CrossMark

SYSTEMS-LEVEL QUALITY IMPROVEMENT

A Multilayer Perceptron Based Smart Pathological Brain Detection System by Fractional Fourier Entropy

Yudong Zhang[1,2,3,4] · Yi Sun[5,6] · Preetha Phillips[7] · Ge Liu[8] · Xingxing Zhou[1,5] · Shuihua Wang[1,2]

Abstract This work aims at developing a novel pathological brain detection system (PBDS) to assist neuroradiologists to interpret magnetic resonance (MR) brain images. We simplify this problem as recognizing pathological brains from healthy brains. First, 12 fractional Fourier entropy (FRFE) features were extracted from each brain image. Next, we submit those features to a multi-layer perceptron (MLP) classifier. Two improvements were proposed for MLP. One improvement is the pruning technique that determines the optimal hidden neuron number. We compared three pruning techniques: dynamic pruning (DP), Bayesian detection boundaries (BDB), and Kappa coefficient (KC). The other improvement is to use the adaptive real-coded biogeography-based optimization (ARCBBO) to train the biases and weights of MLP. The experiments showed that the proposed FRFE + KC-MLP + ARCBBO achieved an average accuracy of 99.53 % based on 10 repetitions of K-fold cross validation, which was better than 11 recent PBDS methods.

Introduction

Pathological brain detection system (PBDS) can help physicians interpret medical brain images accurately [1–3]. In hospitals, the picture archiving and communication system (PACS) can provide either 3D brain or only a single slice that is associated to the foci within the brain [4–6]. Nevertheless, scanning the whole 3D brain is expensive and time-consuming [7–9], hence, we proposed a PBDS for single slice brain images.

At present, neuroradiologists used many neuroimaging methods to detect the brains by two ways: structural and functional. The structural imaging measures the inner of the brain structure, while the functional imaging measures its functions.

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ Shuihua Wang
  wangshuihua@njnu.edu.cn

1 School of Computer Science and Technology, Nanjing Normal University, Nanjing, Jiangsu 210023, China

2 Key Laboratory of Statistical Information Technology and Data Mining, State Statistics Bureau, Chengdu, Sichuan 610225, China

3 Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry Of Education, Jilin University, Changchun, Jilin 130012, China

4 State Key Lab of CAD & CG, Zhejiang University, Hangzhou, Zhejiang 310027, China

5 Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing, Nanjing, Jiangsu 210042, China

6 State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

7 School of Natural Sciences and Mathematics, Shepherd University, Shepherdstown, WV 25443, USA

8 Department of Psychiatry, College of Physicians, & Surgeons, Columbia University, New York, NY 10032, USA

In hospitals, structural imaging is commonly used by magnetic resonance imaging (MRI), since it displays better resolution for brain soft tissues and it does not relate to any radiations, compared to traditional X-ray and computed tomography (CT) [10].

In recent years, various PBDSs were developed by scholars [11–13]. They can provide user-friendly, professional, and even personalized assistance [14–16]. Their accurate performances motivate an increasing willingness for neuroradiologists to make decisions, and for patients to monitor healthy conditions regularly [17–19], with the help of PBDSs.

For instances, El-Dahshan, Hosny and Salem (2010) [20] employed techniques of discrete wavelet transform (DWT) and principal component analysis (PCA), and used K-nearest neighbor (KNN). Dong, et al. (2011) [21] employed a scaled conjugate gradient (SCG) method to train the artificial neural network (ANN). Das, Chowdhury and Kundu (2013) [22] combined Ripplet transform (RT) with PCA and least square SVM (LS-SVM). Wu (2012) [23] used support vector machine (SVM) to develop PBDS. Saritha, Paul Joseph and Mathew (2013) [24] used and spider-web plots (abbreviated as SWP) and wavelet-entropy (abbreviated as WE). They then employed probabilistic neural network (PNN) as the classifiers. El-Dahshan, et al. (2014) [25] employed the feedback pulse-coupled neural network (PCNN) to preprocess the brain images. Then they combined DWT and PCA to extract features. They finally employed back propagation neural network (BPNN). Wang, et al. (2015) [26] employed stationary wavelet transform (abbreviated as SWT) to take place of traditional DWT. Afterwards, to train the classifier, they designed a new training algorithm, viz., the hybridization of PSO and ABC (shorted as HPA). Sun, et al. (2015) [27] combined Hu moment invariants (HMI) with wavelet entropy. Generalized eigenvalue proximal SVM (GEPSVM) was employed. Wibmer, et al. (2015) [28] proposed a novel Haralick texture as image feature. Dong, Ji and Yang (2015) [29] proposed a new image feature as wavelet packet Tsallis entropy (WPTE) and wavelet packet Shannon entropy (WPSE). They proved WPTE is the extension of WPSE, i.e., WPSE is a particular case of WPTE. In this study, we employed WPTE to extract features. Sheejakumari and Gomathi (2015) [30] proposed an improved PSO and used neural network, in order to classify healthy and pathological tissues. Dong, et al. (2015) [31] used stationary wavelet transform (SWT) and PCA. Hemanth, et al. (2014) [32] used iteration-free artificial neural network for abnormal brain image classification. Zhang, et al. (2015) [33] tested wavelet packet Tsallis entropy (WPTE) and fuzzy SVM (FSVM).

Nevertheless, the classification accuracy of above methods do not come across realistic requirement (high accuracy and fast detection speed), those methods can still be enhanced [34]. Yang, et al. (2015) [35] proposed a novel feature and named it as fractional Fourier entropy (FRFE). Their method had been proven to be better than most of existing PBDSs. This study continued to use FRFE.

In addition, multilayer perceptron (MLP) belongs to feedforward neural network (FNN), and MLP has obtained successful applications in various fields. In this study, we proposed two improvements for MLP. We compared three pruning techniques and introduced a relatively new algorithm to train its weights and biases.

The structure of this paper is organized as below: Section 2 provides the materials. Section 3 shows how to extract features by FRFE. Section 4 describes the mechanism of MLP and presented two improvements. Section 5 shows the experimental results and discussions. Section 6 concludes the paper. The abbreviation is listed in the end of this paper.

Materials

In PBDS, there are three open access dataset, which contain different numbers of brain magnetic resonance (MR) images. Dataset I (D_I) contains 66 brain images, Dataset II (D_II) contains 160 images, and Dataset III (D_III) contains 255 images.

Figure 1 shows the samples of MR brains, which are all T2-weighted and with sizes of 256x256. Here T2-weighted (spin-spin) relaxation is to give better image contrast, so as to show different anatomical structures clearly. Note that all pathological brains in Fig. 1 suffer from structural alternation, which is the basis of the success of our PBDS. Here Meningioma, glioma, sarcoma are of neoplastic disease. AD, AD with VA, PD, HD are of degenerative disease. MS are of inflammatory disease. SDH is of cerebrovascular disease. Therefore, the chosen images are of various types of brain diseases.

