IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON HEALTH INFORMATICS FOR THE DEVELOPING WORLD

Received July 23, 2017, accepted August 8, 2017, date of publication August 25, 2017, date of current version September 19, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2741521

A New Hybrid Intelligent Framework for Predicting Parkinson's Disease

ZHENNAO CAI[1,2], JIANHUA GU[1], AND HUI-LING CHEN[2]

[1]School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China
[2]College of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou 325035, China

Corresponding author: Hui-Ling Chen (chenhuiling.jlu@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61303113, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY17F020012 and Grant LY14F020035, and in part by the Science and Technology Plan Project of Wenzhou, China, under Grant ZG2017019 and Grant H20110003.

ABSTRACT Parkinson's disease (PD) is a progressive neurodegenerative motor system disorder. Early diagnosis of PD is important to control the symptoms appropriately. Recent voice and speech recognition techniques provide alternative solutions for PD screening. In this paper, an optimal support vector machine (SVM) based on bacterial foraging optimization (BFO) was established to predict PD effectively. The effectiveness of the proposed method, BFO-SVM, was validated on a PD data set based on vocal measurements. The proposed method was compared with two of the most frequently used parameter optimization methods, including an SVM based on the grid search method and an SVM based on particle swarm optimization. Additionally, to further boost the prediction accuracy, the relief feature selection was employed prior to the BFO-SVM method, consequently the RF-BFO-SVM was proposed. The experimental results have demonstrated that the proposed framework exhibited excellent classification performance with a superior classification accuracy of 97.42%.

INDEX TERMS Bacterial foraging optimization, disease diagnosis, medical diagnosis, parameter optimization, Parkinson's disease diagnosis, support vector machines.

I. INTRODUCTION

At present, Parkinson's disease (PD) has become the world's second major neurodegenerative disease. Although the cause of PD remains unknown, its symptoms can be alleviated significantly if the illness is detected at an early stage [1]. Recent estimates indicate that approximately 90% of patients with PD exhibit symptoms of dysphonia [2]. Thus, vocal measurements could be used as an effective diagnostic tool for PD. In recent years, numerous studies concerning the use of dysphonic indicators to diagnose and monitor PD have been conducted [3], [4]. In one study, Little *et al.* [4] proposed to use feature selection technique combined with a support vector machine (SVM) classifier to identify PD patients. The resulting model detected PD patients with an accuracy of 91.4%. In another study, Shahbaba and Neal [5] used the Dirichlet process mixtures for PD diagnosis, the experimental results have shown that the proposed nonlinear model yielded a superior classification accuracy of 87.7% compared to other machine learning methods. Das [6] compared the PD diagnostic capabilities of artificial neural networks (ANN), DMneural, and regression and decision trees. The results have shown that the ANN yielded the most accurate diagnostic results with an overall accuracy of 92.9%. In addition, Sakar and Kursun [7] developed a PD diagnostic tool with a classification accuracy of 92.75% using a combination of mutual information and SVM. In another study, Psorakis *et al.* [8] proposed an improved multiclass multikernel relevance vector machines (mRVMs) to detect the PD. It yielded an overall PD classification accuracy rate of 89.47%. Guo *et al.* [9] proposed to use genetic programming and the expectation maximization algorithm (GP-EM) to develop a diagnostic tool for PD with a classification accuracy of 93.1%. More recently, Luukka [10] employed the similarity classifier combined with fuzzy entropy measurements-based feature selection to detect PD; the resulting diagnostic method achieved a mean classification accuracy of 85.03% using only two dysphonic features. Li *et al.* [11] developed a PD diagnostic tool with a classification accuracy of 93.47% using a fuzzy-based non-linear transformation method and SVM classifier. In another study, Ozcift and Gulten [12] proposed to use the rotation forest ensemble classifier combined with a correlation based

feature selection (CFS) algorithm to identify patients with PD; the resulting model has produced a classification accuracy of 87.13%. In addition, AStröm and Koker [13] developed a parallel feed-forward neural network with a PD classification accuracy of 91.20%. Spadoto *et al.* [14] employed evolutionary-based techniques and the Optimum-Path Forest (OPF) classifier to identify PD patients. The resulting method exhibited a classification accuracy of 84.01%. Emary *et al.* [15] proposed a diagnostic tool for PD by combining the fuzzy k-nearest neighbor (FKNN) approach and a principle component analysis (PCA). The resulting PCA-FKNN approach yielded a classification accuracy of 96.07%. Zuo *et al.* [16] devised an effective PD aided diagnostic system based on PSO enhanced FKNN with a mean accuracy of 97.47%. Hariharan *et al.* [17] proposed a hybrid method by combining several feature pre-processing methods with several classifiers including a least-square SVM, probabilistic neural networks, and general regression neural network to diagnose the PD; the resulting method resulted in a classification accuracy of 100%. Furthermore, Gök et al. [18] developed a discriminative model with a classification accuracy of 98.46% using a rotation forest ensemble KNN classifier. Peker *et al.* [19] proposed to combine a minimum redundancy maximum relevance attribute selection algorithm with the complex-valued artificial neural network to detect the PD, the classification accuracy of 98.12% has been obtained by the proposed methodology. Chen *et al.* [20] has proposed to use the extreme learning machine (ELM) and kernel extreme learning machine (KELM) for early diagnosis of PD. The experimental results have shown that the proposed KELM in combination with feature selection method can achieve very promising classification accuracy with the highest accuracy of 96.47% and average accuracy of 95.97% over 10 runs of 10-fold CV.

As shown above, SVM is one of the most popular and effective machine learning methods used to diagnose PD. However, comparatively few studies concerning the use of SVM as a diagnostic tool for PD have been conducted. SVM [21], [22], is used to identify the tradeoff between training set error minimization and margin maximization in order to achieve optimal generalization ability while preventing over-fitting. Additionally, convex quadratic programming is used in SVM to prevent the selection of local minima. Due to these properties, SVM has been applied to many classification problems [23]–[27]. SVM has been proven to be particularly useful in medical diagnostic problems [26]–[31]. However, SVM classifiers could still be drastically improved. For example, selecting the proper parameters can significantly improve the classification accuracy of SVM [32]. Thus, the values of certain parameters, such as the penalty and kernel function parameters, should be carefully selected before applying an SVM to any practical problem. SVMs with radial basis function (RBF) kernels have been reported to be well-suited for classification problems and are often the first techniques to be applied [33]. Therefore, this study primarily focused on identifying the optimal parameter values of an RBF kernel function (i.e., C and γ). Traditionally, these parameters have been selected with the grid-search method [33] and gradient descent method [34]–[36]. However, these methods are all easy to be stuck into local optima. In recent years, research has shown that metaheuristics, such as genetic algorithms [37], particle swarm optimization (PSO) [38]–[40], fruit fly optimization [41], whale optimization algorithm [42], differential flower pollination [43], are more apt at identifying global optimum solutions than the aforementioned traditional methods. The bacterial foraging optimization (BFO) method, a relatively new swarm-intelligence has been successfully applied to many real-world optimization problems, such as optimal controller design [44], artificial neural network learning [45], stock market index prediction [46], automatic circle detection in digital images [47], harmonic estimation [48], aluminum electrolysis production process [49], structural learning of Bayesian networks [50] and active power filter design [51]. Therefore, in this paper, the maximum classification performance of an SVM was explored by using the BFO strategy to simulate the foraging behavior of E. coli bacteria and its interactions with the surrounding environment. The primary target of this study was to investigate the maximum generalization capabilities of SVM to effectively identify the PD patients. Additionally, in order to further improve the diagnostic accuracy, we have performed the feature selection using Relief prior to the BFO-SVM prediction model. In the proposed RF-BFO-SVM method, an objective function was designed using the cross validation (CV) classification accuracy of an SVM in order to analyze its generalization capabilities. The experimental results indicated that the established RF-BFO-SVM can result in a higher diagnostic accuracy than two other commonly used methods, including the SVM based on grid search (Grid-SVM), SVM based on PSO (PSO-SVM), KELM [52] and random forest (RF) [53]. It should be noted that this work is an elaboration of our previously published conference paper [54] and that further details regarding the underlying mechanisms of the proposed method and experimental process have been provided.

