based on the fitness function. In the second part, three different solutions are generated based on the best solution found in the population and the random decision structure. In the third part, from among the four available solutions, the best ones based on the fitness function are selected as the best solution for the population. This algorithm effectively introduces randomness and diversity into the population, which in turn contributes to reducing the convergence speed and improving the overall performance of the IBWOA algorithm in solving the feature selection problem for PD detection. In this description, we elaborate on how the coefficients B_1, B_2, and B_3 play a crucial role in creating three auxiliary solutions based on the best solution found in the IBWOA. These coefficients are randomly assigned values between 0 and 1, corresponding to the elements of the best population solution. The generation of the first auxiliary solution involves comparing B_1 to a threshold of 0.5. If B_1 is greater than or equal to 0.5, the respective element in the best solution is flipped (subtracting the original value from 1). Otherwise, the element remains unchanged. This process is applied to all elements in the best solution. For the second auxiliary solution, if B_2 is greater than or equal to 0.5, the corresponding element in the initial half of the best solution is flipped. However, if B_2 is less than 0.5, the element stays the same. Similarly, in the case of the third auxiliary solution, if B_3 is greater than or equal to 0.5, the corresponding element in the latter half of the best solution is flipped. Otherwise, the element remains untouched. This method introduces randomness and diversity when generating new solutions, while still considering the best solution found thus far. It helps maintain a balance between exploring diverse solutions and exploiting the best solution, ultimately improving the performance of the IBWOA algorithm in addressing the feature selection problem for Parkinson's disease detection. Fig. 4 shows the flowchart of IBWOA. Then, Fig. 5 illustrates a sample of the execution of the strategy defined to reduce the speed of convergence.

Algorithm 2: The strategy defined to reduce the speed of convergence in IBWOA

Input:
Initial \\ population of whales
Output:
Optimal \\ the best whale of the population

```
best_whale1= the best whale of population based on fitness function
for i=1:N
   B1=rand(0,1)
   if B1>=0.5
     best_whale2(1,i)=1- best_whale1(1,i)
  else
     best_whale2(1,i)=best_whale1(1,i)
  end if
end for
for i=1:(N/2)
  B2=rand(0,1)
  if B2>=0.5
     best_whale3(1,i)=1- best_whale1(1,i)
  else
     best_whale3(1,i)=best_whale1(1,i)
  end if
end for
for i=(N/2+1):N
  B3=rand(0,1)
  if B3>=0.5
     best_whale4(1,i)=1- best_whale1(1,i)
  else
     best_whale4(1,i)=best_whale1(1,i)
  end if
end for
Optimal=best(best_whale1, best_whale2, best_whale3, best_whale4)
```

5. Experimental results

The proposed hybrid intelligent system was implemented using MATLAB simulator software version 2017. The empirical experiment was conducted on AMD R7 3800XT with 8GB of RAM and Windows 10. The detailed parameter setting for IBWOA is presented in Table 4.

5.1. Dataset

This paper evaluates the results on the PD dataset taken from UCI Machine Learning Repository [47]. The total number of samples in this dataset is 195, and the total number of features is 22. Table 5 describes all the features of the PD dataset.

5.2. Evaluation metrics

Classification accuracy (ACC), error rate (ER), sensitivity, and specificity are used in this study as evaluation metrics. Evaluation metrics are defined based on Equations (23–26):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100(\%) \tag{23}$$

$$ER = \frac{FP + FN}{TP + TN + FP + FN} \times 100(\%) \tag{24}$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100(\%) \tag{25}$$

$$Specificity = \frac{TN}{TN + FP} \times 100(\%) \tag{26}$$

Where:

- ***TP***: *Actual class (patient) and predicted (patient)*
- ***TN***: *Actual class (healthy) and predicted (healthy)*
- ***FP***: *Actual class (healthy) and predicted (patient)*
- ***FN***: *Actual class (patient) and predicted (healthy)*

5.3. Configuration of classifiers

Table 6 gives the information of the basic classifiers. In the proposed ensemble learning strategy, each individual classifier is trained separately and makes predictions independently. The final decision is then made by a voting system that combines the predictions of the selected classifiers.

5.4. Experiment 1

In the first experiment, the performance of the basic classifiers was evaluated on the PD dataset with all features. Table 7 presents the classification accuracy of different basic classifiers. Table 8 then illustrates the ranking of the basic classifiers based on classification accuracy. The SVM classifier with a radial basis function (RBF) kernel and a gamma parameter value of 1 and the RF classifier with 50 trees achieved the highest detection accuracy of 92.8205%.

5.5. Experiment 2

In the second experiment, the performance of the ensemble classifier was evaluated on the PD dataset with all features. Table 9 shows the status of basic classifiers in the proposed ensemble classifier. Among all the expert systems in the classification pool, five expert systems were selected based on Algorithm 1, in the proposed ensemble classifier. As shown in Table 10, the expert systems where average accuracy is higher than or equal to the defined threshold level are selected in the proposed ensemble classifier. Fig. 6 presents the classification accuracy of the proposed ensemble classifier. The average accuracy of the proposed ensemble classifier is equal to 93.3333%, which shows the ensemble classifier has better performance compared to the best basic classifier.

In the context of the proposed ensemble classifier and considering that the odd individual classifiers participate in the final voting process according to equation 22, the voting result will indeed never be 0. As a result, the voting process will not be blocked, and the ensemble classifier will always produce a definitive output based on the collective decision of the individual base classifiers. This ensures that the proposed system can effectively classify instances without encountering a tie or stalemate during the voting process.

5.6. Experiment 3

In this experiment, the performance of the feature selection was evaluated on the PD dataset. Table 10 gives the details of the feature selection results based on BWOA. Next, Table 11 shows the features selected by BWOA in each fold. Fig. 7 displays the frequency of features selected by BWOA. The most important features selected by BWOA were F1, F3, F5, F7, F16, F17, F18, F19, and F22. The feature selection rate in the BWOA approach on the PD dataset was 40.9091%. Table 12 summarizes the details of the feature selection results based on IBWOA. After that, Table 13 gives the features selected by IBWOA in each fold. Fig. 8 portrays the frequency of features selected by IBWOA. The most important features selected by IBWOA were F1, F4, F5, F16, F17, F18, F21, and F22. The feature selection rate in the IBWOA approach on the PD dataset was 36.3636%.

To determine the most frequently selected features from a total of 10 simulation runs, various threshold levels were analyzed, including 6, 7, 8, 9, and 10. The simulation runs consisted of 5 runs of the metaheuristic algorithm using an S-shaped transfer function on 5-

folds, and 5 runs of the metaheuristic algorithm using an V-shaped transfer function on 5-folds. The results showed that a threshold level of 7 yielded the best performance in feature selection.

5.7. Comparison

Fig. 9 illustrates the performance of the feature selection approaches. Feature selection using BWOA based on the S-shaped transfer function achieved the average results of 94.87178%, 93.88506%, and 97.77778% regarding accuracy, sensitivity, and specificity, respectively. On the other hand, feature selection using BWOA based on the V-shaped transfer function achieved the average results of 95.89744%, 95.26436%, and 97.77778% regarding the above constructs. In addition, the results indicate that IBWOA based on the S-shaped transfer function achieved the average results of 96.41026%, 95.93102%, and 97.77778% regarding accuracy, sensitivity, and specificity, while based on the V-shaped transfer function, the average results were 96.92308%, 96.62068%, and 97.77778% regarding the above constructs. As revealed by the results, the best performance belongs to IBWOA based on the V-shaped transfer function. In the proposed intelligent diagnosis system, eight features, i.e., F1, F4, F5, F16, F17, F18, F21 and F22, were selected by IBWOA as the optimal features in the PD dataset.

Figs. 10 to 12 compare the performance of the proposed method with that of other existing methods in regard to accuracy, sensitivity, and specificity. These figures demonstrate that the proposed method achieved the highest accuracy and specificity among all. Although the proposed method ranked second regarding sensitivity, it exhibited a more balanced performance across all the three evaluation criteria. This balance is important because it indicates that the proposed method is capable of detecting both positive and negative cases accurately and, at the same time, maintaining a low false positive rate. Overall, these results suggest that the proposed method is a competitive approach for the task of PD detection. The high performance and balance of the proposed method could have a significant impact on healthcare applications, potentially leading to earlier and more accurate detection of PD.

6. Conclusions

This paper developed an intelligent system for PD diagnosis, called Improved Binary Whale Optimization Algorithm (IBWOA). It was found to have high potential to solve feature selection problems. In addition, an efficient ensemble classifier was presented to evaluate the fitness of the solutions in the proposed meta-heuristic algorithm. It was observed that the proposed intelligent diagnosis system achieved the highest classification accuracy of 96.9231% with 5-fold cross validation. IBWOA consists of two main phases, namely feature selection and ensemble learning, and each of these phases is independently effective in improving system performance. The improved meta-heuristic algorithm also succeeded in solving the feature selection problem for PD, with high accuracy.