Feature extraction

Fractional fourier transform

Suppose we have a function $x(t)$, we have its a-angle fractional Fourier transform (FRFT) F as:
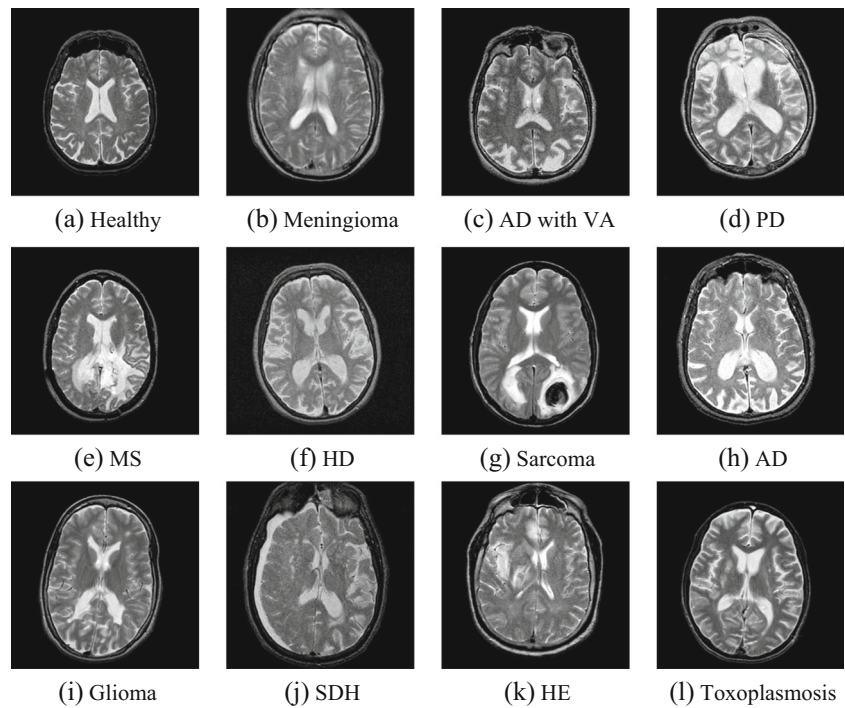
$$F_a(u) = \int_{-\infty}^{\infty} x(t) Z\left(t, u \middle| a\right) \mathrm{d}t \tag{1}$$

here t represents the time, and u denotes the frequency. Z is defined as the transform kernel.

$$Z\left(t, u \middle| a\right) = \sqrt{1 - j \cot a} \times \exp\left(j\pi\left(t^2 \cot a - 2ut \csc a + u^2 \cot a\right)\right) \tag{2}$$

Here j represents the imaginary unit. A problem exists that both cot and csc will diverge for a is set the

Fig. 1 Sample of MR brains. (*AD* Alzheimer's disease, *PD* Pick's disease, *MS* multiple sclerosis, *HE* Herpes encephalitis, *HD* Huntington's disease, *SDH* subdural hematoma, *VA* visual agnosia)

(a) Healthy   (b) Meningioma   (c) AD with VA   (d) PD

(e) MS   (f) HD   (g) Sarcoma   (h) AD

(i) Glioma   (j) SDH   (k) HE   (l) Toxoplasmosis

value of a multiple of π. By taking knowledge from the limitation, equation (2) can be transformed to [36]

$$
Z\big(t,u\big|a\big)=\begin{cases}
\dfrac{\mathbb{D}\ (t-u)}{\sqrt{1-j\cot a}}\times & a\ /\ \pi\ =\ 2\ m \\[4pt]
\exp\begin{pmatrix} j\pi\big(t^2\cot a \\ -2ut\csc a \\ +u^2\cot a\big)\end{pmatrix} & a\ /\ \pi\ \neq\ m \\[4pt]
\mathbb{D}\ (t+u) & a/\pi=(2m+1)\pi
\end{cases}
\tag{3}
$$

where $\mathbb{D}$ represents the Dir0ct delta function and m an arbitrary integer. For 2D FRFT, we have not only angle (denoted by a) for x-axis, but also another angle (denoted by b) for y-axis.

To show the connection between standard Fourier transform (SFT) and FRFT, we showed in Fig. 2 a given rectangular function rect(t) defined as

$$
\mathrm{rect}(t)=\begin{cases}
0 & \big|t\big|>1/2 \\
1/2 & \big|t\big|=1/2 \\
1 & \big|t\big|<1/2
\end{cases}
\tag{4}
$$

In Fig. 2, we present the FRFT results with angles from 0 to 1 with equal increase of 0.1. Remember that the SFT of rect(t) is sinc(u). In this figure, the red line represents the real part while the blue line the imaginary part. It is easily observed that the FRFT result approximate to the SFT result when the value of a increases to 1. This falls within the theoretical prediction.

Another point can be deduce is that adding an extra parameter a can provide more information than SFT does.

Fractional fourier entropy

Yang, et al. (2015) [35] combined FRFT with Shannon entropy, and they proposed a novel image feature based on analysis of MR brain images. They named this new feature as FRFE. Suppose Shannon entropy operation is defined as H, FRFE operation is defined as E, FRFT operation is defined beforehand as F, we have

$$
E=H\cdot F
\tag{5}
$$

Nevertheless, Yang, et al. (2015) [35] used Welch's t-test (WTT) and found only 12 different angle combinations are effective features for brain images. Those angle combinations are listed in Table 1. Therefore, our FRFE followed this setting and we defined it as

$$
E(x)=H\left[\underset{(a,b)\in S}{\cup}F(x)\right]
\tag{6}
$$

here x denotes any brain image (pathological or healthy) and S denotes the angle combination set

$$
S=\Big\{(0.6,1),(0.7,1),(0.8,0.9), \\
(0.8,1.0),(0.9,0.8),(0.9,0.9), \\
(0.9,1.0),(1.0,0.6),(1.0,0.7), \\
(1.0,0.8),(1.0,0.9),(1.0,1,0)\Big\}
\tag{7}
$$