In summary, the main contributions of this study are as follows:

a) First, in order to fully exploit the potential of the SVM classifier, we introduce BFO strategy to adaptively determine the two key parameters of SVM.
b) The resulting model combined with feature selection, RF-BFO-SVM is applied to discriminate the persons with PD from the healthy ones for the first time on the two commonly used PD datasets.
c) The proposed RF-BFO-SVM manages to achieve better classification performance, and offers more stable and robust results when compared to other advanced machine learning methods such as PSO-SVM, Grid-SVM, KELM and RF.

The remainder of this paper is structured as follows. Background information regarding SVM and BFO is presented in

Section 2. In Section 3, the detailed implementation of the proposed diagnostic system is described. The experimental design is discussed in Section 4. In Section 5, the experimental results and a discussion on the proposed method are provided. Lastly, in Section 6, the conclusion and future work are summarized.

II. BACKGROUND

A. SUPPORT VECTOR MACHINES (SVM)

This section gives a brief description of SVM. More details can be referred to literature [21], [55], which provides a complete description of the SVM theory. SVM is a kind of classification algorithm, which is devoted to improving the generalization ability by seeking the minimum structural risk in the learning machine. The core idea lies in the fact that it is the maximum margin strategy, which can be finally transformed into solving a convex quadratic programming problem. Owing to this outstanding property, SVM has found its applications in a wide range of fields [26–27,56–63].

In a binary classification task, the samples are separated with a hyperplane $w^T x + b = 0$, w is a d-dimensional coefficient vector that is normal to the hyper plane, and b is the offset from the origin, x are data points. The main task of SVM is to get the results of w and b. In linear case, w can be solved by introducing Lagrangian multipliers. The data points on the maximum border are called support vectors. As a result, the solution of w takes the following form: $w = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$, where n is the number of SVs, yi are the labels corresponding samples x. After then b can be derived from $y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 = 0$, where are support vectors. After w and b are determined, the linear discriminant function can be given by Eq. (1).

$$g(\mathbf{x}) = sgn\left(\sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b\right) \tag{1}$$

In non-linear cases, a general idea of kernel trick is introduced. And then the decision function can be expressed as follows:

$$g(\mathbf{x}) = sgn\left(\sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{2}$$

Generally, any positive semi-definite functions that satisfy the Mercer's condition can be used as kernel functions [64], such as the polynomial kernel ($K(x, x_i) = ((x^T x_i) + 1)^d$) and the Gaussian kernel ($K(x, x_i) = \exp\ (-\gamma\ \|x - x_i\|^2)$) as shown in Table 1.

B. BACTERIAL FORAGING OPTIMIZATION (BFO)

BFO consists of three principle mechanisms, including chemotaxis, reproduction, and elimination-dispersal [44]. In this section, BFO is briefly described. For more detail of BFO refer to the [44], [65].

TABLE 1. Four common kernel functions

Kernel types	Kernel functions
Linear kernel	$K(x,x_i)=(x^T x_i)$
Polynomial kernel	$K(x,x_i)=((x^T x_i)+1)^d$
Radial based kernel (RBF)	$K(x,x_i)=\exp(-\gamma\|x-x_i\|^2)$
Sigmoid kernel	$K(x,x_i)=\tanh((x^T x_i)+b)$

1) CHEMOTAXIS

An E.coli cell moving with its flagella is simulated in this step. An E.coli bacterium can move in two different ways, including moving and tumbling. The bacterium can move in a randomly-selected direction for a period of time or adjust the parameters of its movement, namely the direction and step length of the next movement, in order to adjust its position. The bacterium can alternate between these two operations for the duration of its lifetime. Swaying corresponds to the evaluation of an individual bacterium's current surrounding environment and cell-to-cell signaling process. Suppose that θi(j,k,l) represents the ith bacterium at the jth chemotactic, kth reproductive, and lth elimination-dispersal steps. In addition, assume that C(i) is the size of a step taken in a random direction (dcti) specified by the tumble (unit of step length). The computational chemotactic movement of the bacterium can be expressed as

$$\begin{aligned} \theta^i(j+1,k,l) &= \theta^i(j,k,l) + C(i) * dct_i \\ dct_i &= \frac{\Delta(i)}{\sqrt{\Delta^T(i)\,\Delta(i)}} \end{aligned} \tag{3}$$

where Δ denotes a vector in a random direction with elements that lie within the range of $[-1, 1]$.

2) SWARMING

During chemotactic movement, a bacterium releases an attractant in order to remain distant from the best individual within its population. In addition, the bacterium releases a repellant in order to maintain population diversity. The cell-to-cell signaling of an E. coli swarm can be expressed as

$$\begin{aligned} &J_{cc}(\theta, P(j,k,l)) \\ &= \sum_{i=1}^{S} J_{cc}\left(\theta, \theta^i(j,k,l)\right) \\ &= \sum_{i=1}^{S}\left[-d_{att}\exp\left(-w_{att}\sum_{m=1}^{p}\left(\theta_m - \theta_m^i\right)^2\right)\right] \\ &\quad + \sum_{i=1}^{S}\left[h_{repe}\exp\left(-w_{repe}\sum_{m=1}^{p}\left(\theta_m - \theta_m^i\right)^2\right)\right] \end{aligned} \tag{4}$$

where $Jcc(\theta, P(j,k,l))$ denotes the fitness function value to be added to the actual fitness function in order to present a time-varying fitness function, S denotes the swarm size of the population, p denotes the number of dimensions to be optimized in each bacterium, and $\theta = [\theta_1, \theta_2, \ldots, \theta_p]^T$ denotes a point in the p-dimensional search domain. The coefficients *datt*, *watt*, *hrepe*, *wrepe* must be properly selected. In 2002, Liu and Passino [65] incorporated a new function ($Jar(\theta)$) into BFO in order to represent any environmentally-dependent cell-to-cell signaling, such that

$$J_{ar}(\theta) = \exp(M - J(\theta)) J_{cc}(\theta) \tag{5}$$

where M is a tunable parameter, and $Jcc(\theta)$ is given in Eq. (4). The minimal value of $J(i,j,k,l) + Jar(\theta i)$ was considered during swarming.

3) REPRODUCTION

After a period of chemotactic movement, the BFO algorithm performs the reproduction step. First, the BFO ranks all of the individuals according to the sum of the evaluation results within the period of chemotactic movement. Then, the last half of the individuals is eliminated, and each of the remaining half are copied in order to maintain the size of the population. The individuals in the new generation are allowed to engage in chemotactic movement for another period. After several reproduction steps, the bacteria gather into several clusters, decreasing the diversity of the population.

4) ELIMINATION-DISPERSAL

In order to prevent the occurrence of premature phenomena, the BFO performs an elimination-dispersal step. In the step, some bacteria are randomly reinitialized with a minute probability of survival, while others are randomly initialized throughout the search space. Then, the individuals in the new generation of bacteria are allowed to engage in chemotactic movement for another period.

