Pathological Brain Detection in Magnetic Resonance Imaging Scanning by Wavelet Entropy and Hybridization of Biogeography-Based Optimization and Particle Swarm Optimization

**Yudong Zhang[1, 5, *], Shuihua Wang[1, 2], Zhengchao Dong[3],
Preetha Phillip[4], Genlin Ji[1], and Jiquan Yang[5]**

Abstract—(Background) We proposed a novel computer-aided diagnosis (CAD) system based on the hybridization of biogeography-based optimization (BBO) and particle swarm optimization (PSO), with the goal of detecting pathological brains in MRI scanning. (Method) The proposed method used wavelet entropy (WE) to extract features from MR brain images, followed by feed-forward neural network (FNN) with training method of a Hybridization of BBO and PSO (HBP), which combined the exploration ability of BBO and exploitation ability of PSO. (Results) The 10 repetition of k-fold cross validation result showed that the proposed HBP outperformed existing FNN training methods and that the proposed WE + HBP-FNN outperformed fourteen state-of-the-art CAD systems of MR brain classification in terms of classification accuracy. The proposed method achieved accuracy of 100%, 100%, and 99.49% over Dataset-66, Dataset-160, and Dataset-255, respectively. The offline learning cost 208.2510 s for Dataset-255, and merely 0.053s for online prediction. (Conclusion) The proposed WE + HBP-FNN method achieves nearly perfect detection pathological brains in MRI scanning.

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a low-risk, fast, non-invasive imaging technique that produces high quality images of the anatomical structures of the human body, especially in the brain, and provides rich information for clinical diagnosis and biomedical research [1]. Soft tissue structures are clearer and more detailed with MRI than other imaging modalities [2]. Numerous researches were performed, trying not only to improve the magnetic resonance (MR) image quality, but also to seek novel methods for easier and quicker pre-clinical diagnosis from MR images.

The problem arose that existing manual methods of analysis and interpretation were tedious, time consuming, costly, and irreproducible, due to the huge amount of imaging data. This necessitated the requirement to develop automatic computer-aided diagnosis (CAD) tool. One of the most discriminant features of a normal brain was its symmetry that was obvious in either the axial or the coronal direction. However, the asymmetry along an axial MR brain suggested a pathological brain. This symmetry-asymmetry can be modelled by various image processing techniques and can be used to detect foci as either tumors or neurodegenerative diseases [3]. Note that the CAD is to assist radiologists in the way of providing a second opinion. The final decision is made by radiologists [4].

The aim of this paper was to improve the performances of existing CAD systems for pathological brain detection. The main contribution of this study lied in the improvements of methodology: (i) We used wavelet entropy to replace commonly used discrete wavelet transform. (ii) We used feed-forward neural network (FNN) as classifier, and proposed a novel optimization algorithm based on biogeography-based optimization (BBO) and particle swarm optimization (PSO), in order to train the FNN. (iii) We used k-fold stratified cross validation to enhance the generalization ability of the CAD system.

The structure of the rest is organized as follows. Section 2 summarizes the state-of-the-art methods. Section 3 describes the methodology of the proposed system. Section 4 contains the experiment results. Section 5 discusses the results and the methodology. Section 6 concludes the paper and presents future work. For the ease of reading, we explained the acronyms in Table A1 (please refer to the appendix).

2. LITERATURE REVIEW

2.1. Existing Pathological Brain Detection Systems

In the last decade, various CAD systems were proposed for pathological brain detection. Chaplot et al. [5] used the approximation coefficients obtained by discrete wavelet transform (DWT), and employed the self-organizing map (SOM) neural network and support vector machine (SVM). Maitra and Chatterjee [6] employed the Slantlet transform, which is an improved version of DWT. Their feature vector of each image is created by considering the magnitudes of Slantlet transform outputs corresponding to six spatial positions chosen according to a specific logic. Then, they used the common back-propagation neural network (BPNN). El-Dahshan et al. [7] extracted the approximation and detail coefficients of 3-level DWT, reduced the coefficients by principal component analysis (PCA), and used feed-forward back propagation artificial neural network (FP-ANN) and K-nearest neighbors (KNN) classifiers. Zhang et al. [8] proposed using DWT for feature extraction, PCA for feature reduction, and FNN with scaled chaotic artificial bee colony (SCABC) as classifier. Based on it, Zhang et al. [9] suggested to replace SCABC with scaled conjugate gradient (SCG) method. Ramasamy and Anandhakumar [10] used fast-Fourier-transform based expectation-maximization Gaussian mixture model for brain tissue classification of MR images. Zhang and Wu [11] proposed to use kernel SVM (kSVM), and they suggested three new kernels as homogeneous polynomial (HPOL), inhomogeneous polynomial (IPOL), and Gaussian radial basis (GRB). Saritha et al. [12] proposed a novel feature of wavelet-entropy (WE), and employed spider-web plots (SWP) to further reduce features. Afterwards, they used the probabilistic neural network (PNN). Zhang et al. [13] suggested that removing spider-web-plot yielded the same classification performance. Das et al. [14] proposed to use Ripplet transform (RT) + PCA + least square SVM (LS-SVM), and the 5×5 CV shows high classification accuracies. Kalbkhani et al. [15] modelled the detail coefficients of 2-level DWT by generalized autoregressive conditional heteroscedasticity (GARCH) statistical model, and the parameters of GARCH model are considered as the primary feature vector. Their classifier was chosen as KNN and SVM models. Padma and Sukanesh [16] used combined wavelet statistical texture features, to segment and classify Alzheimer's disease (AD) benign and malignant tumor slices. El-Dahshan et al. [17] used the feedback pulse-coupled neural network for image segmentation, the DWT for features extraction, the PCA for reducing the dimensionality of the wavelet coefficients, and the FBPNN to classify inputs into normal or abnormal. The classification accuracy on both training and test images is 99%, which was significantly good. Zhang et al. [18] used discrete wavelet packet transform (DWPT), and harnessed Tsallis entropy (TE) to replace traditional Shannon entropy (SE) with the aim of obtaining features from DWPT coefficients. Then, they used a generalized eigenvalue proximal SVM (GEPSVM). Zhou et al. [19] used wavelet-entropy as the feature space, then they employed a Naive Bayes classifier (NBC) classification method. Their results over 64 images showed that the sensitivity of the classifier is 94.50%, the specificity 91.70%, the overall accuracy 92.60%. Damodharan and Raghavan [20] combined tissue segmentation and neural network for brain tumor detection. Yang et al. [21] selected wavelet-energy as the features, and introduced biogeography-based optimization (BBO) to train the SVM. Their method reached 97.78% accuracy on 90 T2-weighted MR brain images. Zhang et al. [22] used wavelet-energy (WEnergy) and SVM. Their results only obtained 82.69% accuracy. Nazir et al. [23] suggested to use filters for the removal of noises, and extracted color moments as mean features. Finally, they achieved an overall accuracy of 91.8%. Zhang et al. [24] suggested to use a 3D eigenbrain method to detect subjects and