Fig. 2 FRFT of rect function (*a* changes from 0 to 1)
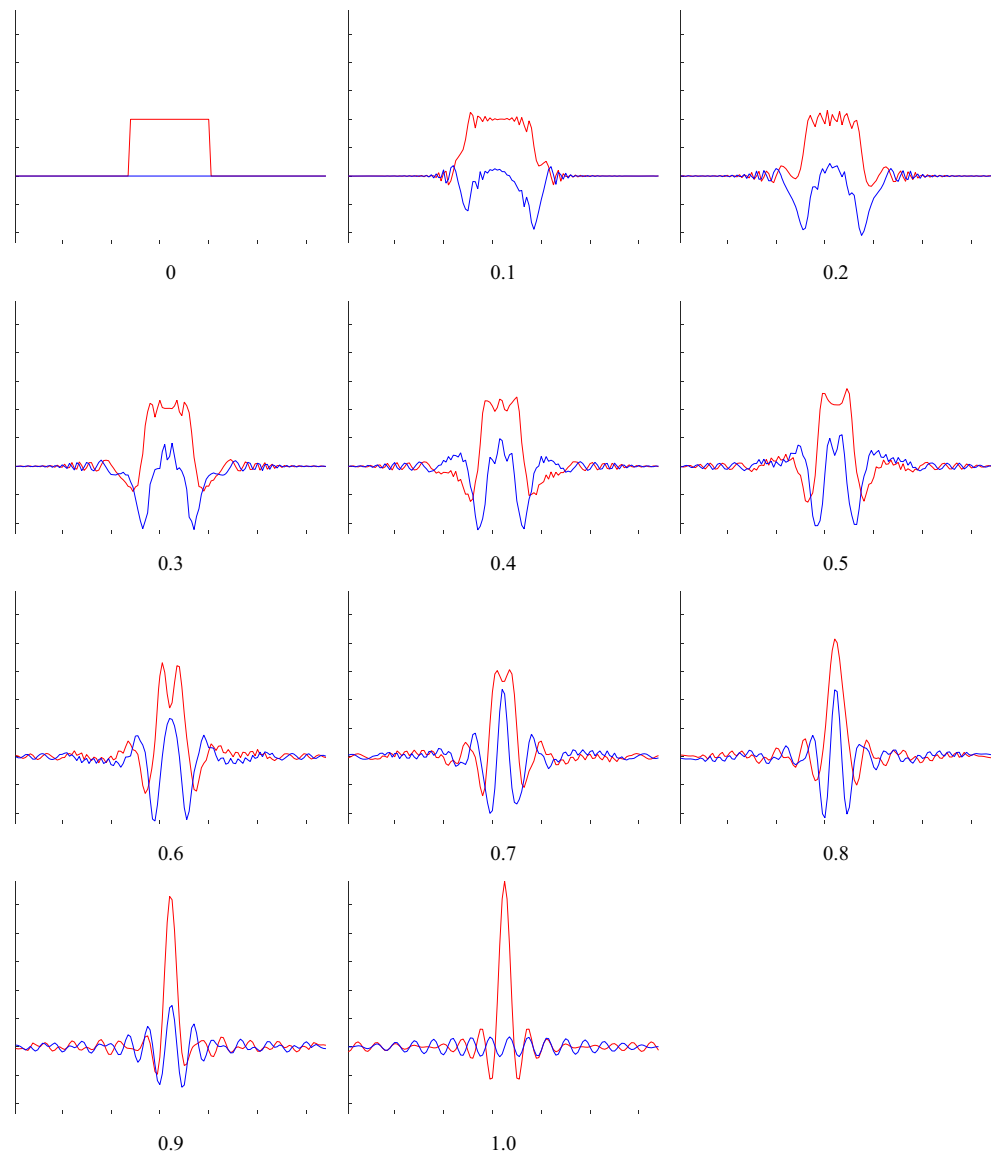
Multi-layer perceptron

A multilayer perceptron (MLP) is a type of neural network, which maps given input data to expected target data. MLP consists of multiple layers of nodes in a directed graph, each layer connecting fully to the next. In this study, we used the common one-hidden layer MLP, so our modal consists of an input layer with $d = 12$ nodes, a hidden layer with unknown neurons with size of M, and an output layer with $c = 1$ neuron with values of either true (denoting pathological) or false (denoting healthy).

For generality (See Fig. 3), suppose $[x(n), t(n)]$ denotes the n-th training sample, where $x(n) = [x_1(n), x_2(n), …, x_d(n)]^T$ $(n = 1, 2, …, N)$ denotes the input vector with d-dimension, and $t(n) = [t_1(n), t_2(n), …, t_c(n)]^T$ the target of c-dimension. The training of MLP is an

optimization problem of minimizing the sum of mean-squared error (MSE) E between the target $t_k(n)$ and realistic output $y_k(n)$.

$$E = \sum_{n=1}^{N} \sum_{k=1}^{c} (y_k(n) - t_k(n))^2 \tag{8}$$

Assume g is the activation function in hidden layer, k the dimension of target, h the activation function in output layer, A the weights connecting the input to hidden layers and B the weights connecting hidden to output layers, we have

$$y_k(n) = h\left(\sum_{j=0}^{M} B_{kj} z_j(n) \right) \tag{9}$$

Table 1 Angle combination of FRFE for brain images

Combination	a	b
1	0.6	1
2	0.7	1
3	0.8	0.9
4	0.8	1.0
5	0.9	0.8
6	0.9	0.9
7	0.9	1.0
8	1.0	0.6
9	1.0	0.7
10	1.0	0.8
11	1.0	0.9
12	1.0	1.0

where $z_j(n)$ represents the output of j-th neuron in the hidden layer and $j = 1, 2, …, M$, with definition of

$$z_j(n) = g\left(\sum_{i=0}^{d} A_{ji}x_i(n)\right) \quad (10)$$

Nevertheless, traditional MLP suffers from two shortcomings: (i) it is difficult to determine the optimal hidden neuron number; and (ii) the weight training may be trapped into local minimum points. To solve above problems, we make two improvements in this work.

Pruning technique

The first major problem is to define the number of hidden neurons. One popular technique is pruning technique (PT) that forces hidden neuron number to be more than necessary, however, this will leads to a sparsely-connected network with most
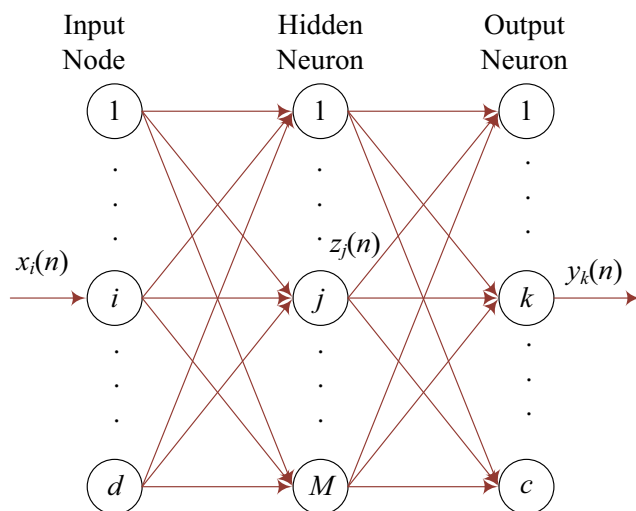

Fig. 3 Structure of one-hidden-layer MLP

weights near-zero. Hence, iterative methods were proposed that removes a neuron with the lowest (or the largest) score in each step, until the error estimation increases. In what below, we will introduce how to define error estimation e and the score function S.