C. RELIEF METHOD

Relief is a well-known multivariate filtering feature selection algorithm proposed by Kira and Rendel [66] in 1992. It is also a feature weighting algorithm based on sample learning. The Relief algorithm selects the neighbors that participate in the weight calculation by calculating the distance between two samples. Because the features involved in the calculation of the distance will affect the relative distance of the samples, thus affecting the choice of neighbors, and ultimately playing a role of the evaluation of the feature weight. Therefore, the interaction between features is realized in the process of calculating neighborhoods. Relief measures the distinguishing ability of a feature by examining the difference between a similar neighbor sample and a heterogeneous neighbor sample. If the difference between the similar samples is small, while the difference between the heterogeneous samples is large, the variable has a strong ability to distinguish.

Given the sample set $S = \{s_1, s_2, \cdots, s_m\}$, each sample contains n features, $s_i = \{s_{i1}, s_{i2}, \cdots, s_{in}\}$, $1 \le i \le m$. The values of all features are scalar or numeric. The class label of si is ci, $c_i \in C$, is a set of class labels. The difference between the two samples si and $sj (1 \le i \ne j \le m)$ in the kth$(1 \le k \le n)$ feature is defined as following.

If the kth feature is scalar,

$$diff(k, s_i, s_j) = \begin{cases} 0 & s_{ik} = s_{jk} \\ 1 & s_{ik} \ne s_{jk} \end{cases} \tag{6}$$

If the kth feature is numeric,

$$diff(k, s_i, s_j) = \left| \frac{s_{ik} - s_{jk}}{\max_k - \min_k} \right| \tag{7}$$

where $maxk$ and $mink$ are respectively the maximum and minimum values of the kth feature in the sample set.

The Relief algorithm first randomly selects a sample s_i from the sample set and two samples closest to s_i from each of the two-class samples. Samples of the same type with s_i are represented by Hit, and samples of the different type with s_i are represented by Miss, using Hit and Miss to update the weight of the kth $(1 \le k \le n)$ feature according to the following formula:

$$w_k = w_k - diff(k, s_i, Hit)/t + diff(k, s_i, Miss)/t \tag{8}$$

where w_k represents the weight of the kth feature, t is the sampling times.

III. THE DEVELOPED HYBRID FRAMEWORK

The RF-BFO-SVM framework was developed to effectively discriminate the PD patients from the healthy controls by combing feature selection with the BFO based SVM. A flowchart of the proposed RF-BFO-SVM is shown in Fig. 1. In the proposed framework, firstly, the vocal data was normalized by scaling to the range [−1,1], and the redundant and nonrelevant features was removed using Relief method, a well-known filtering feature selection algorithm. Secondly, the parameters of an SVM were optimized dynamically by BFO. Then, the optimal parameters were fed into SVM to train an optimal diagnostic model. The developed framework primarily consisted of parameter optimization procedure and performance evaluation procedure.

In the parameter optimization procedure, the main objective is to evaluate the performance of each set of candidate parameters via using the BFO algorithm. The average predictive accuracy over the 5-fold cross validation was set to be the fitness function. The pseudo-code of the parameter optimization procedure is presented below.

While in the performance evaluation procedure, the main aim is to evaluate the overall performance of the SVM classifier with the obtained optimal parameters. The pseudo-code of the performance evaluation procedure is presented below.

IV. EXPERIMENTAL DESIGN

A. DATA DESCRIPTION

The Parkinson's disease data set was created by Little *et al.* [4] and is online at the UCI machine learning

Algorithm 1

step 1. Initialize the parameters *p, S, Nc, Ns, Nre, Ned, Ped*, and θ^i
where
p: number of dimensions in the search space,
S : swarm size of the population,
Nc: number of chemotactic steps,
Ns: swimming length,
Nre: number of reproduction steps,
Ned: number of elimination-dispersal events,
Ped: elimination-dispersal probability, and
$C(i)$: size of a step taken in a random direction specified by the tumble.

step 2. Elimination-dispersal loop: $l = l + 1$.
step 3. Reproduction loop: $k = k + 1$.
step 4. Chemotaxis loop: $j = j + 1$.

(a) **For** $i = 1, 2, \ldots, S$, perform a chemotactic step for bacterium i as follows.
(b) Train the SVM and compute the fitness $(J(i, j, k, l))$
Let, $J(i, j, k, l) = J(i, j, k, l) + J_{ar}(\theta)$, where J_{ar} is defined in Eq. (5).
(c) Let $Jlast = J(i, j, k, l)$. Save this value since a cost better than a run may be identified.
(d) Tumble: generate a random vector $(\Delta(i) \in R^p)$ using each element of $\Delta_m(i), m = 1, 2, \ldots, p$, a uniformly distributed random number on$[-1, 1]$.
(f) Move: let

$$\theta^i(j+1, k, l, di) = \theta^i(j, k, l, di) + C(i) \times \frac{\Delta(i)}{\sqrt{\Delta^T(i)\,\Delta(i)}}$$

(g) Train the SVM and compute the fitness $(J(i, j+1, k, l))$. In addition, let

$$J(i, j+1, k, l) = J(i, j, k, l) + Jar(\theta).$$

(h) Swim.
i) Let $n = 0$;
ii) **While** $n < Ns$
iii) Let $n = n + 1$;
iv) **If** $J(i, j+1, k, l) < Jlast$, let $Jlast = J(i, j+1, k, l)$, and let

$$\theta^i(j+1, k, l, di) = \theta^i(j, k, l, di) + C(i) \times \frac{\Delta(i)}{\sqrt{\Delta^T(i)\,\Delta(i)}}$$

Use $\theta^i(j+1, k, l)$ to train the SVM. Then, computer the new fitness $(J(i, j+1, k, l))$ as shown in (g);
v) Else, let $n = Ns$.
(i) If $i \neq S$, move to the next bacterium $(i + 1)$.

step 5. If $j < Nc$, go to IV.

Algorithm 1 *Continued.*

step 6. Reproduction:
Rank all of the individuals according to the sum of the evaluation results in this period. Then, eliminate the last half of the individuals and copy each of the remaining individuals.
step 7. If $k < Nre$, go to step 3.
step 8. Elimination-dispersal:
For $i = 1, 2, \ldots, S$ with a probability of *Ped*, eliminate and disperse each bacterium.
If $l < Ned$, go to step 2; otherwise, end.

Begin
For $j = 1 : k$
Training set: $k - 1$ subsets;
Testing set: remaining subset;
Train the SVM classifier on the training set using the parameters and feature subsets obtained during the parameter optimization procedure;
the trained SVM classifier to the testing set;
End For;
Return the average predictive accuracy of the SVM over the j testing set;
End.

repository. The aim of this data was to discriminate patients with PD from healthy controls by identifying differences in their vowel vocalizations. The voice recordings of a total of 31 subjects, including 23 patients with PD (16 males and 7 females) and 8 healthy controls (3 males and 5 females) were used for the purposes of this study. In addition, the subjects ranged from 46 to 85 years of age, with a mean age of 65.8 years. Each subject provided an average of six 36-second-long phonations of a vowel, yielding a total of 195 samples. Each recording was subjected to various measurements, such as vocal perturbation and other nonlinear measurements, resulting in the identification of 22 real-value features. Table 2 lists the 22 vocal features and the statistical parameters.

The distribution of the two classes including the PD patients and the healthy controls in the subspace given by the three first principal components is shown in Fig.2. From the figure, we can see that there is a strong overlap between the two class distributions.

B. METHODS FOR COMPARISON

The proposed BFO-SVM classification scheme was compared to other parameter optimization techniques to validate its effectiveness.