brain regions related to AD. The accuracy achieved 92.36 ± 0.94. Harikumar and Kumar [25] analyzed the performance of ANN, in classification of medical images using wavelets as feature extractor. Good classification percentage of 96% was achieved using the RBF when Daubechies (db4) wavelet based feature extraction was used. Wang et al. [26] suggested to use stationary wavelet transform (SWT) to replace DWT, and then they proposed a hybridization of PSO and ABC (HPA) algorithm to train the classifier. Zhang et al. [27] proposed a novel classification system that implemented 3D discrete wavelet transform (3D-DWT) to extract wavelet coefficients the volumetric image. The triplets (energy, variance, and Shannon entropy) of all subbands coefficients of 3D-DWT were obtained as feature vector. They used Winner-Takes-All (WTA) strategy for the multiclass SVM. The results showed an overall accuracy of 81.5%.

All above-mentioned methods achieved promising results, nevertheless, most methods suffered from three points:

First, DWT or similar techniques were commonly used in pathological brain detection; however, they required large storage and were computationally expensive. Therefore, we proposed to extract entropy from the wavelet approximation coefficients [28]. Such a reduction leads to less number of data values for following procedures.

Second, their classifier performed well on training image, but poorly on new query images. To address this, we used K-fold stratified cross validation, with the aim of enhancing the generalization ability of the CAD system. In addition, the feed-forward neural network (FNN) [29] was chosen as the classifier because it had advantages that can classify nonlinear separable patterns and approximate an arbitrary continuous function. However, how to find the optimal parameters of FNN is a difficult task because the search algorithms are easily trapped in local extrema. To address this problem, we proposed a novel optimization algorithm that is a hybridization method of both BBO and PSO.

Third, their algorithms were test on a small dataset (66 images) or even smaller, which may undermine the reliability. Therefore, we tested the proposed method on three benchmark datasets, which consisted of 66, 160, and 255 images, respectively.

2.2. Wavelet-Entropy

The discrete wavelet transform (DWT) is a powerful implementation of the wavelet transform (WT) using the dyadic scales and positions [30]. Applying this technique to MR images, there are 4 subband (LL, LH, HH, HL) images at each scale. The subband LL is used for the next 2D-DWT. The LL subband can be regarded as the approximation component of the image, while the LH, HL, and HH subbands can be regarded as the detailed components of the image. As the level of the decomposition increased, we obtained more compact yet coarser approximation components. Thus, DWT provides a simple hierarchical framework for interpreting the image information.

Usually DWT coefficients cost a mass of memory storage; hence, entropy is used to further extract features from DWT coefficients [31]. Entropy S was a statistical measure of randomness, associated with the order of irreversible processes from a traditional point of view. It was then redefined as as a measure of uncertainty regarding the information content of a system as Shannon entropy [18] as $S = -\sum p_i \log_2(p_i)$, where i represents the greylevel of reconstructed coefficient, p_i the probability of greylevel of i. Here we treated the subband coefficients as a grayscale image.

Following the paper of Saritha et al. [12], we used 6-level db4 wavelet, and the set of wavelet entropy (WE) was calculated as the entropy of wavelet approximation components at 6-levels (LL1, LL2, LL3, LL4, LL5, and LL6). The reasons are threefold: First, average entropy of high-level decomposition approximates to zero [12], hence 6-level is appropriate. Next, db4 is good for smoothly varying signals like MR images of brain [5]. Third, approximation coefficients provide more information than detailed coefficients [28].

2.3. Feed-forward Neural Network

After extracting and reducing the features from the MR brain images, we feed them to the classifier. Classification methods consist of two categories [32]: supervised classification and unsupervised classification. Both achieved satisfactory results; however, supervised classification performed better than unsupervised classification in terms of classification accuracy (successful classification rate) [33].

We chose the latter category and used a typical supervised classification method: feedforward neural network (FNN) as the classifier. It can classify nonlinear separable patterns and approximate an arbitrary continuous function [34], it is widely used in pattern classification and does not need any information about the probability distribution and a priori probabilities of different classes [35].