Error estimation

Apparent rate error (APER) was used as the error estimation e. It can be obtained directly from the confusion matrix. Suppose n_{ij} is at the i-th row and j-th column in the confusion matrix, and obviously n_{ij} denotes the sample number of class i predicted to class j, then we have

$$e_{APER} = \frac{\left(\sum_{i=1}^{c}\sum_{j=1}^{c} n_{ij} - \sum_{i=1}^{c} n_{ii}\right)}{\sum_{i=1}^{c}\sum_{j=1}^{c} n_{ij}} \quad (11)$$

APER presents the proportion in percentage of the incorrectly classified samples. Nevertheless, APER tends to underestimate the true rate error due to overfitting. Therefore, Stratified cross validation (SCV) was employed over those datasets. Table 2 shows the statistical characteristics for each dataset. For D_I, it is composed of 18 healthy and 48 pathological brains. Hence, it is common to segment D_I to 6 folds that each fold contains 3 healthy and 8 pathological brains. For D_II that has 20 healthy and 140 pathological brains, we divide it into 5 folds so that each fold consists of 4 healthy and 28 pathological brains. The same thing is performed for D_III.

Measure of hidden neuron

Three measures of hidden neurons were introduced and would be compared in the experiments. Murase, Matsunaga and Nakade (1991) [37] proposed the dynamic pruning (DP) that scores each hidden neuron j with following equation

$$S_j^{DP} = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{c} B_{kj}^2 z_j^2(n) \quad (12)$$

where S represents the score. Silvestre and Lee Luan (2002) [38] proposed a pruning based on Bayesian detection

Table 2 Statistical characteristics

Name	Healthy	Pathological	Fold
D_I	18	48	6
D_II	20	140	5
D_III	35	220	5

boundary (BDB). The measure is similar to equation (12) except for the drop of quadratic terms as

$$S_j^{BDB} = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{c} B_{kj} z_j(n) \qquad (13)$$

Based on Kappa coefficient (KC), Silvestre and Ling (2014) [39] proposed a relatively new measure method. Usually a higher KC indicates a better classifier. In extreme cases, a zero value of KC means the success of the classifier is by chance, and a value of one indicates perfect classification. KC is defined as

$$S^{KC} = \frac{M \sum_{k=1}^{c} n_{kk} - \sum_{k=1}^{c} n_{k\bullet} n_{\bullet k}}{N^2 - \sum_{k=1}^{c} n_{k\bullet} n_{\bullet k}} \qquad (14)$$

Here $n_{k\bullet}$ represents the sum of k-th row of the confusion matrix, and $n_{\bullet k}$ the sum of k-th column of confusion matrix. The definition of KC of k-th neuron is without neuron k within the network, by deleting all the weights those are linking to neuron k. Finally, the neuron that has the largest KC should be removed, since the network without its present has the best performance.

Training method

The second major problem is to determine its optimal weights. Traditionally, back-propagation (BP) was the most common method to train MLP [40–42]. During the last decade, swarm intelligence methods were employed to train MLP, such as genetic algorithm (GA) [43], improved hybrid GA [44], bacterial chemotaxis optimization [45], particle swarm optimization [46], and cuckoo optimization [47]. Biogeography-based optimization (BBO) [48] was a novel swarm intelligence method and had been reported to present superior performance to other swarm intelligence approaches.

Theory of BBO

The Biogeography-based optimization (BBO) was proposed to solve optimization problems, based on the study of geographical distribution of species. It has three main operators: migration, mutation, and elitism [49]. The objective function is transformed as habitat suitability index (HSI), and the search space is transformed as suitability index variables (SIV) [50].

Migration Migration modifies each individual in the habitat at random. Suppose s denotes the species number, S the maximum number of species, then the emigration rate (a) and immigration rate (b) have a connection as

$$b(s) = B \times \frac{1-s}{S} \qquad (15)$$

$$a(s) = \frac{A \times s}{S} \qquad (16)$$

here A and B represents the maximal values of emigration and immigration possibilities, respectively. In the special case of $A = B$, we have

$$a(s) + b(s) = A = B \qquad (17)$$

Mutation Mutation occurs in the SIV level. Suppose the mutation rate is represented as w, and

$$w(s) = \frac{1-p(s)}{P} \times W \qquad (18)$$

here $p(s)$ represents the solution probability of s. P represents the maximum value of p, and W is the maximum mutation rate. The mutation is implemented by

$$D_i' = D_i + \text{rand}(0, 1) \times (D_{i,\max} - D_{i,\min}) \qquad (19)$$

where D_i represents the decision variable in the search space, and $D_{i,max}$ and $D_{i,min}$ represents the lower and upper bounds of the i-th decision variable.

Elitism Elitism occurs in SIV level as mutation. It aims to keep the best solutions within the ecosystem from mutation operator [51]. Suppose the number of elitism is l, then we perform elitism by taking $b = 0$ for the l elites.

Adaptive real-coded BBO

Real-coded technique has been introduced to improve the performance of BBO. Gong, et al. (2010) [52] extended original BBO and presented a real-coded biogeography-based optimization with mutation (RCBBO). Later, Kumar and Premalatha (2015) [53] introduced adaptive mechanism into RCBBO, and proposed adaptive RCBBO (ARCBBO).

ARCBBO suggested two improvements to improve the standard BBO. First, ARCBBO denotes individuals by real parameter vector; hence, Equation (19) should be modified. Kumar and Premalatha (2015) [53] proposed a probability based Gaussian mutation to improve the convergence characteristics as

$$D_i' = D_i + N(m, \sigma_i^2) \qquad (20)$$

where N represents the Gaussian random number with mean of m and variance of σ^2. m is assigned with the value of zero. Secondly, adaptive mechanism is introduced to the Gaussian

Table 3 Implementation of ARCBBO

Step 1	Parameter Initialization. Initialize the elite number l, the maximum species count S, the maximum migration rate A, the maximum immigration rate B, the maximum iterative number K.
Step 2	Initialization. Generate a set of habitats randomly.
Step 3	Evaluation. Compute habitat suitability index (HSI) values for each habitat
Step 4	Update. For each habitat, update its species number s, emigration rate a, and immigration rate b.
Step 5	Emigration. Adjust the whole ecosystem by migration on the basis of values of a and b.
Step 6	Mutation. Adaptive Gaussian mutation was performed on the worst half set of ecosystem.
Step 7	Elitism. Force $b = 0$ for the l elites
Step 8	Output. Offer the best habitat if termination criterion is reached. Otherwise, go to Step 3.