The developed BFO-SVM was first compared to the grid-search method (Grid-SVM herein), the most frequently used conventional parameter optimization technique. In the grid-search method, the SVM selects the pair of parameters (C, γ) with the highest CV accuracy as the final RBF kernel. Comparing various pairs of exponentially growing sequences of

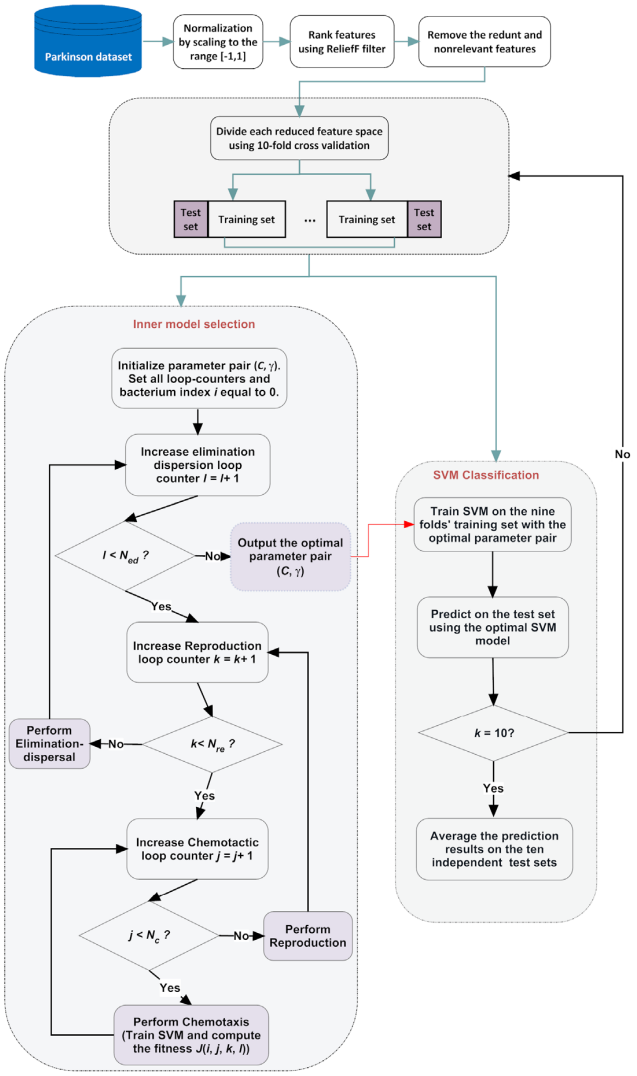


FIGURE 1. Flowchart of the proposed hybrid framework.

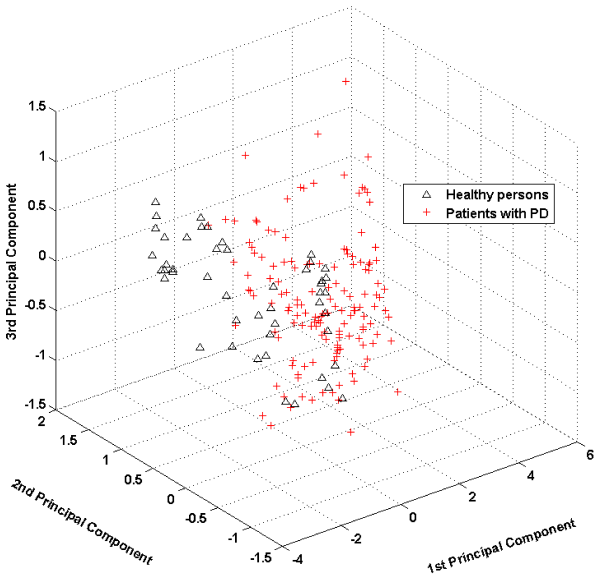


FIGURE 2. Distribution of the two classes in the subspace formed by the three principle components.

FIGURE 3. Illustration of the 5-fold CV process.

C and γ is considered to be a practical means of identifying optimal parameters for an SVM with RBF kernels [33].

The proposed BFO-SVM classification scheme was also compared to the PSO-based parameter optimization technique (PSO-SVM herein), a global optimization-based methodology. PSO [67] is a population based algorithm inspired by the social behaviors of natural swarms, such as flocks of birds and schools of fish. In PSO, the search space is explored by a number of particles, which are updated based on the optimum performance of each particle. In this study, a linearly decreasing inertia weight [68], [69] was adopted in order to promote the global and local search capabilities of the PSO at the beginning and end of the search process, respectively.

In order validate the effectiveness of the developed BFO-SVM approach, we have also compared the proposed method with other advanced machine learning methods including KELM and RF.

In Section 5, the BFO-SVM, Grid-SVM, PSO-SVM, KELM, and RF approaches are compared.

C. EXPERIMENTAL SETUP

The involved classification models and Relief feature selection were implemented in MATLAB. The LIBSVM package was used to construct the SVM model [33]. For KELM, the code was implemented in MATLAB which were available from http://www3.ntu.edu.sg/home/egbhuang. For RF, the software package at http://code.google.com/p/randomforest-matlab/ was used.

Normalization was performed before the classification process. The data was scaled into the interval [0, 1]. The computational analysis was conducted on a Windows Server 2008 R2 operating system with Intel (R) Xeon (R) CPU E5-2660 v3 (2.60 GHz) and 16GB of RAM.

In order to get the unbiased results, k-fold CV [70] was used to evaluate the performance of relevant methods. The 5-fold CV process is illustrated in Fig. 3. The data was divided into five subsets; each iteration, one of the five subsets was applied to the test set and the other four subsets were used

TABLE 2. The 22 vocal attributes of PD data set.

Label	Attribute	Discription	Mean ± Std.
F_1	MDVP:Fo(Hz)	Average vocal fundamental frequency	154.2286 ± 41.3901
F_2	MDVP:Fhi(Hz)	Maximum vocal fundamental frequency	197.1049 ± 91.4915
F_3	MDVP:Flo(Hz)	Minimum vocal fundamental frequency	116.3246 ± 43.5214
F_4	MDVP:Jitter(%)	Several measures of variation in fundamental frequency	0.0062 ± 0.0048
F_5	MDVP:Jitter(Abs)		0.000044 ± 0.0000348
F_6	MDVP:RAP		0.0033 ± 0.0030
F_7	MDVP:PPQ		0.0034 ± 0.0028
F_8	Jitter:DDP		0.0099 ± 0.0089
F_9	MDVP:Shimmer	Several measures of variation in amplitude	0.0297 ± 0.0189
F_{10}	MDVP:Shimmer(dB)		0.2823 ± 0.1949
F_{11}	Shimmer:APQ3		0.0157 ± 0.0102
F_{12}	Shimmer:APQ5		0.0179 ± 0.0120
F_{13}	MDVP:APQ		0.0241 ± 0.0169
F_{14}	Shimmer:DDA		0.0470 ± 0.0305
F_{15}	NHR	Two measures of ratio of noise to tonal components in the voice	0.0248 ± 0.0404
F_{16}	HNR		21.8860 ± 4.4258
F_{17}	RPDE	Two nonlinear dynamical complexity measures	0.4985 ± 0.1039
F_{18}	D2		0.7181 ± 0.0553
F_{19}	DFA	Signal fractal scaling exponent	-5.6844 ± 1.0902
F_{20}	Spread1	Three nonlinear measures of fundamental frequency variation	0.2265 ± 0.0834
F_{21}	Spread2		2.3818 ± 0.3828
F_{22}	PPE		0.2066 ± 0.0901

for training. Then, the average of the classification accuracy values was computed. One of the advantages of this method was that all of the test sets were evaluated independently,

TABLE 3. Common BFO parameters.

S	8
Nc	25
Ns	4
Ned	2
Nre	3
ped	0.25
$datt$	0.1
$watt$	0.2
$Wrepe$	10
$hrepe$	0.1
$C(i)$	0.1

TABLE 4. Confusion matrix.