The training vectors (i.e., features) were presented to the FNN in batch mode. The general one-hidden-layer model of FNN consists of a structure of three layers: input layer, hidden layer, and output layer. Nodes of each adjacent layer are connected completely and directly to form the links. Each link has a weighted value that presents the relational degree between two nodes.

The training of FNN can be regarded as an optimization problem. The variables of the optimization problem can be presented as the combination of the connection weight matrix between the input and hidden layers, and between the hidden and output layers. The objective function is written as the average mean squared error (MSE) of the difference between output and target value.

In this study, considering that there are 6 features remaining, the structure of the FNN is set to 6-10-1. The input neurons N_I correspond to the number of features as 6. The number of hidden neurons N_H is set to 10 by information-entropy criteria [36]. The number of output neurons N_O corresponds to the number of classes as 1 (true for pathological brain and false for normal brain).

2.4. Traditional Optimization Methods

There were a variety of optimization methods for training the weights of FNN, such as back-propagation (BP) [37], momentum BP (MBP) [38], genetic algorithm (GA) [39], simulated annealing (SA) [40], artificial bee colony (ABC) [41], BBO [42], particle swarm optimization (PSO) [43]. Among them, the BBO and PSO perform excellent for various applications.

2.4.1. Biogeography-Based Optimization

Biogeography-based optimization (BBO) was inspired by biogeography, which describes speciation and migration of species between isolated habitats, and the extinction of species [44]. Habitats friendly to life are termed to have a high habitat suitability index (HSI), and vice versa. Features that correlate with HSI are called suitability index variables (SIV). Like other bio-inspired algorithms, the SIV and HSI are considered as search space and objective function, respectively [45].

Habitats with high HSI have not only a high emigration rate, but also a low immigration rate, because they already support many species. Species that migrate to this kind of habitat will tend to die even if it has high HSI, because there is too much competition for resources from other species. On the other hand, habitats with low HSI have both a high emigration rate and a low immigration rate; the reason is not because species want to immigrate, but because there is a lot of resources for additional species [46].

Following common convention, we assumed a linear relation between rates and number of species, and gave the definition of the immigration and emigration rates of habitats that contains S species as follows:

$$\lambda_S = I(1 - S/S_{\max}), \quad \mu_S = ES/S_{\max} \tag{1}$$

where the λ and μ represent the immigration and emigration probability, respectively. I and E represent the maximum immigration and emigration rate, respectively. $S_{\max}$ is the maximum number of species which the habitat can support, and S_0 the equilibrium species count. The emigration and immigration rates are used to share information between habitats. Consider the special case $E = I$, we have $\lambda_S + \mu_S = E$.

Next, consider the probability P_S that the habitat contains exactly S species. P_S changes from time t to time $(t + \Delta t)$ as

$$P_S(t + \Delta t) = P_S(t)\left(1 - \lambda_S \Delta t - \mu_S \Delta t\right) + P_{S-1}\lambda_{S-1}\Delta t + P_{S+1}\mu_{S+1}\Delta t \tag{2}$$

The left term of the equation represents the probability of having S species at time $(t + \Delta t)$. The right term of the equation consists of three explanations [42]: (1) There were S species at time t, and no immigration or emigration occurred during Δt; (2) There were $(S - 1)$ species at time t, and only one species immigrated; (3) There were $(S + 1)$ species at time t, and one species emigrated.

Now, we need to discuss how the biogeography theory can be applied to optimization problems. Emigration and immigration rates of each habitat are used to share information among the ecosystem. With modification probability P_{mod}, we modify solutions H_i and H_j in the way that we use the immigration rate of H_i and emigration rate of H_j to decide some SIVs of H_j be migrated to some SIVs of H_i.

Mutation was simulated in the SIV level. Very high and very low HSI solutions are equally improbable, nevertheless, medium HSI solutions are relatively probable. Above idea can be implemented as a mutation rate m, which is inversely proportional to the solution probability P_S.

Elitism was also included in standard BBO, in order to retain the best solutions in the ecosystem. Hence, the mutation approach will not impair the high HSI habitats. Elitism is implemented by setting $\lambda = 0$ for the p best habitats, where p is a predefined elitism parameter. In closing, the pseudocode of BBO is listed in Table 1.

2.4.2. Particle Swarm Optimization

Particle swarm optimization (PSO) performs searching via a swarm of particles that updates from iteration to iteration. To seek the optimal solution, each particle moves in the direction to its previously best (*pbest*) position and the global best (*gbest*) position in the swarm.

$$pbest(i,t) = \underset{k=1,\ldots,t}{\arg\min} \left[f\left(P_i(k)\right) \right], \quad i \in \{1, 2, \ldots, N_P\} \tag{3}$$

$$gbest(t) = \underset{\substack{i=1,\ldots,N_P \\ k=1,\ldots,t}}{\arg\min} \left[f\left(P_i(k)\right) \right] \tag{4}$$

where i denotes the particle index, N_P the total number of particles, t the current iteration number, f the fitness function, and P the position. The velocity V and position P of particles are updated by the following equations.

$$V_i(t+1) = \omega V_i(t) + c_1 r_1 \left(pbest(i,t) - P_i(t)\right) + c_2 r_2 \left(gbest(t) - P_i(t)\right) \tag{5}$$

$$P_i(t+1) = P_i(t) + V_i(t+1) \tag{6}$$

Table 1. Pseudocode of BBO.