mutation, in order to improve the worst half population set by changing σ_i adaptively:

$$\sigma_i = \beta(k) \times \sum_{i=1}^{n} \left(\frac{F_i}{f_{\min}} \right) \times \left(D_{i,\max} - D_{i,\min} \right) \qquad (21)$$

where $f_{\min}$ represents the minimum fitness value among the whole ecosystem. F_i represents the fitness value of i-th habitat. $\beta(k)$ represents an adaptive parameter at k-th iteration with the form of

$$\beta(k) = 1 - \frac{0.995}{K} \times k \qquad (22)$$

where K denotes the maximum iterative number. Note that the above adaptive mutation is only for mutation operator. For the ecosystem initialization, we still use random generator. In all, Table 3 shows the pseudocode of ARCBBO. We divide it into eight steps.

Results and discussions

Our PBDS is composed of four parts: FRFE, MLP, PT, and ARCBBO. For the PT, we introduced three different measures
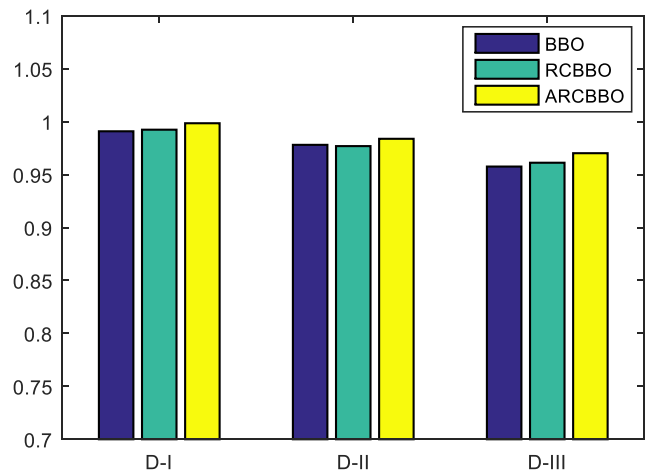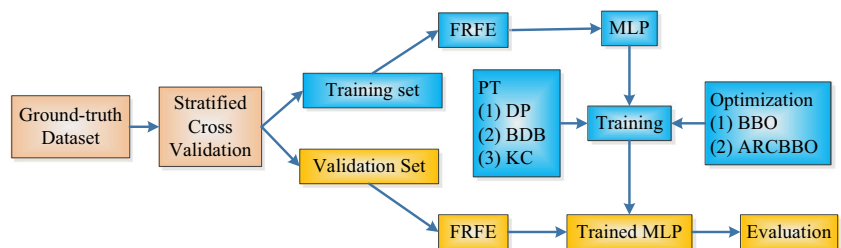

Fig. 5 Comparison among BBO, RCBBO, and ARCBBO (No Pruning Technique)

of DP, BDB, and KC, respectively. Figure 4 shows the diagram of this proposed PBDS.

For statistical analysis, the stratified cross validation (SCV) was used [54]. 6-fold SCV was employed for D_I, and then 5-fold SCV was used for D_II and D_III. Here pathological (P) brains were assigned as true, and the healthy (H) brains were assigned as false. The experiment all run 10 times. The effectiveness of FRFT and FRFE were already reported in [35].

ARCBBO versus BBO and RCBBO

In the first experiment, we compared ARCBBO with standard BBO and RCBBO. We set the hidden neuron number to 20, and no pruning technique was employed. The average accuracy of 10 runs of K-fold SCV was listed in Fig. 5.

The average accuracies of D_I, D_II, and D_III by BBO were 99.09, 97.81, and 95.76 %, respectively. The average accuracies of D_I, D_II, and D_III by RCBBO achieved 99.24, 97.69, and 96.12 %, respectively. What is more, using ARCBBO, the average accuracies of D_I, D_II, and D_III were increased to 99.85, 98.38, and 97.02 %, respectively.

Figure 5 shows that the comparison among the accuracy results obtained by BBO, RCBBO and ARCBBO validated that the ARCBBO is more effective in training MLP for PBDS than both BBO and RCBBO. The reason is because real-coded and adaptive mechanism in ARCBBO can improve
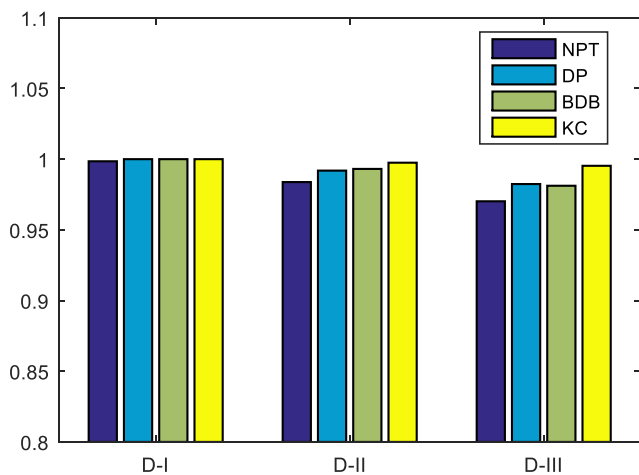
Fig. 4 Diagram of our PBDS

Fig. 6 Pruning technique comparison (*NPT* no pruning technique)

population diversity and exploration ability, hence, leading to better convergence and robustness than BBO. In the future, we will also test other improved BBO variants, such as multiobjective BBO [55], grouping BBO [56], optimal integrated BBO [57], etc.

Pruning technique comparison

In this second experiment, we compared there different pruning techniques (PTs). ARCBBO was chosen since it has been proven to have better performance than BBO in Section 5.1. The comparison was based on 10 repetitions of K-fold SCV. APER was chosen as the error estimation. The results were shown in Fig. 6.

Here, NPT denotes no pruning technique. For D_I, the NPT, DP, BDB and KC approach obtains average accuracy of 99.85, 100.00, 100.00, and 100.00 %, respectively. For D_II, the NPT, DP, BDB and KC approach obtains average accuracy of 98.38, 99.19, 99.31, and 99.75 %, respectively. For D_III, the

DP, BDB and KC approach obtain average accuracy of 97.02, 98.24, 98.12, and 99.53 %, respectively.

The pruning technique comparison in Fig. 6 suggests that using pruning technique will get better performance than not using pruning technique. The reason is the MLP will contains plenty of near-zero weights and biases if just assign a large hidden neuron number, and thus exists overfitting for validation sets. After employed pruning technique, the unnecessary neuron number will be removed, and thus overfitting will be avoided. The comparison also demonstrates KC method is superior to DP and BDB methods.