	Predicted PD patient	Predicted healthy person
Actual PD patient	True Positive (TP)	False Negative (FN)
Actual healthy person	False Positive (FP)	True Negative (TN)

improving the reliability of the results. The experimental process was designed using a two-loop scheme, which is commonly used to prevent over-fitting [68], [71]. The inner loop was used to determine the optimal parameters of the SVM classifier, while the outer loop was used to determine the performance of the SVM classifier. For the purposes of this study, a 10-fold CV scheme was used for the outer loop, and a 5-fold CV scheme was used for the inner loop, respectively.

The detailed parameter settings of a BFO are shown in Table 3. The number of iterations and particles in the PSO-SVM were defined as 250 and 8, respectively. In addition, $c_1 = 2, c_2 = 2$. The searching ranges $C \in \{2^{-5}, 2^{-3}, \ldots, 2^5\}$ and $\gamma \in \{2^{-5}, 2^{-3}, \ldots, 2^5\}$ were used for the SVM methods and KELM method. The searching ranges for the two parameters of RF including ntree (number of trees) and mtry (the number of variables) were chosen from the range of {50, 100, 150, . . . , 500} and {1, 2, 3, 4, 5}, respectively. The experimental results show that RF achieved the best performance when ntree = 500, mtry = 4.

D. PERFORMANCE METRICS

The performance of the three methods were validated via the classification accuracy, sensitivity, and specificity measurements. Table 4 displays an illustration of a confusion matrix, wherein TP denotes the number of PD patients correctly classified as such (true positives), FN denotes the number of PD patients classified as healthy controls (false negatives), TN denotes the number of healthy

TABLE 5. Results of BFO-SVM on the PD data set with different values of $C(i)$.

Chemotactic step size parameter $C(i)$	BFO-SVM		
	Accuracy (%)	Sensitivity (%)	Specificity (%)
0.05	94.84(6.43)	97.41(6.11)	88.00(17.79)
0.1	**96.89(4.34)**	**98.75(3.95)**	**90.83(16.87)**
0.15	95.39(4.53)	97.44(4.33)	90.42(10.77)
0.2	94.42(4.97)	97.57(4.19)	85.50(13.17)
0.25	93.40(6.28)	96.08(5.48)	87.71(17.36)
0.3	94.47(4.97)	96.02(4.82)	90.64(12.32)

controls classified as such (true negatives), and FP represents the number of healthy controls classified as PD patients (false positives).

The accuracy, sensitivity, and specificity can be defined as follows

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (9)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (10)$$

$$Specificity = \frac{TN}{FP + TN} \times 100\% \quad (11)$$

V. EXPERIMENTAL RESULTS

A. EXPERIMENT I: CLASSIFICATION WITHOUT FEATURE SELECTION

The performance of BFO can be influenced by the chemotaxis step size $C(i)$ parameter, which plays an essential role in controlling the search abilities of BFO. Thus, in this study, the effects of $C(i)$ on the performance of the proposed BFO-SVM method were firstly investigated. Although $C(i)$ can be initialized with a biologically motivated value, that value may not be optimal for certain applications [44]. The detailed relationship between different values of $C(i)$ and the performance of the BFO-SVM is shown in Table 5. The average results and their standard deviations (in parentheses) are presented in the table. As shown, the BFO-SVM performed best when $C(i) = 0.1$. Therefore, a value of 0.1 was assigned to $C(i)$ for the purposes of this study.

To evaluate the effectiveness of the proposed method, the BFO-SVM was applied to the PD data set. Table 6 displays the confusion matrix with different performance metrics and optimal pairs of parameters obtained by the BFO-SVM in each fold. As shown, confused results primarily occurred in the 1st, 2nd, 4th, and 7th folds. Two PD patients

TABLE 6. Results obtained by the proposed BFO-SVM when applied to the PD data set.

Fold No.	BFO-SVM						
	Confusion matrix		Accuracy	Sensitivity	Specificity	C ($\times 10^4$)	γ
1	**14**	2	0.8947	0.8750	1.0000	1.4250	3.7726
	0	**3**					
2	**13**	0	0.9474	1.0000	0.8333	2.2658	3.3197
	1	**5**					
3	**17**	0	1.0000	1.0000	1.0000	2.2930	3.1821
	0	**3**					
4	**15**	0	0.9474	1.0000	0.7500	1.2372	4.6147
	1	**3**					
5	**11**	0	1.0000	1.0000	1.0000	2.8519	3.5251
	0	**8**					
6	**14**	0	1.0000	1.0000	1.0000	1.4252	3.8452
	0	**6**					
7	**16**	0	0.9000	1.0000	0.5000	2.0478	3.2553
	2	**2**					
8	**14**	0	1.0000	1.0000	1.0000	0.1331	3.7350
	0	**6**					
9	**15**	0	1.0000	1.0000	1.0000	2.2390	3.7168
	0	**5**					
10	**16**	0	1.0000	1.0000	1.0000	0.8046	4.9414
	0	**3**					
Avg.			**0.9689**	**0.9875**	**0.9083**	**1.6723**	**3.7908**
Dev.			**0.0434**	**0.0395**	**0.1687**	**0.8182**	**0.5739**

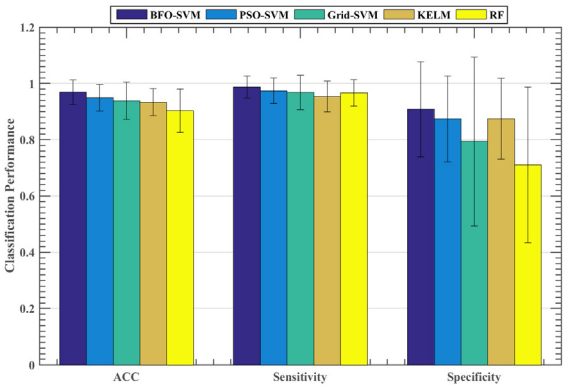


FIGURE 4. Classification performance comparison among the five methods.

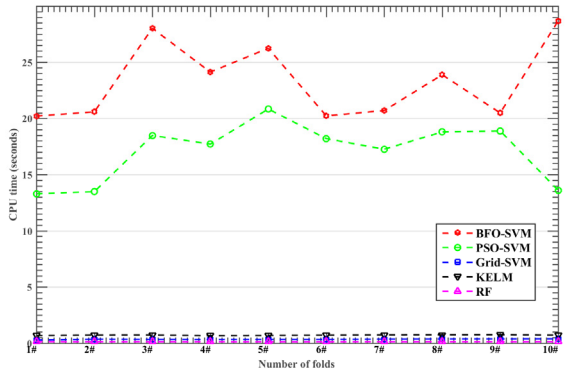


FIGURE 5. CPU cost comparison among the five methods.

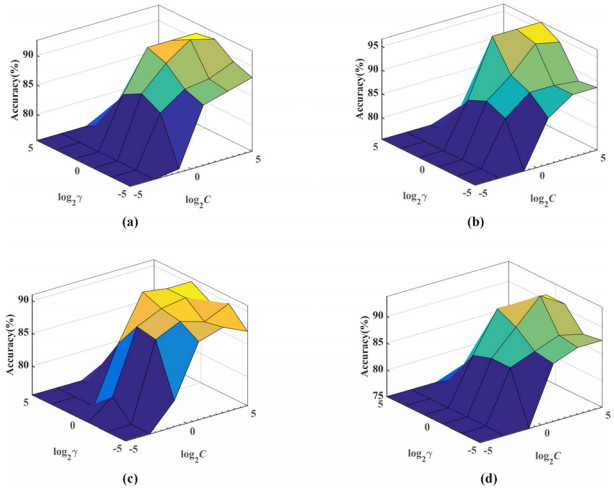


FIGURE 6. 3D view of parameters selection of the Grid-SVM for several folds.