Algorithm — BBO	
Step 1	Initialize BBO parameters, which include a problem-dependent method of mapping problem solutions to SIVs and habitats, the modification probability P_{mod}, the maximum species count $S_{\max}$, the maximum migration rates E and I, the maximum mutation rate $m_{\max}$, and elite number p.
Step 2	Initialize a random set of habitats.
Step 3	Compute HSI for each habitat.
Step 4	Computer S, λ, and μ for each habitat.
Step 5	Modify the whole ecosystem by migration based on P_{mod}, λ and μ.
Step 6	Mutate the ecosystem based on mutate probabilities.
Step 7	Implement elitism.
Step 8	If termination criterion was met, output the best habitat, otherwise jump to Step 3.

Table 2. Pseudocode of PSO.

Algorithm — PSO	
Step 1	Initialize particle's position, and the *pbest*, *gbest*, and velocity.
Step 2	Repeat until a termination criteria is met
	a. Pick random numbers: r_1, $r_2 \sim U(0, 1)$.
	b. Update particle's velocity and position;
	c. Update *pbest* and *gbest*;
Step 3	Output *gbest(t)* that holds the best found solution.

where V denotes the velocity and ω the inertia weight used to balance the global exploration and local exploitation. r_1 and r_2 are uniformly distributed random variables within range [0, 1]. c_1 and c_2 are positive constant parameters called "acceleration coefficients" [47].

It is common to set an upper bound for the velocity parameter. "Velocity clamping" [48] was used as a way to limit particles flying out of the search space. Another method is the "constriction coefficient" strategy, proposed by Clerc and Kennedy [49], as an outcome of a a theoretical analysis of swarm dynamic, in which the velocities are constricted too. The pseudocode of PSO is listed in Table 2.

2.5. Biography Particle Swarm Optimization

Mo and Xu [50] proposed a hybridization method of both BBO and PSO, and termed it as biography particle swarm optimization (BPSO). In their method, the particle position strategy was used to modify the habitats not selected for immigration. The pseudocode of BPSO is listed in Table 3.

Table 3. Pseudocode of BPSO.

Algorithm — BPSO
Step 1 Initialize parameters
Step 2 Evaluate the fitness for each individual
Step 3 While termination criteria is not met, do
Step 3.1 Record the previous best and neighborhood best positions of each habitat.
Step 3.2 Map the HIS to number of species S, immigration rate λ, and emigration rate μ.
Step 3.3 If a habitat is selected to be immigrated, migrate randomly selected SIVs
Else update the position of current habitat by particle position strategy
End if
Step 3.4 Mutate the worst half of the population
Step 3.5 Evaluate the fitness of each individual and sort them from best to worst, save the best two habitats
Step 4 End while

3. METHODOLOGY

Unfortunately, the GA, SA, PSO and BBO algorithms all demand expensive computational costs and can still be easily trapped into the local best, hence would probably end up without finding the optimal weights of the FNN.

Recently, hybridization optimization methods were proposed, since they usually combined the advantages of two successful optimization algorithms. In this study, we aimed to combine BBO and PSO, because the former has effective search ability (by migration operations) while its exploration ability is poor [51], and the latter has good exploitation ability but may get stuck into local minima.

3.1. Proposed Hybridization Method

Although Mo and Xu [50] proposed the BPSO method, we proposed a novel hybridization method that can further enhance the optimization performance by introducing a novel variable *elig*. We termed it as Hybridization of BBO and PSO (HBP) algorithm. Two improvements were proposed: (i) we divided the whole population into two sub-populations, one implements PSO and other BBO. (ii) We introduced in the crossover (recombination) operation [52] from GA, to produce a new offspring *elig* from the *elite* of BBO and the *gbest* of PSO. The former improvement combined the exploration ability of BBO and exploitation ability of PSO, while the latter enhanced the variety of population, providing the HBP with better global search and exploitation abilities.

Suppose that $elig(i)$ is the ith component (dimension) of *elig*, and the same for *elite* and *gbest*, then the crossover was implemented as

$$elig(i) = \begin{cases} elite(i) & \text{rand} < P_{elig} \\ gbest(i) & \text{otherwise} \end{cases} \tag{7}$$

Table 4. Pseudocode of HBP.

Algorithm — HBP
Step 1 Initialize the HBP parameters, and determine *gbest* of PSO and *elite* of BBO.
Step 2 Repeat until termination criterion is met
Step 3 Apply the crossover procedure to the *gbest* of PSO and *elite* of BBO to generate *elig*.
Step 4 Update the velocities and positions of the particles, and update the *pbest* of the particles and the *gbest* of the population.
Step 5 Implement immigration, emigration, mutation and elitism of BBO. Determine the *elite* of BBO.
Step 6 Return to Step 2.
Step 7 Output the best of *elig*.
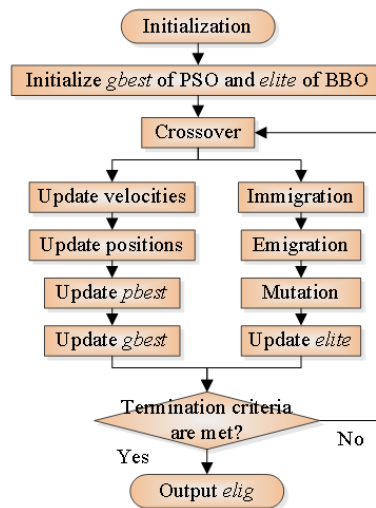
Figure 1. Flowchart of the proposed HBP.

where P_{elig} is defined as the ratio of fitness values of *elite* to the sum of *elite* and *gbest*.

$$P_{elig} = \frac{f_{elite}}{f_{elite} + f_{gbest}} \qquad (8)$$

The pseudocode of HBP is listed in Table 4.