The best proposed approach

From above, we will know the best proposed approach is "FRFE + KC-MLP + ARCBBO". In this section, we report in Table 4 the classification results of each run and each fold over the largest dataset D_III. Here we can see in the first run, for example, our algorithm successes in predicting 50 instances in Fold 1, and all 51 instances for other four folds. Hence, our algorithm achieves an accuracy of 99.61 % for the first run. Summarizing all 10 runs, the average accuracy of our algorithm is 99.53 %.

Classifier comparison

In the fourth experiment, we compared the best proposed classifier "KC-MLP + ARCBBO", with native Bayesian classifier (NBC) [35] and support vector machine (SVM) [35]. All methods used FRFE and ran 10 times of K-fold SCV. Table 5 shows the comparison results.

The classifier comparison in Table 5 shows that the proposed KC-MLP + ARCBBO gives better classification performance than both NBC and SVM. This indicates

Table 4 Classification results over D_III

Run	F1	F2	F3	F4	F5	All
1	50 (98.04 %)	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	254 (99.61 %)
2	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	50 (98.04 %)	51 (100.00 %)	254 (99.61 %)
3	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	50 (98.04 %)	51 (100.00 %)	254 (99.61 %)
4	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	50 (98.04 %)	50 (98.04 %)	253 (99.22 %)
5	51 (100.00 %)	50 (98.04 %)	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	254 (99.61 %)
6	51 (100.00 %)	50 (98.04 %)	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	254 (99.61 %)
7	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	50 (98.04 %)	254 (99.61 %)
8	50 (98.04 %)	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	254 (99.61 %)
9	50 (98.04 %)	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	50 (98.04 %)	253 (99.22 %)
10	51 (100.00 %)	51 (100.00 %)	51 (100.00 %)	50 (98.04 %)	50 (98.04 %)	253 (99.22 %)

(*F* fold)

Table 5 Classifier comparison (FRFE were used for all)

Classifier	D_I	D_II	D_III
NBC [35]	97.12 %	95.94 %	95.69 %
SVM [35]	100.00 %	99.69 %	98.98 %
KC-MLP + ARCBBO (Proposed)	100.00 %	99.75 %	99.53 %

that MLP may have a potential to excel NBC and SVM, only if the users can carefully tune its hidden neuron number and training algorithm.

Comparison to state-of-the-art approaches

In the fifth experiment, we compared FRFE + KC-MLP + ARCBBO, with 11 approaches including DWT + PCA + KNN [20], DWT + PCA + SCG-ANN [21], RT + PCA + LS-SVM [22], DWT + PCA + SVM [23], DWT + SE + SWP + PNN [24], PCNN + DWT + PCA + BPNN [25], SWT + PCA + HPA-ANN [26], WE + HMI + GEPSVM + RBF [27], WPTE + GEPSVM [29], SWT + PCA + GEPSVM [31], WPTE + FSVM [33]. Table 6 showed the comparison results together with the feature number. Here we only report the results over D_III, since the other two datasets are too small. The abbreviations can be found in Table 7.

Table 6 shows that the proposed FRFE + KC-MLP + ARCBBO achieved the highest average accuracy of 99.53 %, better than 11 state-of-the-art approaches, such as DWT + PCA + KNN [20] with an average accuracy of 96.79 %, DWT + PCA + SCG-ANN [21] with an average accuracy of 98.82 %, RT + PCA + LS-SVM [22] with an

average accuracy of 99.39 %, DWT + PCA + SVM [23] with an average accuracy of 94.29 %, WE + SWP + PNN [24] with an average accuracy of 98.86 %, PCNN + DWT + PCA + BPNN [25] with an average accuracy of 98.24 %, SWT + PCA + HPA-ANN [26] with an average accuracy of 99.45 %, WE + HMI + GEPSVM + RBF [27] with an average accuracy of 98.63 %, WPTE + GEPSVM [29] with an average accuracy of 99.33 %, SWT + PCA + GEPSVM [31] with an average accuracy of 99.02 %, and WPTE + FSVM [33] with an average accuracy of 99.49 %. The improvements may be small in degree, but it was obtained by 10 repetitions of K-fold SCV. Hence, this improvement of our method is reliable.

Conclusions and future researches

This paper proposed a new PBDS of "FRFE + KC-MLP + ARCBBO". The experiments validated its effectiveness as achieved an average accuracy of 99.53 %. Our contributions lie in three points. We compared three different pruning techniques for MLP and showed KC is the most effective. Besides, we introduced the ARCBBO and proved it give better performance than BBO. Finally, the proposed PBDS is superior to 11 state-of-the-art PBDS methods.

In the future, we will include images obtained by other modalities, such as MRSI [58]. Further, other advanced pruning techniques will be tested. Deep learning [59] will be considered after we obtain enough brain images. Internet of things [60] will be another potential research field to embed this PBDS.

Table 6 Classification comparison over D_III

Existing approaches	Feature Number	Average Accuracy
DWT + PCA + KNN [20]	7	96.79 %
DWT + PCA + SCG-ANN [21]	19	98.82 %
RT + PCA + LS-SVM [22]	9	99.39 %
DWT + PCA + SVM [23]	19	94.29 %
WE + SWP + PNN [24]	3	98.86 %
PCNN + DWT + PCA + BPNN [25]	7	98.24 %
SWT + PCA + HPA-ANN [26]	7	99.45 %
WE + HMI + GEPSVM [27]	14	98.63 %
WPTE + GEPSVM [29]	16	99.33 %
SWT + PCA + GEPSVM [31]	7	99.02 %
WPTE + FSVM [33]	16	99.49 %
FRFE + KC-MLP + ARCBBO (Proposed)	16	r

(Bold means the best)

Compliance with Ethical Standards

Conflict of Interest We have no conflicts of interest to disclose with regard to the subject matter of this paper.