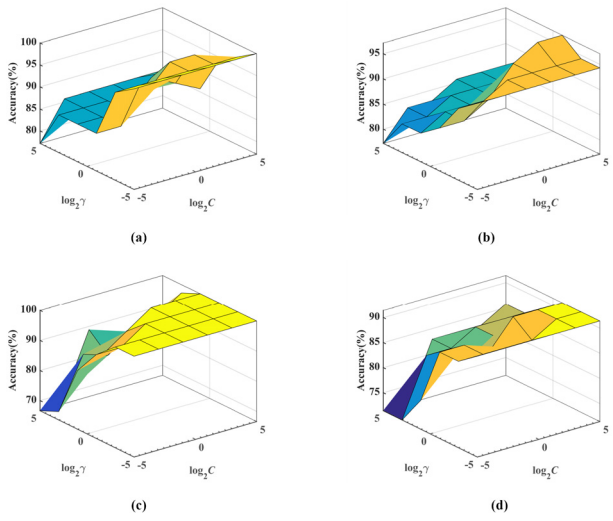


FIGURE 7. 3D view of parameters selection of the KELM for several folds.

were misclassified as healthy controls, while four healthy controls were misclassified as PD patients. The classification results and CPU costs obtained by the five methods are illustrated in Fig. 4 and Fig. 5, respectively. 3D view of parameters

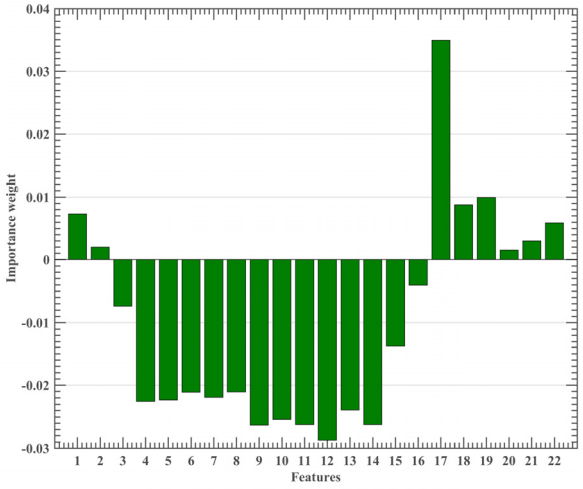


FIGURE 8. The weight of each feature obtained by the relief algorithm.

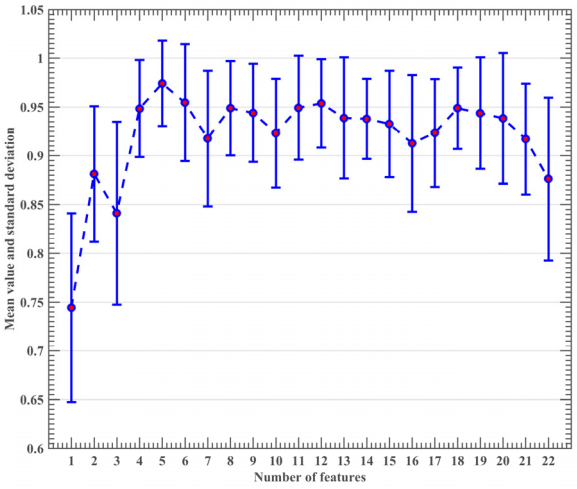


FIGURE 9. The relationship between different feature subsets and the classification accuracies of BFO-SVM.

selection of the SVM and KELM by applying the grid search method to the PD data set for several folds are illustrated in Fig.6 and Fig. 7, respectively.

As shown in Fig. 4, the performance of the proposed BFO-SVM outperformed the other four methods in terms of ACC, sensitivity and specificity. The average predictive accuracy of the BFO-SVM was approximately 96.89%, while the average classification accuracies of the PSO-SVM, Grid-SVM, KELM and RF were approximately 94.89%, 93.87%, 93.34% and 90.32%, respectively. As shown, the BFO-SVM also achieved a higher sensitivity and specificity than those of the PSO-SVM, Grid-SVM, KELM, and RF. The average sensitivity of the BFO-SVM over all ten folds was approximately 98.75%, while those of the PSO-SVM, Grid-SVM, KELM and RF were approximately 97.41%, 96.83%, 95.33% and 96.62%, respectively. The average specificity of the BFO-SVM over all ten folds was approximately 90.83%, while those of the PSO-SVM, Grid-SVM, KELM and RF

TABLE 7. Feature subsets produced by relief algorithm.

Size	Feature subset
1	{ F17}
2	{ F17, F19}
3	{ F17, F19, F18}
4	{ F17, F19, F18, F1}
5	{ F17, F19, F18, F1, F22}
6	{ F17, F19, F18, F1, F22, F21}
7	{ F17, F19, F18, F1, F22, F21, F2}
8	{ F17, F19, F18, F1, F22, F21, F2, F20}
9	{ F17, F19, F18, F1, F22, F21, F2, F20, F16}
10	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3}
11	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15}
12	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8}
13	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8, F6}
14	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8, F6, F7}
15	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8, F6, F7, F5}
16	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8, F6, F7, F5, F4}
17	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8, F6, F7, F5, F4, F13}
18	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8, F6, F7, F5, F4, F13, F10}
19	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8, F6, F7, F5, F4, F13, F10,F14 }
20	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8, F6, F7, F5, F4, F13, F10,F14, F11}
21	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8, F6, F7, F5, F4, F13, F10,F14, F11, F9}
22	{ F17, F19, F18, F1, F22, F21, F2, F20, F16, F3, F15, F8, F6, F7, F5, F4, F13, F10,F14, F11, F9, F12}

were approximately 87.42%, 79.40%, 87.50% and 71.00%, respectively.

As shown, we can find that the performance of the BFO-SVM and PSO-SVM is superior than the Grid-SVM. This indicated that swarm intelligence based optimization methods, such as BFO and PSO, are significantly more effective at SVM parameter tuning than the grid search technique, likely due to the susceptibility of the grid search method to local optima. Moreover, the standard deviations of the three performance metrics of the Grid-SVM were higher than those of the BFO and PSO based models. It reveals that the swarm intelligence optimized SVM models can offer much more stable results than the grid search optimized SVM model. We can also find that the standard deviations of the BFO-SVM on the three performance metrics were lower than or comparable with those of the other four methods. It indicates that the proposed BFO-SVM has resolved the PD diagnostic problem more effectively than the PSO-SVM and produces much more robust results. However, it should be noted that the proposed BFO-SVM is much more time-consuming than other four methods as shown in Fig.5. RF performs the fastest among the five methods, followed by Grid-SVM, KELM, PSO-SVM, and BFO-SVM.

TABLE 8. Results of the RF-BFO-SVM model with the best features.

Fold	ACC	Sensitivity	Specificity
1#	1.0000	1.0000	1.0000
2#	1.0000	1.0000	1.0000
3#	1.0000	1.0000	1.0000
4#	1.0000	1.0000	1.0000
5#	1.0000	1.0000	1.0000
6#	0.9000	0.6000	0.8000
7#	0.9474	0.7500	0.8750
8#	1.0000	1.0000	1.0000
9#	1.0000	1.0000	1.0000
10#	0.8947	0.8000	0.8643
Mean	0.9742	0.9929	0.9150
Std.	0.0437	0.0226	0.1454

B. EXPERIMENT II: CLASSIFICATION WITH FEATURE SELECTION

To investigate the best feature subset for the diagnosis of PD, we have implemented the proposed method combined with feature selection. The Relief method was implemented in order to rank the features. The weight of each feature is computed as displayed in Fig. 8. According to the weight of the feature, we can obtain 22 different feature subsets by adding features one-by-one from higher to lower rank. The BFO-SVM classifier can be constructed on each feature subset in an incremental manner. As we can see that it produced 22 different incremental feature subsets as shown in Table 7. Additionally, we studied the relationship between RF-BFO-SVM's classification performance and the different features. As shown in Fig.9, we can see that as the number of features increases, RF-BFO-SVM's performance changes.