By introducing the variable *elig*, a strong relationship is offered between the *elite* in BBO and the *gbest* in PSO, i.e., the *elig* works as a communication component. From another point of view, the exploitation ability of BBO is improved through the direct use of information provided by *gbest*, meanwhile, the exploration ability of PSO is enhanced through through the *elite* information. The *gbest* of PSO is replaced with a more powerful variable *elig*, which is also given to replace *elite* in BBO. The flowchart of the proposed HBP is displayed in Figure 1.

3.2. Dataset and CV Setting

Three benchmark MR image datasets of different sizes, (Dataset-66, Dataset-160, and Dataset-255), were used in this study. All datasets consist of T2-weighted MR brain images in axial plane and 256×256 in-plane resolution, which were downloaded from the website of Harvard Medical School with URL of http://med.harvard.edu/AANLIB.

Among them, the Dataset-66 and Dataset-160 were already widely used in pathological brain detection. They consist of normal brain images and pathological brain images from seven types of diseases: glioma, meningioma, Alzheimer's disease, Alzheimer's disease plus visual agnosia, Pick's disease, sarcoma, and Huntington's disease.
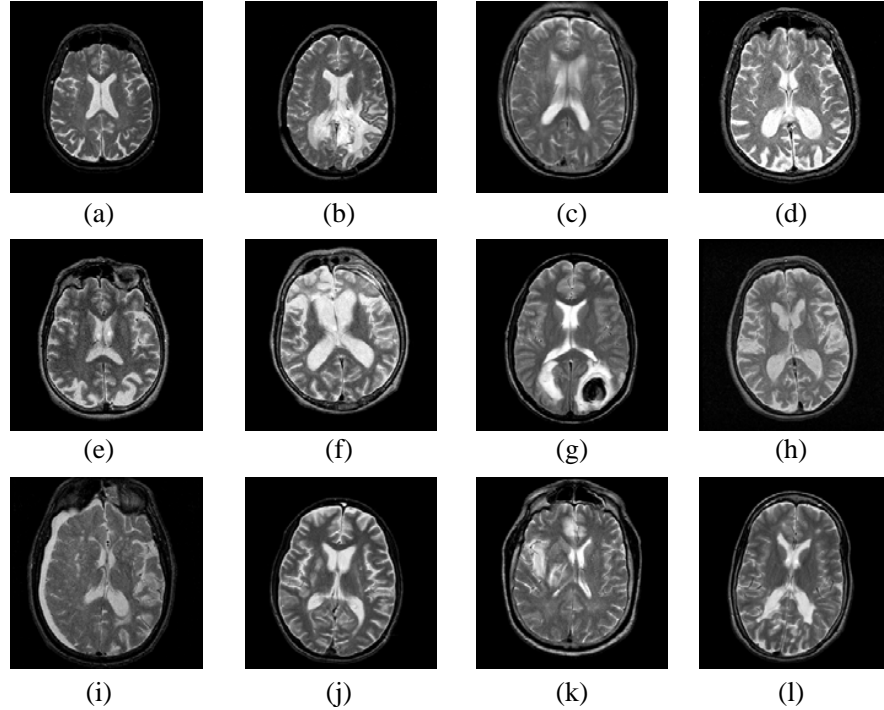
Figure 2. Sample of 1 normal brain and 11 types of diseases. (a) A normal brain. (b) Glioma. (c) Meningioma. (d) AD. (e) AD with visual agnosia. (f) Pick's disease. (g) Sarcoma. (h) Huntington's disease. (i) Chronic subdural hematoma. (j) Cerebral toxoplasmosis. (k) Herpes encephalitis. (l) Multiple sclerosis.

Table 5. Statistical characteristics and CV setting of three datasets.

Dataset	Total		Training		Validation		K-Fold
	Normal	Pathological	Normal	Pathological	Normal	Pathological	
Dataset-66	18	48	15	40	3	8	6-
Dataset-160	20	140	16	112	4	28	5-
Dataset-255	35	220	28	176	7	44	5-

In addition, Das et al. [14] proposed a third dataset "Dataset-255", which contains 11 types of diseases, among which 7 types are the same as above, and 4 new diseases (chronic subdural hematoma, cerebral toxoplasmosis, herpes encephalitis, and multiple sclerosis) were included. Figure 2 shows samples of a normal brain and 11 types of diseases.

Following common convention and ease of stratified cross validation, 6-fold stratified cross validation (CV) was used for Dataset-66, and 5-fold stratified CV was used for the other two datasets. Table 5 shows the statistical characteristics and CV setting of the three datasets used in this experiment. We did not use a 10-fold CV because of the stratification, i.e., we would like to guarantee each fold contains roughly the same proportions of either 7 or 11 diseases.