Appendix

Table 7   Abbreviation list

Abbreviation	Definition
ABC	Artificial bee colony
AD	Alzheimer's disease
ANN	Artificial neural network
APER	Apparent rate error
ARCBBO	Adaptive RCBBO
BBO	Biogeography-based optimization
BDB	Bayesian detection boundary
BPNN	Backpropagation neural network
DP	Dynamic pruning
DWT	Discrete wavelet analysis
FNN	Feed-forward neural network
FRFE	Fractional Fourier entropy
FRFT	Fractional Fourier transform
FSVM	Fuzzy SVM
GA	Genetic algorithm
GEPSVM	Generalized eigenvalue proximal SVM
HD	Huntington's disease
HE	Herpes encephalitis
HMI	Hu's moment invariant
HPA	Hybridization of PSO and ABC
HSI	Habitat suitability index
LS	Least square
KC	Kappa coefficient
KNN	K-nearest neighbor
MLP	Multilayer perceptron
MR (I)	Magnetic resonance (imaging)
MS	Multiple sclerosis
MSE	Mean-squared error
NBC	Native Bayesian classifier
NPT	No pruning technique
PACS	Picture archiving and communication 1system
PBDS	Pathological brain detection system
PCA	Principal component analysis
PCNN	Pulse coupled neural network
PD	Pick's disease
PNN	Probabilistic neural network
PT	Pruning technique
RCBBO	Real-coded BBO
RT	Ripplet transform
SCG	Scaled conjugate gradient
SCV	Stratified cross validation
SDH	Subdural hematoma
SIV	Suitability index variables
SWP	Spider web plot
SWT	Stationary wavelet transform
SVM	Support vector machine
VA	Visual agnosia
WE	Wavelet entropy
WPTE	Wavelet packet Tsallis entropy

References

1. D'Angelino, R. H. R., Pituco, E. M., and Villalobos, E. M. C. et al., Detection of bovine leukemia virus in brains of cattle with a neurological syndrome: pathological and molecular studies. *Biomed. Res. Int.* 6, 2013.
2. Zhang, Y., Wang, S., Dong, Z., et al., Pathological brain detection in magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimization. *Prog. Electromagn. Res.* 152:41–58, 2015.
3. Yasmin, M., Sharif, M., Mohsin, S., et al., Pathological brain image segmentation and classification: a survey. *Curr. Med. Imag. Rev.* 10(3):163–177, 2014.
4. Kim, S. J., Kim, S. J., Park, J. S., et al., Analysis of age-related changes in Asian facial skeletons using 3D vector mathematics on picture archiving and communication system computed tomography. *Yonsei Med. J.* 56(5):1395–1400, 2015.
5. Floyd, D. M., Trepp, E. R., Ipaki, M., et al., Study of radiologic technologists' perceptions of Picture Archiving and Communication System (PACS) competence and educational issues in Western Australia. *J. Digit. Imaging* 28(3):315–322, 2015.
6. Lee, Y. H., Park, E. H., and Suh, J. S., Simple and efficient method for region of interest value extraction from picture archiving and communication system viewer with optical character recognition software and macro program. *Acad. Radiol.* 22(1):113–116, 2015.
7. Liu, G., Phillips, P., and Yuan, T.-F., Detection of Alzheimer's disease by three-dimensional displacement field estimation in structural magnetic resonance imaging. *J. Alzheimers Dis.* 50(1):233–248, 2016.
8. Yoon, J. H., Lee, J. M., Yu, M. H., et al., Fat-suppressed, three-dimensional T1-weighted imaging using high-acceleration parallel acquisition and a dual-echo Dixon technique for gadoxetic acid-enhanced liver MRI at 3T. *Acta Radiol.* 56(12):1454–1462, 2015.
9. Bianchi, A., Tibiletti, M., Kjorstad, A., et al., Three-dimensional accurate detection of lung emphysema in rats using ultra-short and zero echo time MRI. *NMR Biomed.* 28(11):1471–1479, 2015.
10. Chen, Y., Yang, J., Cao, Q., et al., Curve-like structure extraction using minimal path propagation with back-tracing. *IEEE Trans. Image Process.* 25(2):988–1003, 2016.
11. Zhang, Y., Chen, M. and Huang, D. et al., iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. *Future Genera. Comput. Syst.*
12. Sakalauskas, A., Lauckaite, K., Lukosevicius, A., et al., Computer-aided segmentation of the mid-brain in trans-cranial ultrasound images. *Ultrasound Med. Biol.* 42(1):322–332, 2016.
13. Shanthakumar, P., and Kumar, P. G., Computer aided brain tumor detection system using watershed segmentation techniques. *Int. J. Imaging Syst. Technol.* 25(4):297–301, 2015.
14. Zhang, Y., Qiu, M., and Tsai, C. W. et al., Health-CPS: healthcare cyber-physical system assisted by cloud and big data. *IEEE Syst. J.* PP(99) 1–8, 2015.
15. Kostopoulos, S., Konstandinou, C., Sidiropoulos, K., et al., Assessing the performance of four different categories of histological criteria in brain tumours grading by means of a computer-aided diagnosis image analysis system. *J. Microsc.* 260(1):37–46, 2015.
16. Arakeri, M. P., and Reddy, G. R. M., Computer-aided diagnosis system for tissue characterization of brain tumor on magnetic resonance images. *SIViP* 9(2):409–425, 2015.
17. Zhang, Y., Zhang, D. Q., Hassan, M. M., et al., CADRE: cloud-assisted drug REcommendation service for online pharmacies. *Mobile Netw. Appl.* 20(3):348–355, 2015.
18. Zhang, Y., Peng, B., Liang, Y.-X., et al., Image processing methods to elucidate spatial characteristics of retinal microglia after optic nerve transection. *Sci. Rep.* 6:21816, 2016.
19. Zhang, Y., Wang, S., Phillips, P., et al., Three-dimensional eigenbrain for the detection of subjects and brain regions related with Alzheimer's disease. *J. Alzheimers Dis.* 50(4):1163–1179, 2016.
20. El-Dahshan, E. S. A., Hosny, T., and Salem, A. B. M., Hybrid intelligent techniques for MRI brain images classification. *Digit. Sign. Process.* 20(2):433–441, 2010.
21. Dong, Z., Wu, L., Wang, S., et al., A hybrid method for MRI brain image classification. *Expert Syst. Appl.* 38(8):10049–10053, 2011.

22. Das, S., Chowdhury, M., and Kundu, M. K., Brain MR image classification using multiscale geometric analysis of Ripplet. *Prog. Electromagnet. Res.-Pier* 137:1–17, 2013.

23. Wu, L., An MR brain images classifier via principal component analysis and kernel support vector machine. *Prog. Electromagn. Res.* 130:369–388, 2012.

24. Saritha, M., Paul Joseph, K., and Mathew, A. T., Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. *Pattern Recogn. Lett.* 34(16):2151–2156, 2013.

25. El-Dahshan, E. S. A., Mohsen, H. M., Revett, K., et al., Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm. *Expert Syst. Appl.* 41(11):5526–5545, 2014.

26. Wang, S., Dong, Z., Du, S., et al., Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. *Int. J. Imaging Syst. Technol.* 25(2):153–164, 2015.

27. Sun, P., Wang, S., Phillips, P., et al., Pathological brain detection based on wavelet entropy and Hu moment invariants. *Bio-Med. Mater. Eng.* 26(s1):1283–1290, 2015.