The advantages of the proposed method are as follows:

- Improved Feature Selection: The use of an IBWOA for feature selection resulted in an optimal subset of features for PD detection, demonstrating its effectiveness in enhancing system performance.
- Enhanced Classification: The NELS was employed to handle the learning process, contributing to the high classification accuracy of 96.9231% achieved in the study.
- State-of-the-Art Performance: The proposed hybrid intelligent system demonstrated a 0.9231% improvement in detection accuracy compared to the best existing model in the current literature.
- Effective Preprocessing: The utilization of normalization as a preprocessing step ensured that the PD dataset was properly prepared for feature selection and classification, further improving the overall system's performance.

- Robust Optimization Algorithm: The comparison between IBWOA and the original BWOA revealed that IBWOA outperformed BWOA in solving the feature selection problem, indicating its robustness and superiority for this particular application.

The disadvantages of the proposed method are as follows:

- Parameter Tuning: As with most metaheuristic algorithms, the IBWOA requires proper tuning of its parameters to achieve optimal performance. This process can be time-consuming and may require expertise in the field to determine the best parameter configuration. Inappropriate tuning can lead to suboptimal results and may affect the overall performance of the model.
- Computational Load of Ensemble Classifier: The proposed model utilizes a NELS for classification. Ensemble learning techniques often require training multiple models and combining their predictions, which can result in a higher computational load compared to single-model approaches. This increased computational load may lead to longer processing times and demand more computational resources.

As for the limitations in this research on the IBWOA and NELS for PD detection, some potential limitations might include:

- Data constraints: The study relies on a specific PD dataset, which may limit the generalizability of the findings to other datasets or populations.
- Exploratory nature: As the proposed approach is relatively novel, further validation and comparison with alternative methods may be necessary to fully assess its effectiveness and robustness.

Suggested future work based on the IBWOA and NELS for PD detection could include:

- Evaluation on Additional Datasets: Test the proposed method on other PD datasets or even data from related neurological disorders to validate its generalizability and effectiveness in diverse scenarios.
- Exploring Alternative Ensemble Strategies: Experiment with other ensemble learning techniques, such as different combination methods, to identify potential performance improvements and learn more about the impact of various ensemble strategies.
- Investigating Computational Efficiency: Develop methods to address the high computational load of the ensemble classifier, such as parallelization or model optimization, to improve processing times and reduce resource requirements.
- Integration with Clinical Practice: Collaborate with healthcare professionals to assess the feasibility and potential impact of integrating the proposed method into clinical practice for improved Parkinson's disease detection and patient care.
- Expanding to Other Diseases or Applications: Extend the proposed method to other medical conditions or applications, such as the detection of Alzheimer's disease or other neurological disorders.
- Addressing Imbalanced Datasets: Explore strategies to handle imbalanced datasets, which are common in medical applications, such as over sampling, under sampling, or using specialized classification techniques.

References

[1] Ho, A.K., Iansek, R., Marigliani, C., et al. "Speech impairment in a large sample of patients with Parkinson's disease," Behavioural Neurology, 11(3), pp. 131-137, (1998). https://doi.org/10.1155/1999/327643.

[2] Hoffmeister, J.D., Kelm-Nelson, C.A., & Ciucci, M.R. "Quantification of brainstem norepinephrine relative to vocal impairment and anxiety in the Pink1-/-rat model of Parkinson disease," Behavioural Brain Research, 414, pp. 113514, (2021). https://doi.org/10.1016/j.bbr.2021.113514.

[3] Good, A., Earle, E., Vezer, E., et al. "Community choir improves vocal production measures in individuals living with Parkinson's disease," Journal of Voice, pp. 1-50, (2023). https://doi.org/10.1016/j.jvoice.2022.12.001.

[4] Mohammadi, A.G., Mehralian, P., Naseri, A., et al. "Parkinson's disease diagnosis: the effect of auto encoders on extracting features from vocal characteristics," Array, 11, pp. 1-50, (2021). https://doi.org/10.1016/j.array.2021.100079.
[5] Durán-Rosal, A.M., Camacho-Cañamón, J., Gutiérrez, P.A., et al. "Ordinal classification of the affectation level of 3D-images in Parkinson diseases," Scientific Reports, 11(1), pp. 1-13, (2021). https://doi.org/10.1038/s41598-021-86538-y.
[6] Balaji, E., Brindha, D., Elumalai, V.K., et al. "Automatic and non-invasive Parkinson’s disease diagnosis and severity rating using LSTM network," Applied Soft Computing, 108, pp. 1-50, (2021). https://doi.org/10.1016/j.asoc.2021.107463.
[7] Polat, K. & Nour, M. "Parkinson disease classification using one against all based data sampling with the acoustic features from the speech signals," Medical Hypotheses, 140, pp. 1-50, (2020). https://doi.org/10.1016/j.mehy.2020.109678.
[8] Alam, T.M., Iqbal, M.A., Ali, Y., et al. "A model for early prediction of diabetes," Informatics in Medicine Unlocked, 16, pp. 1-50, (2019). https://doi.org/10.1016/j.imu.2019.100204.
[9] Li, J., Luo, T., Zhang, B., et al. "IMOABC: An efficient multi-objective filter–wrapper hybrid approach for high-dimensional feature selection," Journal of King Saud University-Computer and Information Sciences, 36(9), pp. 1-50, (2024). https://doi.org/10.1016/j.jksuci.2024.102205.
[10] Ullah, Z., Naqvi, S.R., Farooq, W., et al. "A comparative study of machine learning methods for bio-oil yield prediction–A genetic algorithm-based features selection," Bioresource Technology, 335, pp. 1-50, (2021). https://doi.org/10.1016/j.biortech.2021.125292.
[11] Abbasi, M.S., Al-Sahaf, H., Mansoori, M., et al. "Behavior-based ransomware classification: A particle swarm optimization wrapper-based approach for feature selection," Applied Soft Computing, 121, pp. 1-50, (2022). https://doi.org/10.1016/j.asoc.2022.108744.
[12] Wang, Y., Ran, S., & Wang, G.G. "Role-oriented binary grey wolf optimizer using foraging-following and Lévy flight for feature selection," Applied Mathematical Modelling, 126, pp. 310-326, (2024). https://doi.org/10.1016/j.apm.2023.08.043.
[13] Arora, S. & Anand, P. "Binary butterfly optimization approaches for feature selection," Expert Systems with Applications, 116, pp. 147-160, (2019). https://doi.org/10.1016/j.eswa.2018.08.051.
[14] Wan, Y., Wang, M., Ye, Z., et al. "A feature selection method based on modified binary coded ant colony optimization algorithm," Applied Soft Computing, 49, pp. 248-258, (2016). https://doi.org/10.1016/j.asoc.2016.07.041.
[15] Hichem, H., Elkamel, M., Rafik, M., et al. "A new binary grasshopper optimization algorithm for feature selection problem," Journal of King Saud University-Computer and Information Sciences, 34(2), pp. 316-328, (2022). https://doi.org/10.1016/j.jksuci.2019.11.007.
[16] Anuradha, K., Krishna, M.V., & Mallik, B. "Bio inspired boolean artificial bee colony based feature selection algorithm for sentiment classification," Measurement: Sensors, 32, pp. 1-50, (2024). https://doi.org/10.1016/j.measen.2024.101034.
[17] Hussien, A.G., Hassanien, A.E., Houssein, E.H., et al. "S-shaped binary whale optimization algorithm for feature selection," In Recent Trends in Signal and Image Processing, pp. 79-87, (2019). https://doi.org/10.1007/978-981-10-8863-6_9.
[18] Li, Y., Yang, L., Wang, P., et al. "Classification of Parkinson's disease by decision tree based instance selection and ensemble learning algorithms," Journal of Medical Imaging and Health Informatics, 7(2), pp. 444-452, (2017). https://doi.org/10.1166/jmihi.2017.2033.
[19] Hosni, M., Abnane, I., Idri, A., et al. "Reviewing ensemble classification methods in breast cancer," Computer Methods and Programs in Biomedicine, 177, pp. 89-112, (2019). https://doi.org/10.1016/j.cmpb.2019.05.019.
[20] Al Essa, H.A., & Bhaya, W.S. "Ensemble learning classifiers hybrid feature selection for enhancing performance of intrusion detection system," Bulletin of Electrical Engineering and Informatics, 13(1), pp. 665-676, (2024). https://doi.org/10.11591/eei.v13i1.5844.
[21] Krawczyk, B., Woźniak, M., & Schaefer, G. "Cost-sensitive decision tree ensembles for effective imbalanced classification," Applied Soft Computing, 14, pp. 554-562, (2014). https://doi.org/10.1016/j.asoc.2013.08.014.
[22] Özçift, A. "Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis," Computers in Biology and Medicine, 41(5), pp. 265-271, (2011). https://doi.org/10.1016/j.compbiomed.2011.03.001.
[23] Zhang, Y., Cao, G., Wang, B., et al. "A novel ensemble method for k-nearest neighbor," Pattern Recognition, 85, pp. 13-25, (2019). https://doi.org/10.1016/j.patcog.2018.08.003.
[24] Abdar, M., & Makarenkov, V. "CWV-BANN-SVM ensemble learning classifier for an accurate diagnosis of breast cancer," Measurement, 146, pp. 557-570, (2019). https://doi.org/10.1016/j.measurement.2019.05.022.
[25] Attallah, O. "Skin cancer classification leveraging multi-directional compact convolutional neural network ensembles and gabor wavelets," Scientific Reports, 14(1), pp. 1-50, (2024). https://doi.org/10.1038/s41598-024-69954-8.