When the number of features arrived at 5, the mean accuracy obtained the highest value with a minimum standard deviation.

Table 8 lists the results of RF-BFO-SVM constructed on the best feature subset with the best five features in terms of three different performance metrics. The table shows that BFO-SVM classification performance has been further improved by Relief feature selection. As shown, after applied feature selection, the ACC, sensitivity, and specificity were further improved by 0.53%, 0.54% and 0.67%, respectively. It indicates that there is lot of redundant and irrelevant information existed in the data. It is also interesting to find that the standard deviation of the proposed method is also smaller than that on the original data. It indicates that the feature selection can aid the classifier to achieve more stable results.

VI. CONCLUSION AND FUTURE WORK

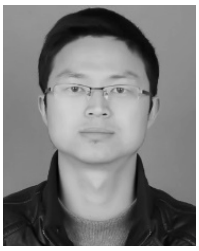
In this study, a new diagnostic method for PD, RF-BFO-SVM, was developed. In the proposed approach, the generalization capabilities of an SVM classifier were maximized by implementing a swarm intelligence technique in order to identify the parameters optimal for the diagnosis of PD. The results of the experiments, in which the BFO-SVM was applied to a PD data set, indicated that the proposed method outperformed PSO-SVM and Grid-SVM techniques in terms of classification accuracy, sensitivity, and specificity. Thus, the proposed RF-BFO-SVM method could be used as a viable decision support tool for PD.

In this study, the classical BFO algorithm was employed for parameter optimization, and the chemotactic step size parameter $C(i)$ was selected as the fixed step height. However, the search capabilities of a BFO algorithm significantly depend upon $C(i)$. Thus, an adaptive BFO algorithm in which the value of $C(i)$ could be adapted based upon a bacterium's surrounding environment could be developed in future studies. In addition, the computational efficiency could be improved by parallel computing methods, and the proposed method could be applied to other medical diagnostic problems in future studies.

REFERENCES

[1] N. Singh, V. Pillay, and Y. E. Choonara, "Advances in the treatment of Parkinson's disease," *Prog. Neurobiol.*, vol. 81, no. 1, pp. 29–44, 2007.
[2] A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of patients with Parkinson's disease," *Behavioural Neurol.*, vol. 11, no. 3, pp. 131–137, 1998.
[3] D. A. Rahn, III, M. Chou, J. J. Jiang, and Y. Zhang, "Phonatory impairment in Parkinson's disease: Evidence from nonlinear dynamic analysis and perturbation analysis," *J. Voice*, vol. 21, no. 1, pp. 64–71, 2007.
[4] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.
[5] B. Shahbaba and R. Neal, "Nonlinear models using Dirichlet process mixtures," *J. Mach. Learn. Res.*, vol. 10, pp. 1829–1850, Aug. 2009.
[6] R. Das, "A comparison of multiple classification methods for diagnosis of Parkinson disease," *Expert Syst. Appl.*, vol. 37, pp. 1568–1572, Mar. 2010.
[7] C. O. Sakar and O. Kursun, "Telediagnosis of Parkinson's disease using measurements of dysphonia," *J. Med. Syst.*, vol. 34, no. 4, pp. 591–599, 2010.
[8] I. Psorakis, T. Damoulas, and M. A. Girolami, "Multiclass relevance vector machines: Sparsity and accuracy," *IEEE Trans. Neural Netw.*, vol. 21, no. 10, pp. 1588–1598, Oct. 2010.
[9] P.-F. Guo, P. Bhattacharya, and N. Kharma, "Advances in detecting Parkinson's disease," in *International Conference on Medical Biometrics*. Berlin, Germany: Springer, 2010, pp. 306–314.
[10] P. Luukka, "Feature selection using fuzzy entropy measures with similarity classifier," *Expert Syst. Appl.*, vol. 38, pp. 4600–4607, Apr. 2011.
[11] D.-C. Li, C.-W. Liu, and S. C. Hu, "A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets," *Artif. Intell. Med.*, vol. 52, pp. 45–52, May 2011.
[12] A. Ozcift and A. Gulten, "Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms," *Comput. Methods Programs Biomed.*, vol. 104, pp. 443–451, Dec. 2011.
[13] F. Åström and R. Koker, "A parallel neural network approach to prediction of Parkinson's disease," *Expert Syst. Appl.*, vol. 38, pp. 12470–12474, 2011.
[14] A. A. Spadoto, R. C. Guido, F. L. Carnevali, A. F. Pagnin, A. X. Falcão, and J. P. Papa, "Improving Parkinson's disease identification through evolutionary-based feature selection," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug./Sep. 2011, pp. 7857–7860.
[15] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary ant lion approaches for feature selection," *Neurocomputing*, vol. 213, pp. 54–65, Nov. 2016.
[16] W.-L. Zuo, Z.-Y. Wang, T. Liu, and H.-L. Chen, "Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach," *Biomed. Signal Process. Control*, vol. 8, no. 4, pp. 364–373, 2013.
[17] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of Parkinson's disease," *Comput. Methods Programs Biomed.*, vol. 113, pp. 904–913, Mar. 2014.
[18] M. Gök, "An ensemble of k-nearest neighbours algorithm for detection of Parkinson's disease," *Int. J. Syst. Sci.*, vol. 46, pp. 1108–1112, Apr. 2015.
[19] M. Peker, B. Sen, and D. Delen, "Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm," *J. Healthcare Eng.*, vol. 6, no. 3, pp. 281–302, 2015.
[20] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," *Neurocomputing*, vol. 184, pp. 131–144, Apr. 2016.
[21] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.
[22] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 144–152.
[23] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[24] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," *Mach. Learn., ECML*, pp. 137–42, 1998.
[25] K.-S. Shin, T. S. Lee, and H.-J. Kim, "An application of support vector machines in bankruptcy prediction model," *Expert Syst. Appl.*, vol. 28, pp. 127–135, Jan. 2005.
[26] H.-L. Chen, D.-Y. Liu, B. Yang, J. Liu, and G. Wang, "A new hybrid method based on local Fisher discriminant analysis and support vector machines for hepatitis disease diagnosis," *Expert Syst. Appl.*, vol. 38, pp. 11796–11803, Sep. 2011.
[27] H.-L. Chen, B. Yang, J. Liu, and D.-Y. Liu, "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis," *Expert Syst. Appl.*, vol. 38, pp. 9014–9022, Jul. 2011.
[28] E. D. Übeyli, "Comparison of different classification algorithms in clinical decision-making," *Expert Syst.*, vol. 24, no. 1, pp. 17–31, 2007.
[29] H.-L. Chen, B. Yang, G. Wang, J. Liu, Y.-D. Chen, and D.-Y. Liu, "A three-stage expert system based on support vector machines for thyroid disease diagnosis," *J. Med. Syst.*, vol. 36, pp. 1953–1963, Jun. 2012.
[30] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Syst. Appl.*, vol. 36, pp. 3240–3247, Mar. 2009.
[31] I. Maglogiannis, E. Zafiropoulos, and I. Anagnostopoulos, "An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers," *Appl. Intell.*, vol. 30, no. 1, pp. 24–36, 2007.
[32] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.*, vol. 15, no. 7, pp. 1667–1689, 2003.