28. Wibmer, A., Hricak, H., Gondo, T., et al., Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *Eur. Radiol.* 25(10):2840–2850, 2015.

29. Dong, Z., Ji, G., and Yang, J., Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with Tsallis entropy and generalized eigenvalue proximal support vector machine (GEPSVM). *Entropy* 17(4):1795–1813, 2015.

30. Sheejakumari, V. and Gomathi B. S., MRI brain images healthy and pathological tissues classification with the aid of improved particle swarm optimization and neural network. *Comput. Math. Methods Med.* 12, 2015.

31. Dong, Z., Liu, A., Wang, S., et al., Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine. *J. Med. Imag. Health Inform.* 5(7):1395–1403, 2015.

32. Hemanth, D. J., Vijila, C. K. S., Selvakumar, A. I., et al., Performance improved iteration-free artificial neural networks for abnormal magnetic resonance brain image classification. *Neurocomputing* 130:98–107, 2014.

33. Zhang, Y.-D., Wang, S.-H., Yang, X.-J., et al., Pathological brain detection in MRI scanning by wavelet packet Tsallis entropy and fuzzy support vector machine. *SpringerPlus* 4(1):716, 2015.

34. Zhang, Y., Chen, M., Mao, S. W., et al., CAP: community activity prediction based on big data analysis. *IEEE Netw.* 28(4):52–57, 2014.

35. Yang, X., Sun, P., Dong, Z., et al., Pathological brain detection by a novel image feature—fractional fourier entropy. *Entropy* 17(12): 7877, 2015.

36. Atangana, A., Jafari, H., and Belhaouari, S. B. et al., Partial fractional equations and their applications. *Math. Problems Eng.* 1, 2015.

37. Murase, K., Matsunaga, Y., and Nakade, Y., A backpropagation algorithm which automatically determines the number of association units. *Neural Netw. 1991. 1991 I.E. Int. Joint Conf.* 1:783–788, 1991.

38. Silvestre, M. R., and Lee Luan, L., Optimization of neural classifiers based on Bayesian decision boundaries and idle neurons pruning. *Pattern Recognit., 2002. Proc. 16th Int. Conf.* 3:387, 2002.

39. Silvestre, M. R., and Ling, L. L., Pruning methods to MLP neural networks considering proportional apparent error rate for classification problems with unbalanced data. *Measurement* 56:88–94, 2014.

40. Khan, Y., Partial discharge pattern analysis using PCA and back-propagation artificial neural network for the estimation of size and position of metallic particle adhering to spacer in GIS. *Electr. Eng.* 98(1):29–42, 2016.

41. Nejad, H. C., Farshad, M., Rahatabad, F. N., et al., Gradient-based back-propagation dynamical iterative learning scheme for the neuro-fuzzy inference system. *Expert. Syst.* 33(1):70–76, 2016.

42. Lin, B. S., Wu, H. D., and Chen, S. J., Automatic wheezing detection based on signal processing of spectrogram and back-propagation neural network. *J. Healthcare Eng.* 6(4):649–672, 2015.

43. Oghaz, M. M., Maarof, M. A., Zainal, A., et al., A hybrid color space for skin detection using genetic algorithm heuristic search and principal component analysis technique. *Plos One* 10(8):21, 2015.

44. Lu, S., Wang, S., and Zhang, Y., A note on the weight of inverse complexity in improved hybrid genetic algorithm. *J. Med. Syst.* 40(6):1–2, 2016.

45. Zhang, Y., and Wu, L., Weights optimization of neural network via improved BCO approach. *Prog. Electromagn. Res.* 83:185–198, 2008.

46. Ji, G., A comprehensive survey on particle swarm optimization algorithm and its applications. *Math. Probl. Eng.* 2015:38, 2015.

47. Bayati, M., Using cuckoo optimization algorithm and imperialist competitive algorithm to solve inverse kinematics problem for numerical control of robotic manipulators. *Proc. Instit. Mech. Eng. Part I-J. Syst. Contrl. Eng.* 229(5):375–387, 2015.

48. Ji, G., Yang, J., Wu, J., et al., Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization. *Entropy* 17(8):5711–5728, 2015.

49. Ma, H. P., Fei, M. R., and Yang, Z. L., Biogeography-based optimization for identifying promising compounds in chemical process. *Neurocomputing* 174:494–499, 2016.

50. Li, B. X., and Low, K. S., Low sampling rate online parameters monitoring of DC-DC converters for predictive-maintenance using biogeography-based optimization. *IEEE Trans. Power Electron.* 31(4):2870–2879, 2016.

51. Ma, H. P., Su, S. F., Simon, D., et al., Ensemble multi-objective biogeography-based optimization with application to automated warehouse scheduling. *Eng. Appl. Artif. Intell.* 44:79–90, 2015.

52. Gong, W. Y., Cai, Z. H., Ling, C. X., et al., A real-coded biogeography-based optimization with mutation. *Appl. Math. Comput.* 216(9):2749–2758, 2010.

53. Kumar, A. R., and Premalatha, L., Optimal power flow for a deregulated power system using adaptive real coded biogeography-based optimization. *Int. J. Electr. Power Energy Syst.* 73:393–399, 2015.

54. Purushotham, S., and Tripathy, B. K., Evaluation of classifier models using stratified tenfold cross validation techniques. In: Krishna, P. V., Babu, M. R., and Ariwa, E. (Eds.), *Global Trends in Information Systems and Software Applications, Pt 2*. Springer-Verlag Berlin, Berlin, pp. 680–690, 2012.

55. Guo, W. A., Wang, L., and Wu, Q. D., Numerical comparisons of migration models for multi-objective biogeography-based optimization. *Inf. Sci.* 328:302–320, 2016.

56. Kim, S. S., Byeon, J. H., Lee, S., et al., A grouping biogeography-based optimization for location area planning. *Neural Comput. Appl.* 26(8):2001–2012, 2015.

57. Yosef, M., Sayed, M. M., and Youssef, H. K. M., Allocation and sizing of distribution transformers and feeders for optimal planning of MV/LV distribution networks using optimal integrated biogeography based optimization method. *Electr. Power Syst. Res.* 128: 100–112, 2015.

58. Dong, Z., Zhang, Y., Liu, F., et al., Improving the spectral resolution and spectral fitting of 1H MRSI data from human calf muscle by the SPREAD technique. *NMR Biomed.* 27(11):1325–1332, 2014.

59. Rusk, N., Deep learning. *Nat. Methods* 13(1):35–35, 2016.

60. Ma, Y. J., Zhang, Y., Dung, O. M., et al., Health internet of things: recent applications and outlook. *J. Internet Technol.* 16(2):351–362, 2015.