[26] Zhang, H., Petitjean, F., & Buntine, W. "Bayesian network classifiers using ensembles and smoothing," Knowledge and Information Systems, 62, pp. 3457-3480, (2020). https://doi.org/10.1007/s10115-020-01458-z.
[27] Åström, F., & Koker, R. "A parallel neural network approach to prediction of Parkinson's disease," Expert Systems with Applications, 38(10), pp. 12470-12474, (2011). https://doi.org/10.1016/j.eswa.2011.04.028.
[28] Ozcift, A. "SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease," Journal of Medical Systems, 36, pp. 2141-2147, (2012). https://doi.org/10.1007/s10916-011-9678-1.
[29] Gök, M. "An ensemble of k-nearest neighbours algorithm for detection of Parkinson's disease," International Journal of Systems Science, 46(6), pp. 1108-1112, (2015). https://doi.org/10.1080/00207721.2013.809613.
[30] Pasha, A., & Latha, P.H. "Bio-inspired dimensionality reduction for Parkinson's disease (PD) classification," Health Information Science and Systems, 8, pp. 1-22, (2020). https://doi.org/10.1007/s13755-020-00104-w.
[31] Tunc, H.C., Sakar, C.O., Apaydin, H., et al. "Estimation of Parkinson's disease severity using speech features and extreme gradient boosting," Medical & Biological Engineering & Computing, 58, pp. 2757-2773, (2020). https://doi.org/10.1007/s11517-020-02250-5.
[32] Soumaya, Z., Taoufiq, B.D., Benayad, N., et al. "The detection of Parkinson disease using the genetic algorithm and SVM classifier," Applied Acoustics, 171, pp. 1-50, (2021). https://doi.org/10.1016/j.apacoust.2020.107528.
[33] Sonawane, B., & Sharma, P. "Speech-based solution to Parkinson's disease management," Multimedia Tools and Applications, 80(19), pp. 29437-29451, (2021). https://doi.org/10.1007/s11042-021-11061-1.
[34] Al-Fatlawi, A.H., Jabardi, M.H., & Ling, S.H. "Efficient diagnosis system for Parkinson's disease using deep belief network," In 2016 IEEE Congress on Evolutionary Computation, pp. 1324-1330, (2016). https://doi.org/10.1109/CEC.2016.7743941.
[35] Alweshah, M., Alkhalaileh, S., Al-Betar, M.A., et al. "Coronavirus herd immunity optimizer with greedy crossover for feature selection in medical diagnosis," Knowledge-Based Systems, pp. 1-50, (2022). https://doi.org/10.1016/j.knosys.2021.107629.
[36] Ali, L., Chakraborty, C., He, Z., et al. "A novel sample and feature dependent ensemble approach for Parkinson's disease detection," Neural Computing and Applications, pp. 1-50, (2022). https://doi.org/10.1007/s00521-022-07046-2.
[37] Govindu, A., & Palwe, S. "Early detection of Parkinson's disease using machine learning," Procedia Computer Science, 218, pp. 249-261, (2023). https://doi.org/10.1016/j.procs.2023.01.007.
[38] Sathiya, T., Reenadevi, R., & Sathiyabhama, B. "Random forest classifier based detection of Parkinson's disease," Annals of the Romanian Society for Cell Biology, pp. 2980-2987, (2021). http://annalsofrscb.ro/index.php/journal/article/view/4912.
[39] Mirjalili, S., & Lewis, A. "The whale optimization algorithm," Advances in Engineering Software, 95, pp. 51-67, (2016). https://doi.org/10.1016/j.advengsoft.2016.01.008.
[40] Beheshti, Z. "A fuzzy transfer function based on the behavior of meta-heuristic algorithm and its application for high-dimensional feature selection problems," Knowledge-Based Systems, 284, pp. 1-50, (2024). https://doi.org/10.1016/j.knosys.2023.111191.
[41] Ghosh, K.K., Guha, R., Bera, S.K., et al. "S-shaped versus V-shaped transfer functions for binary Manta ray foraging optimization in feature selection problem," Neural Computing and Applications, 33(17), pp. 11027-11041, (2021). https://doi.org/10.1007/s00521-020-05560-9.
[42] Chandra, M.A., & Bedi, S.S. "Survey on SVM and their application in image classification," International Journal of Information Technology, 13, pp. 1-11, (2021). https://doi.org/10.1007/s41870-017-0080-1.
[43] Peng, N., Zhang, Y., & Zhao, Y. "A SVM-KNN method for quasar-star classification," Science China Physics, Mechanics and Astronomy, 56, pp. 1227-1234, (2013). https://doi.org/10.1007/s11433-013-5083-8.
[44] Verikas, A., Gelzinis, A., & Bacauskiene, M. "Mining data with random forests: A survey and results of new tests," Pattern Recognition, 44(2), pp. 330-349, (2011). https://doi.org/10.1016/j.patcog.2010.08.011.
[45] Sameh, S.M., Moustafa, H.E.D., AbdelHay, E.H., et al. "An effective chaotic maps image encryption based on metaheuristic optimizers," The Journal of Supercomputing, 80(1), pp. 141-201, (2024). https://doi.org/10.1007/s11227-023-05413-x.
[46] El Ghouate, N., Bencherqui, A., Mansouri, H., et al. "Improving the Kepler optimization algorithm with chaotic maps: comprehensive performance evaluation and engineering applications," Artificial Intelligence Review, 57(11), pp. 1-50, (2024). https://doi.org/10.1007/s10462-024-10857-5.
[47] Machine Learning Repository, Center for Machine Learning and Intelligent Systems, https://archive.ics.uci.edu/ml/datasets/parkinsons. https://doi.org/10.24432/C59C74.