3.3. Proposed System

Based on FNN and HBP optimization methods, we proposed a novel classifier: HBP-FNN. From another point of view, the proposed systems consists of two stages (Figure 3) as feature extraction by WE, and classification by HBP-FNN. The system is in line with existing classification systems except the feature reduction. Note that the number of extracted features was only 6, and it was unnecessary to include a feature-reduction stage. The implementation of the whole system was two-fold: offline learning with
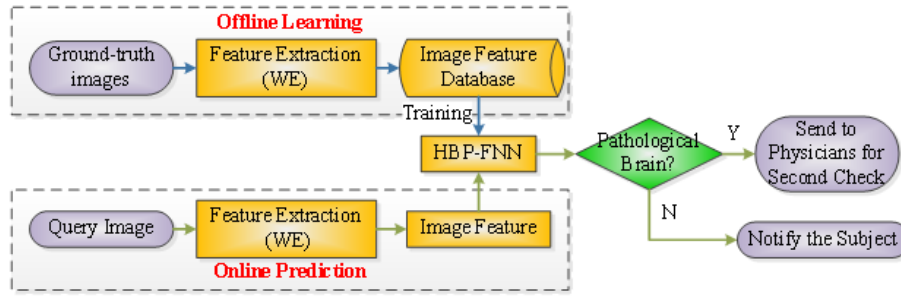
Figure 3. Flowchart of the proposed system.

Table 6. Pseudocodes of the proposed system.

Phase I: Offline learning (users are scientists)
Step 1.   Images were decomposed by DWT, and Entropy was implemented on the approximation coefficients.
a. 6-level db4 wavelet decomposition was performed on each brain image.
b. Entropy was calculated on all approximation components (LL1–LL6)
c. The 6 WEs are stored as the feature database.
Step 2.   The set of WE features along with the corresponding class labels, were used to train the HBP-FNN.
a. Initialize a FNN with structure of 6-10-1.
b. Use k-fold cross validation and feed the feature database to FNN in batch.
c. Train the FNN by the proposed HBP algorithm.
d. Store the trained FNN for online prediction use.
Step 3.   Report the classification performance.
a. Run 10 times since it is a 10-time k-fold cross validation.
b. For one-run, combine the confusion matrix generated by each fold and calculate the accuracy.
c. Average the accuracies of 10 runs and reported the averaged accuracy.
Phase II: Online prediction (users are radiologists)
Step 1.   Users presented to the system the query image to be classified.
a. For a new subject, capture his/her 2D brain image containing the focus.
Step 2.   WE was extracted on every query image, and was input to the previously trained HBP-FNN.
a. Obtain the 6 WEs from the query image by the method in Phase I.
b. Input the 6 WEs to the trained HBP-FNN in Phase I.
Step 3.   The classifier labeled the input query image whether pathological or not.
a. The value from output neuron is binary, either true of false
b. True represents a pathological brain, and false a normal brain.

the aim of training the classifier and online prediction with the aim of detecting existence of diseases for subjects (Table 6). Note that we did not get the size and location of the foci of those diseases.

3.4. Experiment Design

We designed four experiment tasks in this study.

I. We gave a visual illustration of WE features from a normal brain and two types of diseases.

II. We compared the proposed HBP with BP [37], MBP [38], GA [39], SA [40], 2.4.1 [41], BBO [42], PSO [43], and BPSO [50]

III. We compared the proposed HBP-FNN with fourteen state-of-the-art classification methods as DWT + PCA + FP-ANN [7], DWT + PCA + KNN [7], DWT + PCA + SCABC-FNN [8], DWT + PCA + SVM + HPOL [11], DWT + PCA + SVM + IPOL [11], DWT + PCA + SVM + GRB [11], WE + SWP + PNN [12], RT + PCA + LS-SVM [14], PCNN + DWT + PCA + BPNN [17], DWPT

+ SE + GEPSVM [18], DWPT + TE + GEPSVM [18], WE + NBC [19], WEnergy + SVM [22], and SWT + PCA + HPA-FNN [26].

IV. We analyzed the computation time for every step of offline learning and online prediction.

4. RESULTS

The experiments were carried out on the platform of IBM machine with 3 GHz core i3 processor and 8 GB RAM, running under Windows 7 operating system. Matlab 2014a was employed as the develop platform.

4.1. Feature Illustration

First, we carried out 6-level db4 decomposition on a three MR brain images (a normal brain, a Pick's disease, and a Huntington's disease). Table 7 and Table 8 show the approximation components and their corresponding entropy values, respectively.
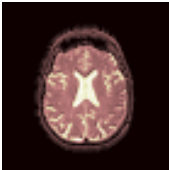
Table 7. The approximation components.
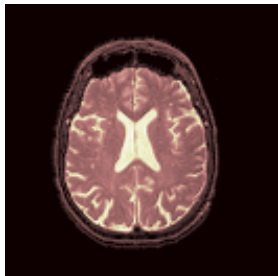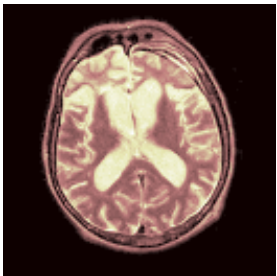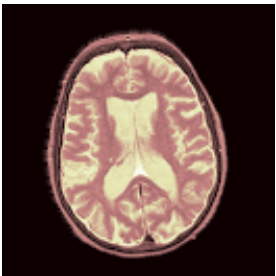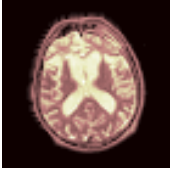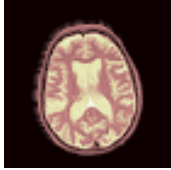
Table 8. WE values of Table 7.