[33] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, 2011.
[34] C. Gold and P. Sollich, "Model selection for support vector machine classification," *Neurocomputing*, vol. 55, pp. 221–249, Sep. 2003.
[35] S. S. Keerthi, "Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1225–1229, Sep. 2002.
[36] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 131–159, 2002.
[37] C.-L. Huang and C.-J. Wang, "A GA-based feature selection and parameters optimization for support vector machines," *Expert Syst. Appl.*, vol. 31, pp. 231–240, Aug. 2006.
[38] S.-W. Lin, K.-C. Ying, S.-C. Chen, and Z.-J. Lee, "Particle swarm optimization for parameter determination and feature selection of support vector machines," *Expert Syst. Appl.*, vol. 35, pp. 1817–1824, Nov. 2008.
[39] H.-L. Chen, B. Yang, G. Wang, S.-J. Wang, J. Liu, and D.-Y. Liu, "Support vector machine based diagnostic system for breast cancer using swarm intelligence," *J. Med. Syst.*, vol. 36, no. 4, pp. 2505–2519, 2012.
[40] H. L. Chen *et al.*, "Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy," *Appl. Math. Comput.*, vol. 239, pp. 180–197, Jul. 2014.
[41] L. Shen *et al.*, "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowl.-Based Syst.*, vol. 96, pp. 61–75, Mar. 2016.
[42] A. Tharwat, Y. S. Moemen, and A. E. Hassanien, "Classification of toxicity effects of biotransformed hepatic drugs using whale optimized support vector machines," *J. Biomed. Inform.*, vol. 68, pp. 132–149, Apr. 2017.
[43] N.-D. Hoang, D. T. Bui, and K.-W. Liao, "Groutability estimation of grouting processes with cement grouts using differential flower pollination optimized support vector machine," *Appl. Soft Comput.*, vol. 45, pp. 173–186, Aug. 2016.
[44] K. M. Passino, "Biomimicry of bacterial foraging for distributed optimization and control," *IEEE Control Syst.*, vol. 22, no. 3, pp. 52–67, Jun. 2002.
[45] M. Ulagammai, P. Venkatesh, P. S. Kannan, and N. P. Padhy, "Application of bacterial foraging technique trained artificial and wavelet neural networks in load forecasting," *Neurocomputing*, vol. 70, pp. 2659–2667, Oct. 2007.
[46] R. Majhi, G. Panda, B. Majhi, and G. Sahoo, "Efficient prediction of stock market indices using adaptive bacterial foraging optimization (ABFO) and BFO based techniques," *Expert Syst. Appl.*, vol. 36, pp. 10097–10104, Aug. 2009.
[47] S. Dasgupta, S. Das, A. Biswas, and A. Abraham, "Automatic circle detection on digital images with an adaptive bacterial foraging algorithm," *Soft Comput.*, vol. 14, no. 11, pp. 1151–1164, 2010.
[48] S. Mishra, "A hybrid least square-fuzzy bacterial foraging strategy for harmonic estimation," *IEEE Trans. Evol. Comput.*, vol. 9, no. 1, pp. 61–73, Feb. 2005.
[49] J. Yi, D. Huang, S. Fu, H. He, and T. Li, "Multi-objective bacterial foraging optimization algorithm based on parallel cell entropy for aluminum electrolysis production process," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2488–2500, Apr. 2016.
[50] C. Yang, J. Ji, J. Liu, J. Liu, and B. Yin, "Structural learning of Bayesian networks by bacterial foraging optimization," *Int. J. Approx. Reason.*, vol. 69, pp. 147–167, Feb. 2016.
[51] S. Mishra and C. N. Bhende, "Bacterial foraging technique-based optimized active power filter for load compensation," *IEEE Trans. Power Del.*, vol. 22, no. 1, pp. 457–465, Jan. 2007.
[52] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.
[53] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[54] Y.-W. Fu, H.-L. Chen, S.-J. Chen, L. Shen, and Q. Li, "A new evolutionary support vector machine with application to Parkinson's disease diagnosis," in *Advances in Swarm Intelligence*. Berlin, Germany: Springer, 2014, pp. 42–49.
[55] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[56] M. Pontil and A. Verri, "Support vector machines for 3D object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 6, pp. 637–646, Jun. 1998.
[57] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1055–1064, Sep. 1999.
[58] H. Drucker, D. Wu, and V. N. Vapnik, "Support vector machines for spam categorization," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1048–1054, Sep. 1999.
[59] S. Hua and Z. Sun, "A novel method of protein secondary structure prediction with high segment overlap measure: Support vector machine approach," *J. Mol. Biol.*, vol. 308, no. 2, pp. 397–407, 2001.
[60] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 389–422, 2002.
[61] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, 2002.
[62] Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu, "Credit rating analysis with support vector machines and neural networks: A market comparative study," *Decision Support Syst.*, vol. 37, no. 4, pp. 543–558, 2004.
[63] J. Mourão-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data," *NeuroImage*, vol. 28, no. 4, pp. 980–995, 2005.
[64] B. Schölkopf, C. J. C. Burges, and A. J. Smola, *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA, USA: MIT Press, 1998.
[65] Y. Liu and K. M. Passino, "Biomimicry of social foraging bacteria for distributed optimization: Models, principles, and emergent behaviors," *J. Optim. Theory Appl.*, vol. 115, no. 3, pp. 603–628, 2002.
[66] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proc. Int. Workshop Mach. Learn.*, 1992, pp. 249–256.
[67] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 4. Nov./Dec. 1995, pp. 1942–1948.
[68] H.-L. Chen *et al.*, "A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method," *Knowl.-Based Syst.*, vol. 24, no. 8, pp. 1348–1359, 2011.
[69] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proc. Congr. Evol. Comput. (CEC)*, 1999, p. 1950.
[70] S. L. Salzberg, "On comparing classifiers: Pitfalls to avoid and a recommended approach," *Data Mining Knowl. Discovery*, vol. 1, no. 3, pp. 317–328, 1997.
[71] H.-L. Chen, D.-Y. Liu, B. Yang, J. Liu, G. Wang, and S.-J. Wang, "An adaptive fuzzy k-nearest neighbor method based on parallel particle swarm optimization for bankruptcy prediction," in *Advances in Knowledge Discovery and Data Mining*, J. Z. Huang, L. Cao, and J. Srivastava, Eds. Berlin, Germany: Springer, 2011, pp. 249–264.



ZHENNAO CAI received the master's degree from the School of Software Engineering, Huazhong University of Science and Technology. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Northwestern Polytechnical University, China. He is currently a Researcher with Wenzhou University, China. His main research interests include computer aided diagnosis, medical imaging computing, machine learning, high performance computing, and big data mining.

JIANHUA GU received the Ph.D. degree from Northwestern Polytechnical University, China, in 1998. He was a Visiting Scholar with Michigan State University, USA, in 2005. He is currently a Professor with the School of Computer Science and Engineering, Northwestern Polytechnical University, China. He has authored or co-authored over 80 scientific papers in international journals and books. His research interests include computer operating system, high performance computing, distributed computing, and big data analysis. He is a Senior Member of the China Computer Federation.

HUI-LING CHEN received the Ph.D. degree from the Department of Computer Science and Technology, Jilin University, China. He is currently a Lecturer with the Department of Computer Science and Technology, Wenzhou University, China. He has authored over 60 papers in international journals and conference proceedings, including the *Expert Systems with Applications*, the *Knowledge-Based Systems*, the *Neurocomputing*, the *Soft Computing*, PloS One, and PAKDD. His current research interests include machine learning and data mining, and their applications, such as medical diagnosis, bankruptcy prediction, and face recognition. He is currently a Reviewer for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, PART B.