Tables and Figures List

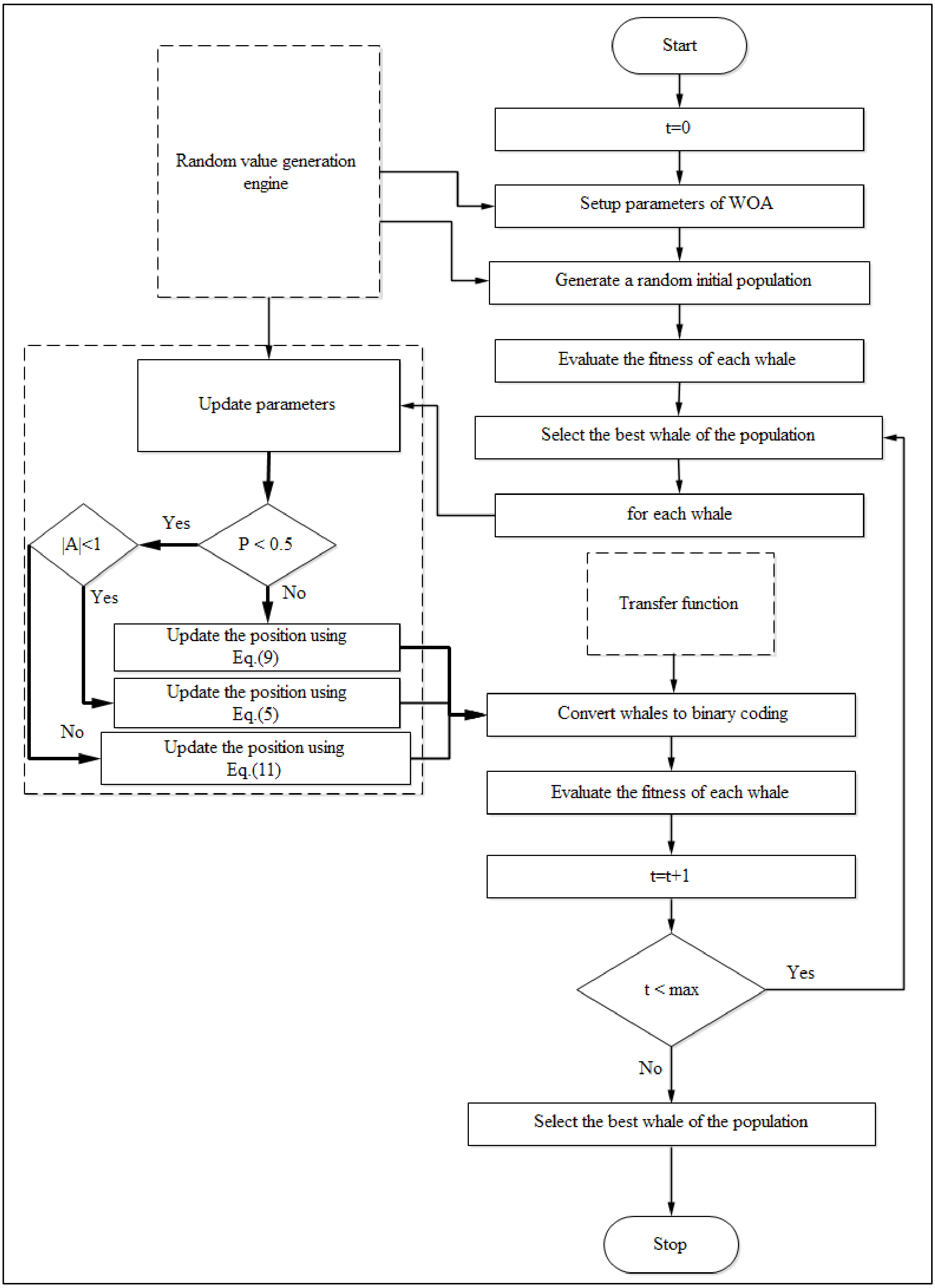

Fig. 1: The flowchart of BWOA

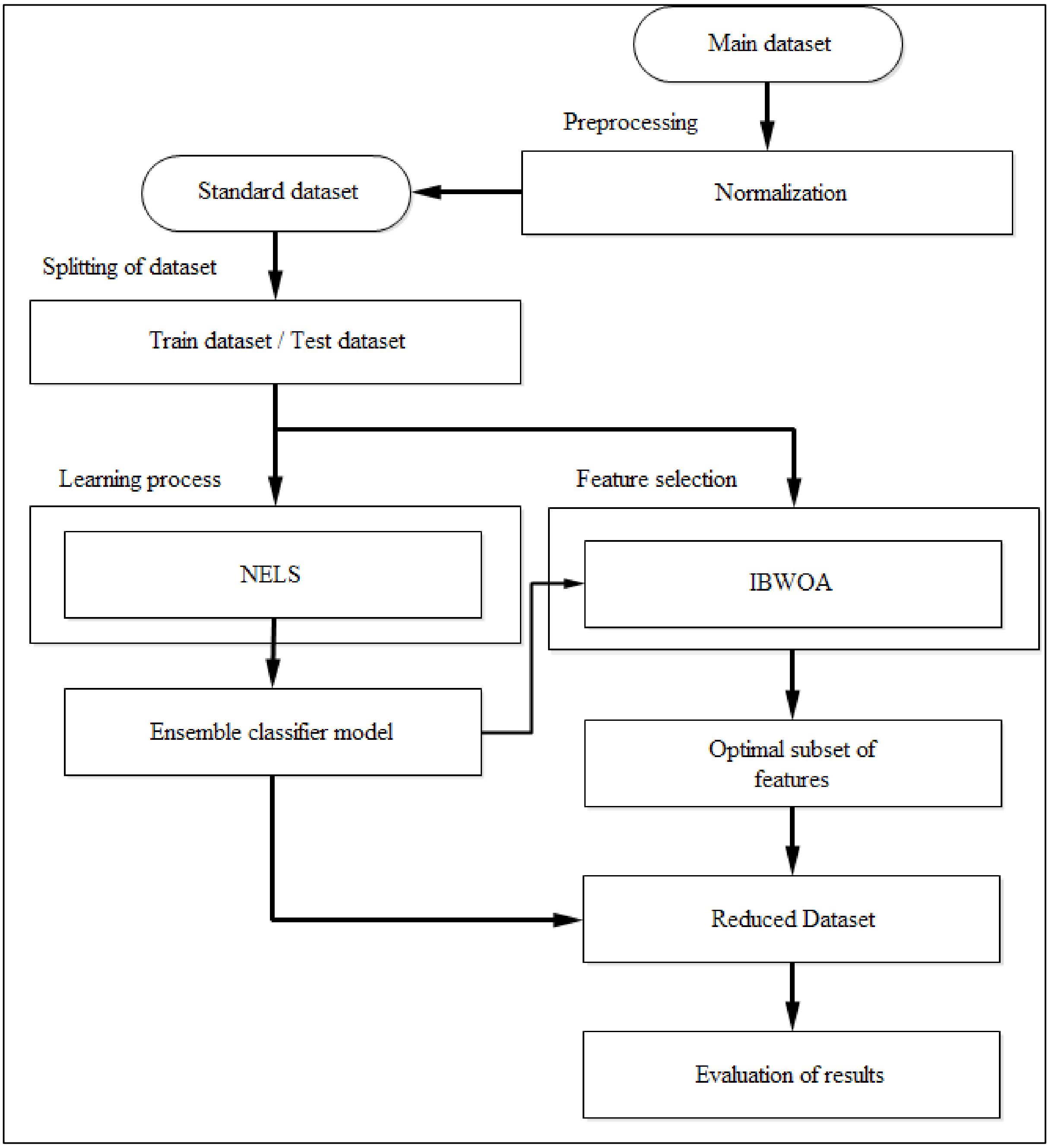

Fig. 2: The overall framework of the proposed hybrid intelligent system

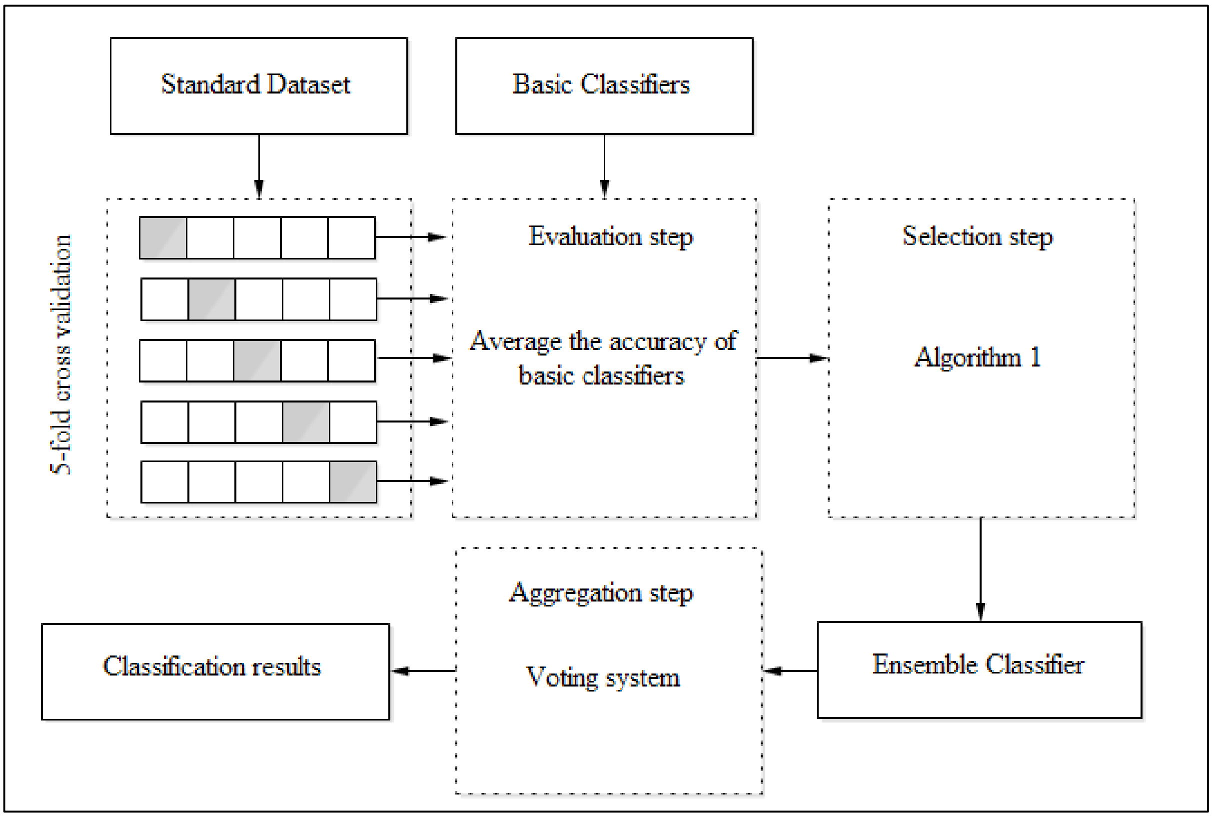