Level	Normal	Pick's	Huntington's
1	3.6039	3.9678	3.7850
2	2.8924	2.5358	2.0946
3	2.1121	1.8152	1.8229
4	1.7614	1.9777	1.9365
5	1.9631	2.4611	2.2714
6	1.9405	2.3301	2.0990

4.2. Optimization Comparison

We compared the proposed HBP method with latest state-of-the-art optimization algorithm, BP [37], MBP [38], GA [39], SA [40], ABC [41], BBO [42], PSO [43], and BPSO [50]. The parameters of these algorithms were obtained using grid-searching method, and are listed in Table 9. For fair comparison, the maximum iterative epochs of all algorithms were all set to 5000. The numbers of population were set to 100. Each algorithm ran 20 times on the training set of MR brain images. The mean and standard deviation (SD) of MSE of all algorithms are listed in Table 10.

Table 9. Parameters of optimization algorithms.

Algorithm	Parameter/Value
BP	learning rate = 0.01,
MBP	learning rate = 0.01, momentum constant = 0.9
GA	population = 100, crossover probability = 0.8, mutation probability = 0.1
SA	population = 100, initial temperature = 100, final temperature = 0, temperature decrease function = "exponential"
ABC	population = 100 (50 employed bees and 50 onlooker bees)
BBO	population = 100, Modification Probability = 0.95, maximum immigration rate = maximum emigration rate = 1, maximum mutation rate = 0.1, elite number = 2
PSO	population = 100, maximal velocity = 1, initial weight = 0.5, acceleration coefficient = 1
BPSO	Population = 100 (50 habitats and 50 particles), other parameters are the combination of parameters of BBO and PSO.
HBP	Population = 100 (50 habitats and 50 particles), other parameters are the combination of parameters of BBO and PSO.

Table 10. Comparison of training algorithms on FNN for MR brain image classification.

Algorithm	Mean of MSE	SD of MSE	Rank
BP [37]	0.1533	0.0693	9
MBP [38]	0.1082	0.0720	8
GA [39]	0.0119	0.0046	6
SA [40]	0.0863	0.0427	7
ANC [41]	0.0072	0.0027	4
BBO [42]	0.0085	0.0035	5
PSO [43]	0.0051	0.0039	3
BPSO [50]	0.0047	0.0025	2
HBP (Proposed)	0.0042	0.0029	1

4.3. Classification Comparison

We compared the proposed method of "WE + HBP-FNN" with fourteen state-of-the-art methods. The full explanations of the abbreviations can be found either in Section 2 or in the appendix. The comparison result is shown in Table 11. Some authors used 5 repetition while other used 10 repetition. In this work, the proposed method used 10 repetitions to get more robust evaluation result.

The number of correct prediction cases and its corresponding accuracy of the proposed WE + HBP-FNN method over Dataset-255 for each fold and each run are listed in Table 12. Note that our method obtained 100.00% accuracy over Dataset-66 and Dataset-160, so their accuracies for each fold and each run are all 100.00%.

Table 11. Average classification accuracy comparison.

Existing approach	# of features	# of runs	Dataset-66	Dataset-160	Dataset-255
*DWT + PCA + FP-ANN [7]	7	5	97.00	96.98	95.29
*DWT + PCA + KNN [7]	7	5	98.00	97.54	96.79
*DWT + PCA + SCABC-FNN [8]	19	5	**100.00**	99.27	98.82
*DWT + PCA + SVM + HPOL [11]	19	5	98.34	96.88	95.61
*DWT + PCA + SVM + IPOL [11]	19	5	**100.00**	98.12	97.73
*DWT + PCA + SVM + GRB [11]	19	5	**100.00**	99.38	98.82
WE + SWP + PNN [12]	3	5	**100.00**	99.88	98.90
*RT + PCA + LS-SVM [14]	9	5	**100.00**	**100.00**	99.39
PCNN + DWT + PCA + BPNN [17]	7	5	**100.00**	98.88	98.43
DWPT + SE + GEPSVM [18]	16	10	99.85	99.62	98.78
DWPT + TE + GEPSVM [18]	16	10	**100.00**	**100.00**	99.33
WE + NBC [19]	7	10	92.58	91.87	90.51
WEnergy + SVM [22]	7	10	82.58	80.13	77.76
SWT + PCA + HPA-FNN [26]	7	10	**100.00**	**100.00**	99.45
Proposed approach	**# of features**	**# of Runs**	**Dataset-66**	**Dataset-160**	**Dataset-255**
WE + HBP-FNN	6	10	**100.00**	**100.00**	**99.49**

(* Some results are obtained in Ref. [14])

Table 12. Results of WE + HBP-FNN over Dataset-255 for each fold and each run (A 10 × 5-fold cross validation).

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Total
Run 1	51 (100.00%)	51 (100.00%)	50 (98.04%)	51 (100.00%)	51 (100.00%)	254 (99.61%)
Run 2	50 (98.04%)	51 (100.00%)	51 (100.00%)	50 (98.04%)	51 (100.00%)	253 (99.22%)
Run 3	51 (100.00%)	51 (100.00%)	51 (100.00%)	51 (100.00%)	51 (100.00%)	255 (100.00%)
Run 4	51 (100.00%)	50 (98.04%)	50 (98.04%)	51 (100.00%)	50 (98.04%)	252 (98.82%)
Run 5	51 (100.00%)	50 (98.04%)	51 (100.00%)	51 (100.00%)	51 (100.00%)	254 (99.61%)
Run 6	51 (100.00%)	51 (100.00%)	51 (100.00%)	51 (100.00%)	51 (100.00%)	255 (100.00%)
Run 7	51 (100.00%)	50 (98.04%)	50 (98.04%)	51 (100.00%)	51 (100.00%)	253 (99.22%)
Run 8	51 (100.00%)	50 (98.04%)	51 (100.00%)	51 (100.00%)	51 (100.00%)	254 (99.61%)
Run 9	51 (100.00%)	51 (100.00%)	51 (100.00%)	51 (100.00%)	50 (98.04%)	254 (99.61%)
Run 10	51 (100.00%)	50 (98.04%)	50 (98.04%)	51 (100.00%)	51 (100.00%)	253 (99.22%)
Total						2537 (99.49%)

Table 13. Computation time analysis.