Fig. 3: The scheme of NELS

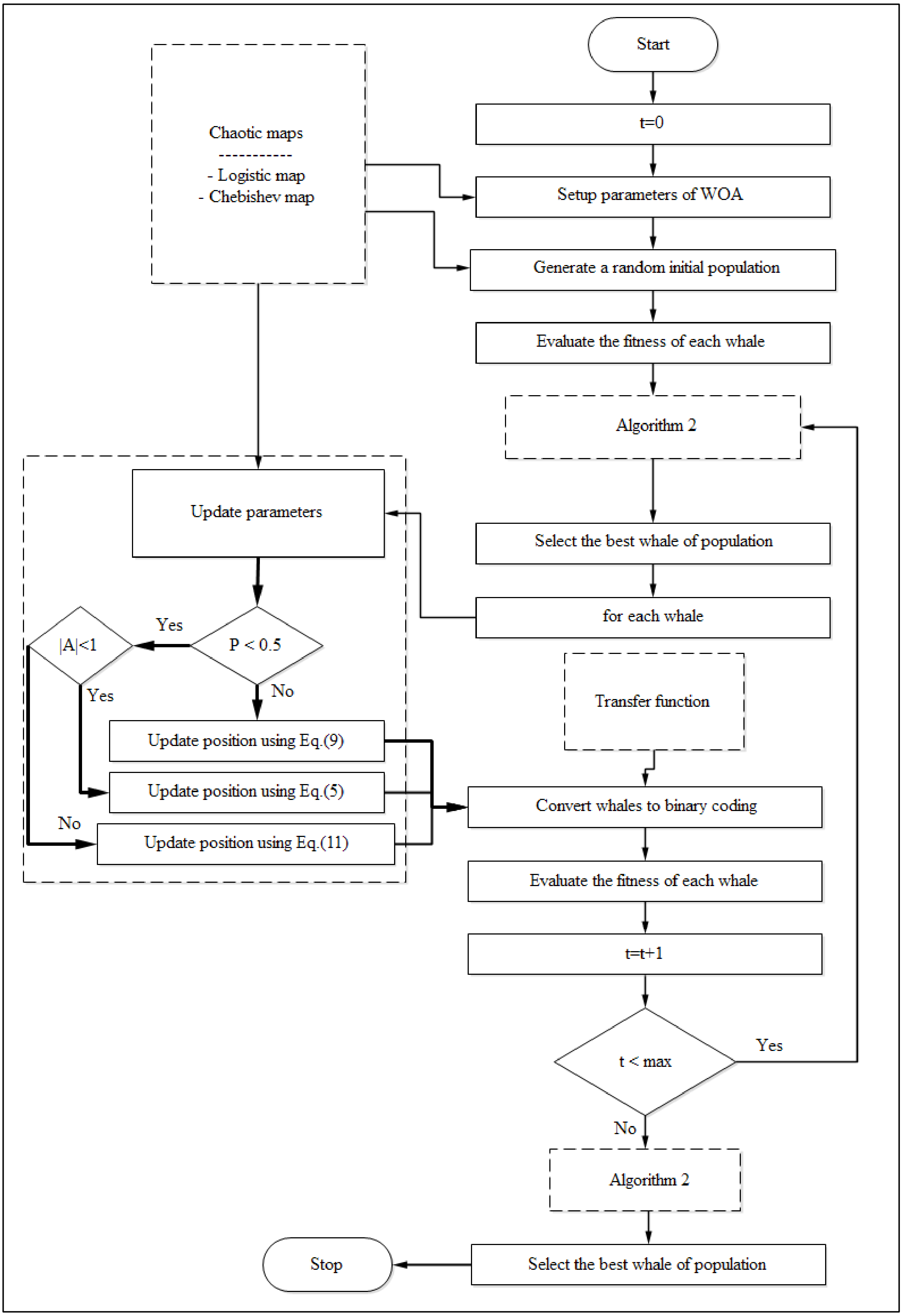

Fig. 4: The flowchart of IBWO

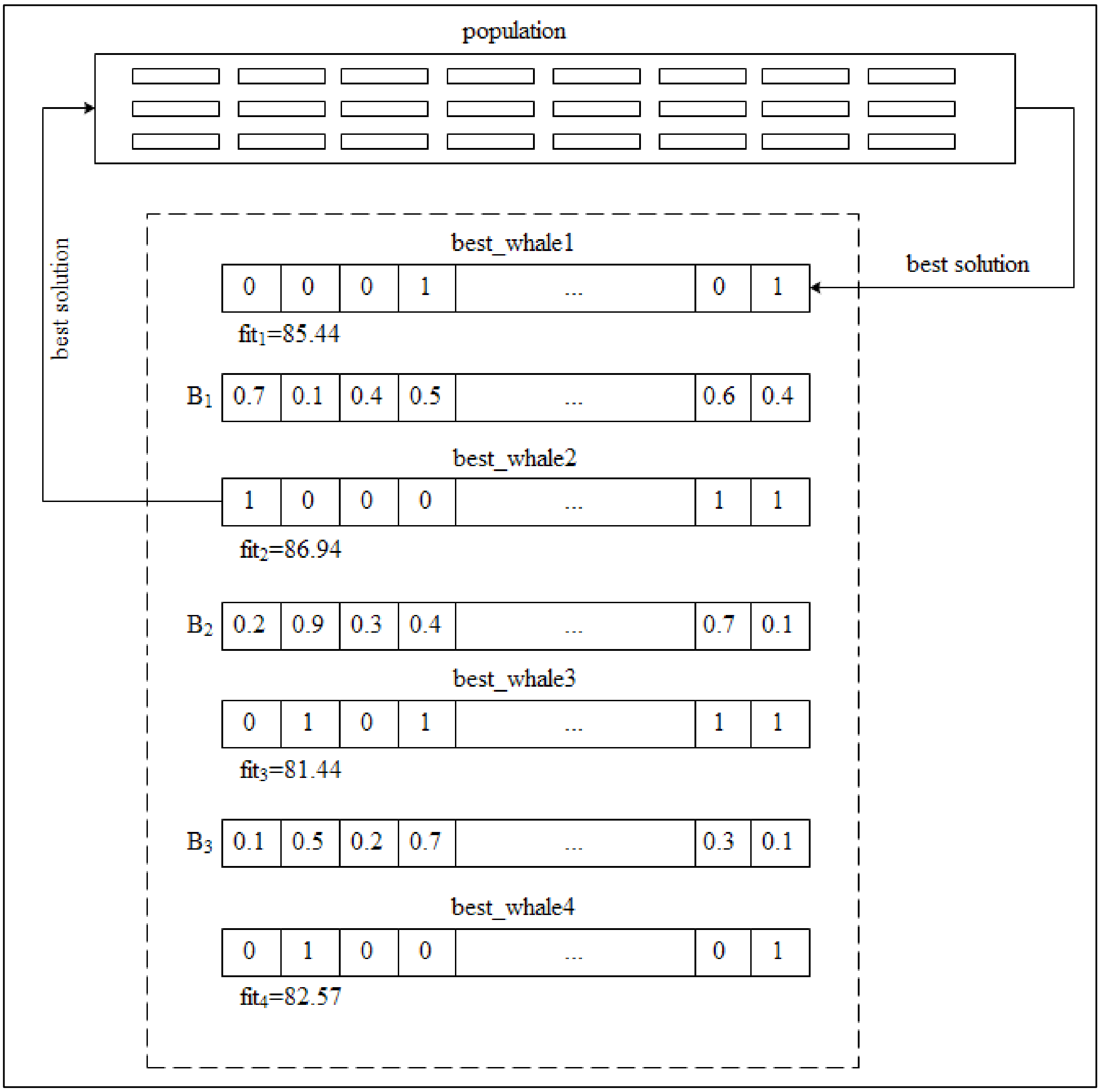

Fig. 5: A sample of the execution of the strategy defined to reduce the speed of convergence

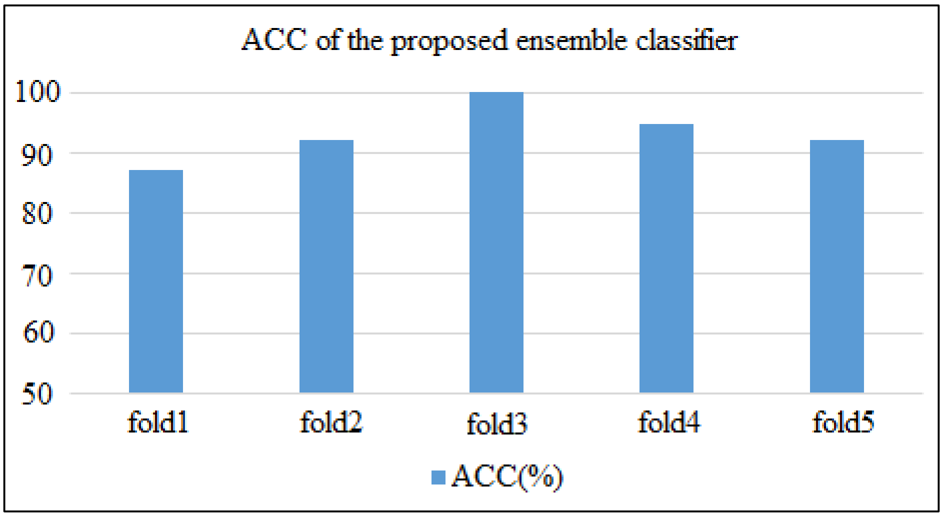

Fig. 6: Classification accuracy of the proposed ensemble classifier

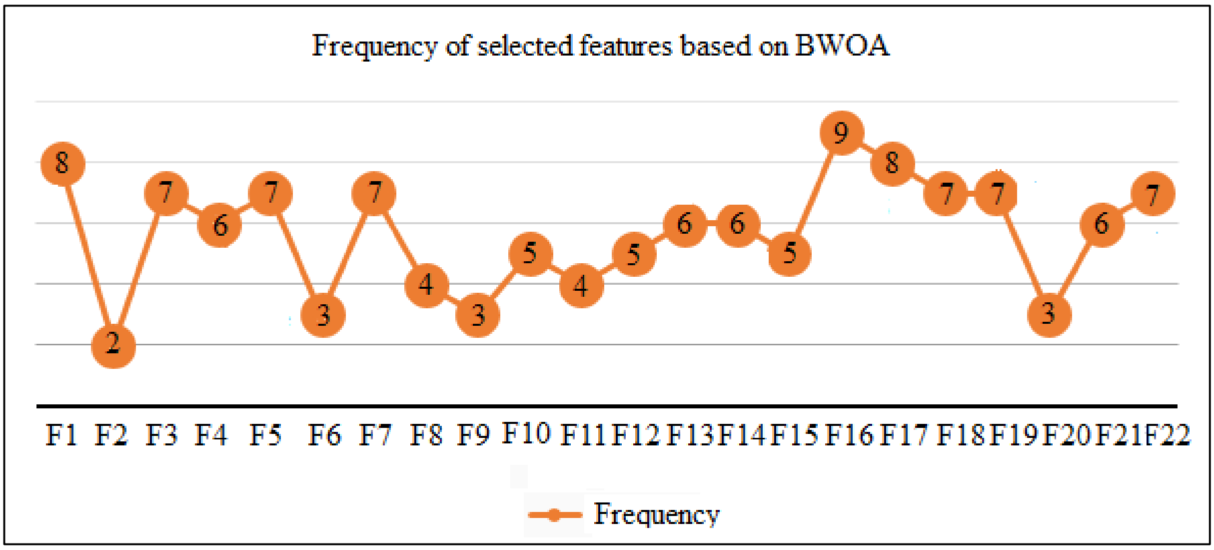

Fig. 7: Frequency of features selected by BWOA

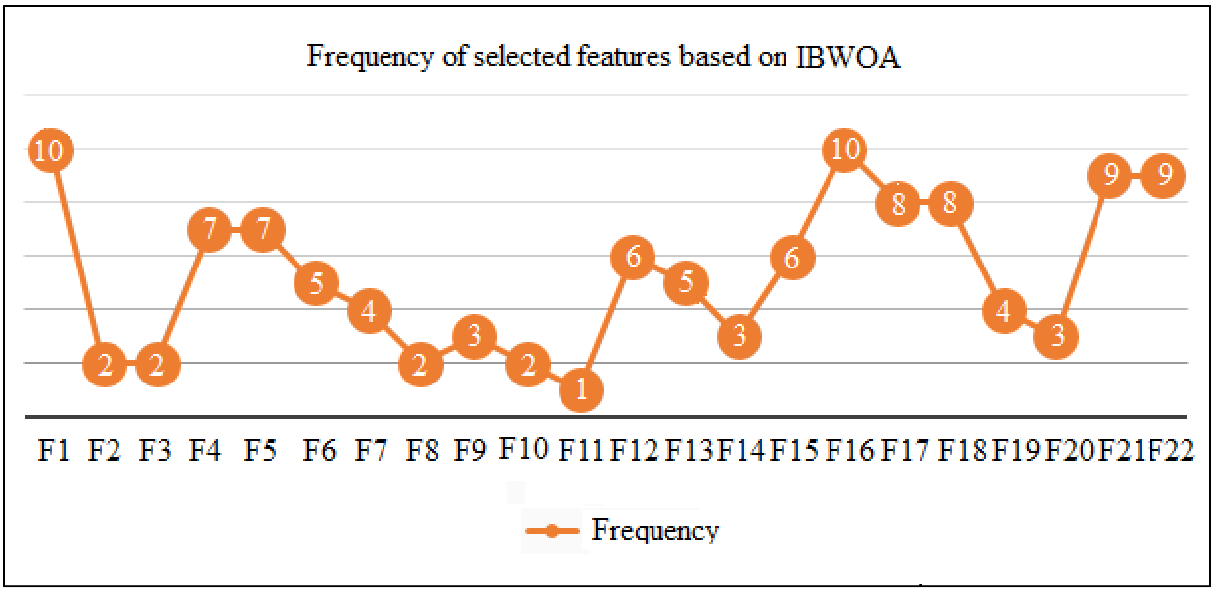

Fig. 8: Frequency of features selected by IBWOA

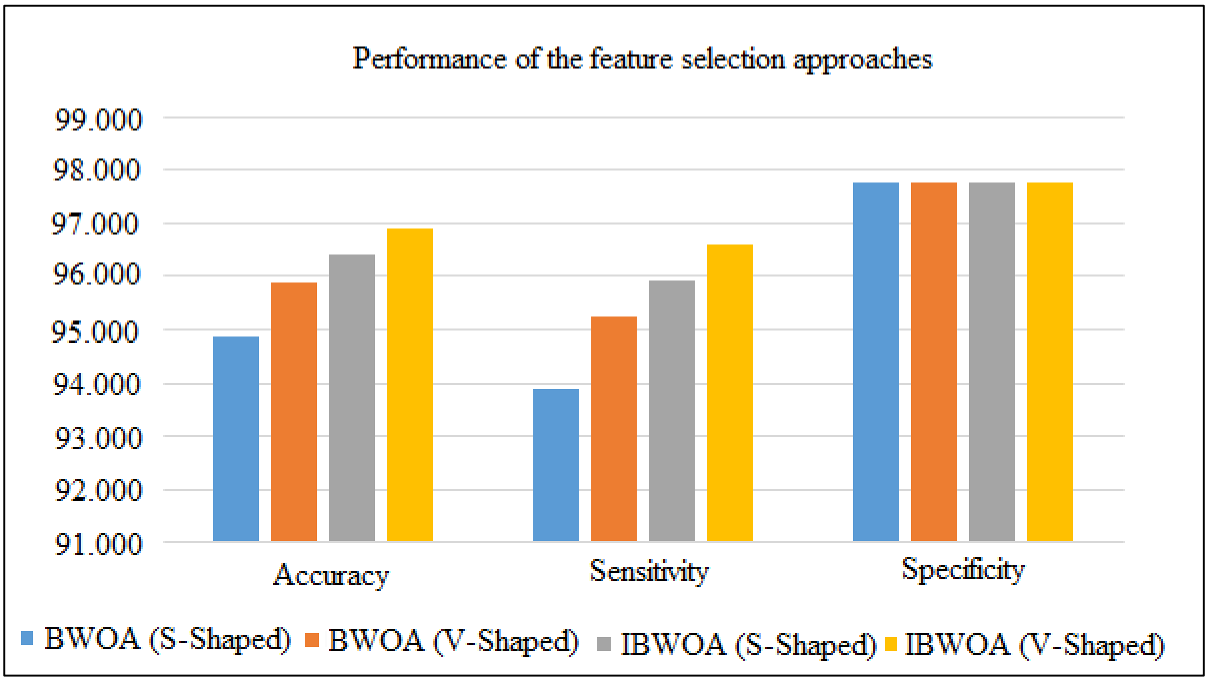

Fig. 9: Performance of the feature selection approaches

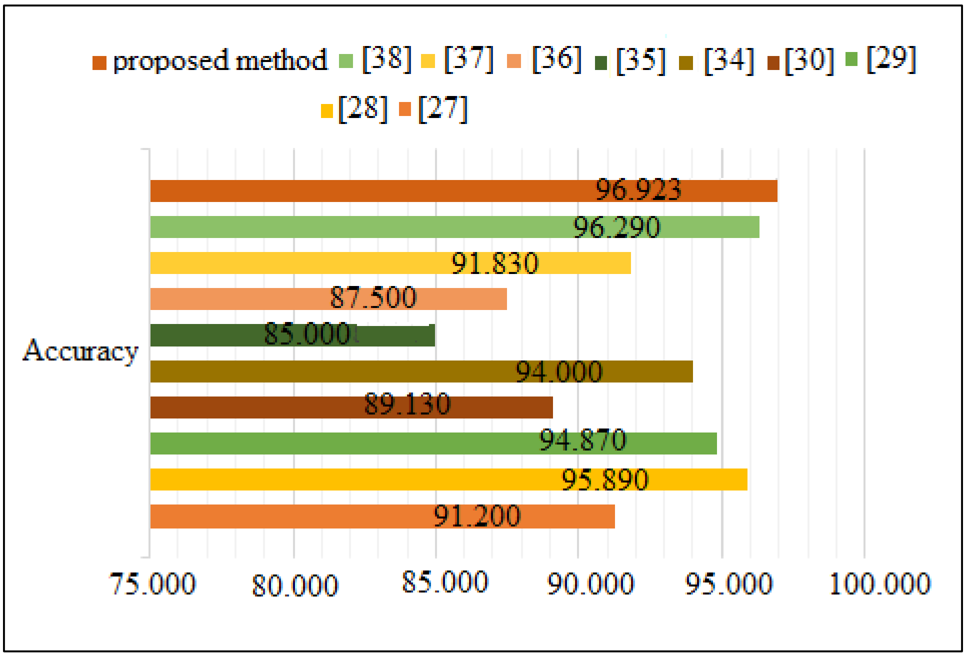

Fig. 10: Comparison of the proposed method with the other methods based on accuracy

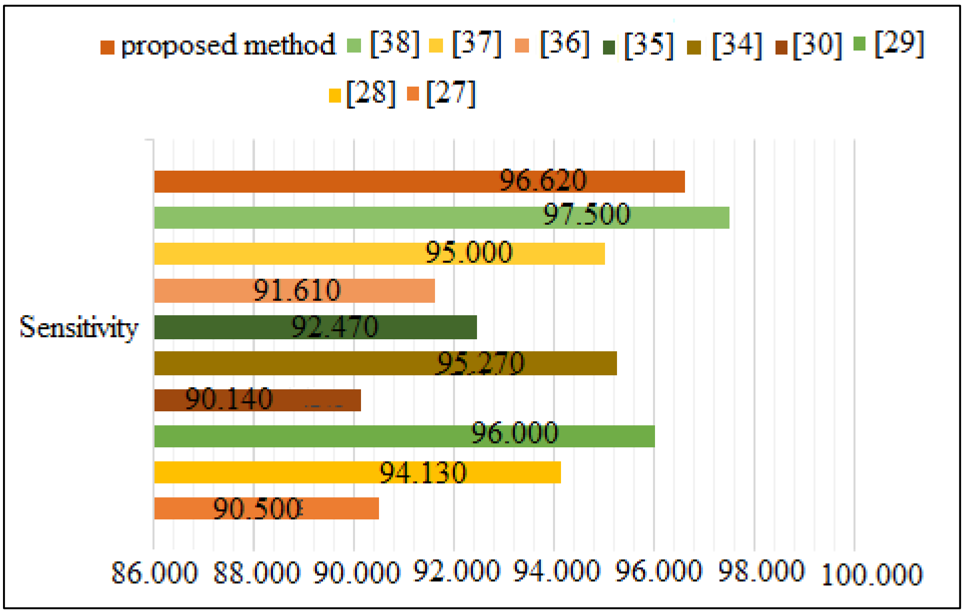

Fig. 11: Comparison of the proposed method with the other methods based on sensitivity

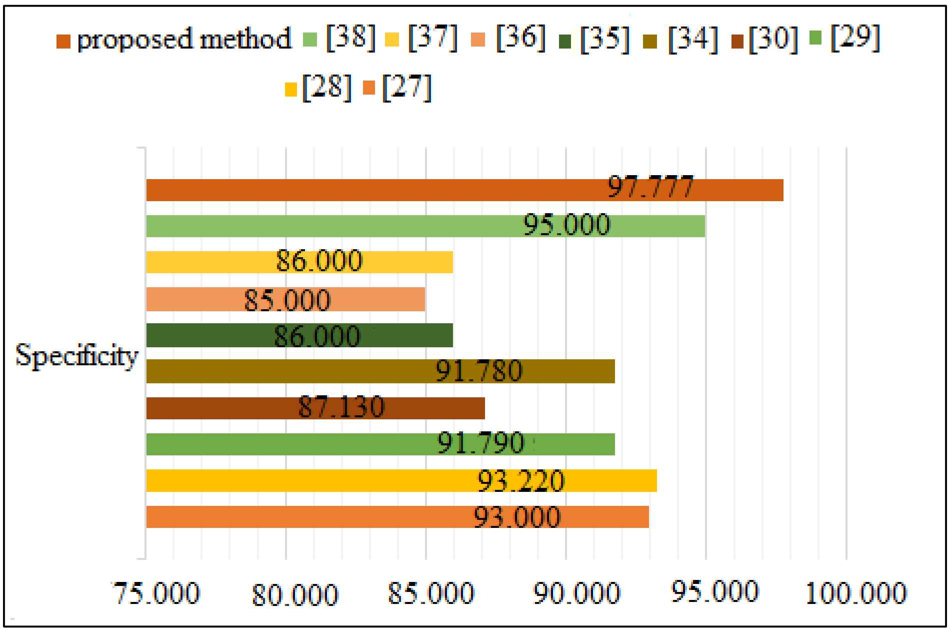

Fig. 12: Comparison of the proposed method with the other methods based on specificity

Table 1: The comparison of the models already proposed in the literature for the PD diagnosis

Source	Preprocessing	Feature selection	Learning process	Dataset	ACC
Ref [27]	×	×	Parallel neural network	UCI	91.20
Ref [28]	×	Linear SVM	Rotation forest ensemble classifier	UCI	95.89
Ref [29]	×	Linear SVM	KNN ensemble classifier	UCI	95.90
Ref [30]	Normalization	GA/BPSO	Basic classifiers	UCI	90.70
Ref [31]	Signal processing	Heuristic algorithm	Tree-based boosting algorithm	Private	96.13
Ref [32]	Signal processing	GA	SVM	private	91.18

Ref [33]	Signal processing	×	Convolutional neural network	private	77.48
Ref [34]	×	×	Deep belief network	UCI	94.00
Ref [35]	×	CHIO	KNN	UCI	88.00
Ref [36]	×	DNN	Ensemble learning strategy	UCI	93.75
Ref [37]	Normalization	PCA	SVM/KNN/RF/logistic regression	UCI	91.83
Ref [38]	×	RFE	RF	UCI	96.00

Table 2: The characteristics of the WOA parameters

NO	Parameters	State	Interval	Characteristic
1	a	Value number	(0,2)	Linear reduction in each iteration
2	P	Random number	(0,1)	Control parameter (Randomly)
3	$\bar{r}$	Random vector	(0,1)	Randomly
4	L	Random number	(-1,1)	Randomly
5	B	Constant number	---	Logarithmic spiral movement of whales
6	LB_i	Real number	---	The lowest boundary
7	UB_i	Real number	---	The highest boundary
8	ps	Integer number	---	Population size
9	max	Integer number	---	Actual number of iterations
10	N	Integer number	---	Dimensions of the problem
11	t	Integer number	---	Current iteration

Table 3: The characteristics of the chaotic maps used in IBWOA

NO	Chaotic map	Purpose	RN_0	Control parameter	Iteration
1	Logistic map	Production of the P parameter	0.51	$\gamma = 3 \cdot 9$	$ps \times max$
2	Logistic map	Production of the $\bar{r}$ parameter	0.52	$\gamma = 3 \cdot 8$	$N \times ps \times max$
3	Logistic map	Population production	0.53	$\gamma = 3 \cdot 7$	$N \times ps$
4	Chebishev map	Production of the L parameter	0.54	$\mu = 2$	$ps \times max$

Table 4: Parameter setup for IBWOA

NO	Parameter	Value
1	ps	50
2	max	50
3	N	22
4	B	2

Table 5: Features of the PD dataset

NO	Index	Attribute
1	F1	MDVP:Fo (Hz)
2	F2	MDVP:Fhi (Hz)
3	F3	MDVP:Flo (Hz)
4	F4	MDVP:Jitter (%)
5	F5	MDVP:Jitter (Abs)
6	F6	MDVP:RAP
7	F7	MDVP:PPQ
8	F8	Jitter:DDP
9	F9	MDVP:Shimmer
10	F10	MDVP:Shimmer (dB)
11	F11	Shimmer:APQ3
12	F12	Shimmer:APQ5
13	F13	MDVP:APQ

14	F14	Shimmer:DDA
15	F15	NHR
16	F16	HNR
17	F17	RPDE
18	F18	D2
19	F19	DFA
20	F20	Spread1
21	F21	Spread2
22	F22	PPE

Table 6: Information of the basic classifiers

No	Name	Classifier	Characteristics	Parameter
1	CL_1	SVM	RBF Kernel	gamma =1
2	CL_2	SVM	RBF Kernel	gamma =2
3	CL_3	SVM	RBF Kernel	gamma =3
4	CL_4	SVM	RBF Kernel	gamma =4
5	CL_5	SVM	RBF Kernel	gamma =5
6	CL_6	KNN	Number of neighbors	K =3
7	CL_7	KNN	Number of neighbors	K =5
8	CL_8	KNN	Number of neighbors	K =7
9	CL_9	KNN	Number of neighbors	K =9
10	CL_{10}	KNN	Number of neighbors	K =11
11	CL_{11}	RF	Number of trees	Ntree =50
12	CL_{12}	RF	Number of trees	Ntree =100
13	CL_{13}	RF	Number of trees	Ntree =150
14	CL_{14}	RF	Number of trees	Ntree =200
15	CL_{15}	RF	Number of trees	Ntree =250

Table 7: Classification accuracy of the basic classifiers

Classifier	fold 1	fold 2	fold 3	fold 4	fold 5
CL_1	87.1795	89.7436	97.4359	94.8718	94.8718
CL_2	87.1795	89.7436	94.8718	84.6154	87.1795
CL_3	84.6154	84.6154	92.3077	84.6154	84.6154
CL_4	82.0513	82.0513	87.1795	79.4872	79.4872
CL_5	82.0513	82.0513	82.0513	79.4872	79.4872
CL_6	76.9231	76.9231	89.7436	87.1795	84.6154
CL_7	79.4872	84.6154	87.1795	84.6154	84.6154
CL_8	82.0513	84.6154	92.3077	82.0513	79.4872
CL_9	79.4872	82.0513	89.7436	82.0513	84.6154
CL_{10}	76.9231	79.4872	92.3077	79.4872	84.6154
CL_{11}	87.1795	89.7436	100	94.8718	92.3077
CL_{12}	84.6154	89.7436	100	89.7436	92.3077
CL_{13}	87.1795	89.7436	100	89.7436	92.3077
CL_{14}	82.0513	89.7436	100	94.8718	92.3077
CL_{15}	82.0513	89.7436	97.4359	92.3077	92.3077

Table 8: Ranking of the basic classifiers based on the classification accuracy

Classifier	AVE of ACC	Rank
CL_1	92.8205	1
CL_2	88.7180	7
CL_3	86.1539	8
CL_4	82.0513	14
CL_5	81.0257	15
CL_6	83.0769	12
CL_7	84.1026	9
CL_8	84.1026	10
CL_9	83.5898	11
CL_{10}	82.5641	13

CL_{11}	92.8205	2
CL_{12}	91.2821	5
CL_{13}	91.7949	3
CL_{14}	91.7949	4
CL_{15}	90.7692	6

Table 9: Status of basic classifiers in the proposed ensemble classifier

Classifier	ACC	T	Status
CL_1	92.8205	90.7692	✓
CL_2	88.7180		×
CL_3	86.1539		×
CL_4	82.0513		×
CL_5	81.0257		×
CL_6	83.0769		×
CL_7	84.1026		×
CL_8	84.1026		×
CL_9	83.5898		×
CL_{10}	82.5641		×
CL_{11}	92.8205		✓
CL_{12}	91.2821		✓
CL_{13}	91.7949		✓
CL_{14}	91.7949		✓
CL_{15}	90.7692		×

Table 10: Feature selection results based on BWOA

fold	Transfer function	ACC (%)	ER (%)	Sensitivity (%)	Specificity (%)	No. of selected features
# 1	S-shaped	89.7435	10.2565	89.6552	88.8889	13
	V-shaped	92.3077	7.6923	93.1034	88.8889	13
# 2	S-shaped	92.3077	7.6923	90	100	12
	V-shaped	92.3077	7.6923	90	100	11
# 3	S-shaped	100	0	100	100	12
	V-shaped	100	0	100	100	13
# 4	S-shaped	97.4359	2.5641	96.6667	100	14
	V-shaped	97.4359	2.5641	96.6667	100	14
# 5	S-shaped	94.8718	5.1282	93.1034	100	11
	V-shaped	97.4359	2.5641	96.5517	100	12

Table 11: Features selected by BWOA

Index	BWOA based on S-Shaped					BWOA based on V-Shaped				
	fold 1	fold 2	fold 3	fold 4	fold 5	fold 1	fold 2	fold 3	fold 4	fold 5
F1	✓	✓	✓	×	✓	✓	✓	✓	×	✓
F2	×	×	×	×	✓	×	×	×	✓	×
F3	✓	✓	×	✓	×	✓	✓	✓	✓	×
F4	✓	×	✓	✓	×	×	×	✓	✓	✓
F5	×	✓	✓	✓	×	✓	✓	✓	×	✓
F6	×	×	×	✓	✓	×	×	×	×	✓
F7	✓	×	✓	✓	✓	×	✓	×	✓	✓
F8	×	✓	×	×	✓	✓	×	✓	×	×
F9	×	×	×	×	✓	×	×	×	✓	✓
F10	×	✓	×	✓	×	✓	×	✓	✓	×
F11	×	✓	×	✓	×	×	✓	✓	×	×
F12	✓	×	✓	×	×	✓	×	✓	✓	×
F13	✓	×	✓	✓	×	×	✓	✓	✓	×
F14	✓	×	✓	✓	×	✓	×	✓	✓	×
F15	×	×	✓	✓	✓	×	✓	×	✓	×
F16	✓	✓	✓	×	✓	✓	✓	✓	✓	✓
F17	✓	✓	×	✓	✓	✓	✓	×	✓	✓

F18	✓	✓	×	✓	✓	✓	✓	×	×	✓
F19	✓	✓	×	✓	×	✓	✓	×	✓	✓
F20	×	✓	✓	×	×	×	×	×	×	✓
F21	✓	✓	✓	×	×	✓	×	✓	✓	×
F22	✓	×	✓	✓	✓	✓	×	✓	×	✓

Table 12: The feature selection results based on IBWOA

fold	transfer function	ACC (%)	ER (%)	Sensitivity (%)	Specificity (%)	No. of selected features
# 1	S-shaped	92.3077	7.6923	93.1034	88.8889	13
	V-shaped	94.8718	5.1282	96.5517	88.8889	12
# 2	S-shaped	94.8718	5.1282	93.3333	100	11
	V-shaped	94.8718	5.1282	93.3333	100	11
# 3	S-shaped	100	0	100	100	10
	V-shaped	100	0	100	100	12
# 4	S-shaped	97.4359	2.5641	96.6667	100	12
	V-shaped	97.4359	2.5641	96.6667	100	13
# 5	S-shaped	97.4359	2.5641	96.5517	100	11
	V-shaped	97.4359	2.5641	96.5517	100	11

Table 13: Features selected by IBWOA

Index	IBWOA based on S-Shaped					IBWOA based on V-Shaped				
	fold 1	fold 2	fold 3	fold 4	fold 5	fold 1	fold 2	fold 3	fold 4	fold 5
F1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
F2	×	×	×	×	✓	×	✓	×	×	×
F3	×	✓	×	×	×	×	×	×	✓	×
F4	×	✓	✓	✓	✓	✓	×	✓	✓	×
F5	✓	✓	×	✓	✓	✓	×	✓	×	✓
F6	×	✓	✓	×	✓	×	✓	×	✓	×
F7	✓	×	✓	×	✓	×	✓	×	×	×
F8	×	×	×	×	×	✓	×	✓	×	×
F9	✓	×	×	×	×	×	✓	×	×	✓
F10	×	×	×	×	×	×	×	✓	✓	×
F11	×	×	×	×	×	×	×	×	×	✓
F12	×	×	✓	✓	×	✓	✓	✓	✓	×
F13	✓	×	×	✓	×	✓	×	×	✓	✓
F14	×	✓	×	×	×	✓	×	✓	×	×
F15	✓	×	×	✓	×	✓	✓	×	✓	✓
F16	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
F17	✓	✓	✓	✓	✓	×	✓	✓	×	✓
F18	✓	✓	✓	✓	✓	×	×	✓	✓	✓
F19	✓	×	×	✓	×	×	×	✓	✓	×
F20	✓	×	×	×	✓	✓	×	×	×	×
F21	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
F22	✓	✓	✓	✓	✓	✓	✓	×	✓	✓

Biographies:

Seyyed Ahmad Hashemi received his M.s degree in Computer Sciences, Iran in 2013. He is currently PhD Candidate in Computer Sciences at Islamic Azad University, Sari Branch. His research interests are in the area of Data Mining and Meta heuristic algorithms.

Faraein Aeini received her Ph.D. degree in Software Engineering, Iran in 2013. She is currently Assistant Professor in Computer Engineering at Islamic Azad University, Sari Branch. Her research interests include Data Mining, Meta heuristic algorithms and manifold learning.

Homayun Motameni received his Ph.D. degree in Software Engineering, Iran in 2013. He is currently is Professor in Computer Engineering at Islamic Azad University, Sari Branch. His research interests are in the area of Software Engineering, Evolutionary Computation, Fuzzy Systems and Meta heuristic algorithms.

Behnam Barzegar received his Ph.D. degree in Software Engineering, Iran in 2013. He is currently is Assistant Professor in Computer Engineering at Islamic Azad University, Babol Branch. His research interests are in the area of Cloud Computing, Performance Analysis, Meta heuristic algorithms.