Offline learning (on Dataset-255)	Time (s)
WE	8.680
HBP-FNN Training	199.571
Online Prediction (on one query image)	**Time (s)**
WE	0.052
FNN Prediction	0.001

4.4. Time Analysis

Finally, computation time was checked as another important measure of the developed system. Table 13 records the consumed time of each step of the proposed "WE + HBP-FNN" method. For the offline learning phase, the two procedures, i.e., WE, and HBP-FNN training, cost 8.680 s and 199.571 s,

respectively. For the online learning phase, the two procedures, i.e., WE, and FNN prediction, cost 0.052 s and 0.001 s, respectively.

5. DISCUSSION

The multi-scale and multi-resolution representations of DWT are obviously shown in Table 7. As the decomposition increases, the size of the approximation coefficients is reduced to approximately 2^{-j} of original size on x and y direction. Here we chose db4 other than the Haar wavelet, because db4 is good for smoothly varying signals as MR images of brain [5]. The values in Table 8 prove the effectiveness of WE, because the WE values of different types have significant differences; hence, the WE is a good informative indicator to detect pathological brains from normal brains, which is coherent with past literature [12]. The reason behind the good performance of WE may fall in its ability to analyze transient features of non-stationary signals, which are abundant in MR brain images.

Results in Table 10 show that the proposed HBP algorithm performs the best, obtaining the least mean MSE of 0.0042. BPSO ranks the second with mean MSE of 0.0047. PSO ranks third with MSE of 0.0051. BBO ranks fifth with MSE of 0.0085. It is unexpected that BBO is worse than ABC, which ranks fourth with MSE of 0.0072. GA and SA rank 6th and 7th, respectively. SA is third worst with MSE of 0.0863, due to its requirement of large iterative epochs [53]. The worst two algorithms are BP and MBP, which were developed to classify linear problems. However, in this study the MR brain classification is a nonlinear problem; hence, the output neurons are working in the saturated nonlinear range, impairing the performance of BP and MBP [54].

From another point of view, this demonstrates the effectiveness of HBP. By using the crossover procedure, a novel variable *elig* is created based on *elite* in BBO and *gbest* in PSO. The proposed HBP not only enhances the variety of the ecosystem (population), but also improves the exploitation ability of BBO and the exploration ability of PSO. The most distinct difference of the proposed HBP with BPSO [50] lies in the proposed new variable *elig* that is the crossover result of *elite* of BBO and *gbest* of PSO. The *elig* is able to enhance the variety of population, providing better global search and exploitation abilities.

It is clear from Table 11 that the proposed "WE + HBP-FNN" obtains the highest accuracy of 100.00%, 100.00%, and 99.49% over Dataset-66, Dataset-160, and Dataset-255, respectively. The results are better than fourteen state-of-the-art algorithms in terms of classification accuracy. The performance of WE + HBP-FNN shows the effectiveness of this technique.

From Table 13, we can calculate that the offline learning costs totally $8.680 + 199.571 = 208.2510$ s. To predict a query image, the computer costs only $0.052 + 0.001 = 0.053$ s, which is rather fast and meets real-time requirement. The offline learning costs 208.2510 s in total, which is rather considerable time. The reason is the training is implemented on 255 images in the dataset. Nevertheless, the prediction procedure does not need to re-train the FNN and can obtain the classification output directly by submitting the WE of query image to the trained FNN.

The proposed method WE + HBP-FNN is similar to WE + SWP + PNN [12], however, we did not use the spider web plot (SWP) method, because we found that procedure did not improve the accuracy in our published work [13]. In addition, we did not choose probabilistic neural network (PNN) because of following three points [54]: (i) FNN can achieve higher accuracy than PNN in general [55]; (ii) PNN is slower than FNN in predicting new cases; and (iii) PNN requires more memory space to store the model.

The proposed method does not contain the image segmentation procedure in [17], because we believe that the image segmentation will impair the information that was contained in original MR image. For example, the maximum gray values will decrease to the number of segments. Another reason is that the goal of image segmentation is to simplify the representation of an image for human eyes. However, computers did not think unsegmented image too complicated. Hence, there is no need to perform the "image segmentation" procedure. The final proof is that most of recent publications about pathological brain detection did not involve segmentation procedure (See Section 2).

The proposed method largely solves the problem raised in the introduction. The contributions of this study center in following five aspects. (i) We used WE that combined wavelet decomposition and entropy to estimate the degree of order/disorder of a signal with a high time-frequency resolution.

(ii) We proposed a novel optimization method as HBP that combined two successful optimization algorithms of BBO and PSO. (iii) We proposed a novel classification method of HBP-FNN. (iv) We proved the HBP performed the best than existing FNN training methods. (iv) We proved the proposed WE + HBP-FNN methods achieved better classification than existing state-of-the-art methods w.r.t. accuracy. (v) We proved the computation time of online prediction of the proposed system reached the requirement of realistic use.

Three limitations are revealed after reconsidering the proposed system. First, FNN established machine-oriented rules not human-oriented rules. It was hard to interpret the model of FNN. Second, how to choose the optimal wavelet and optimal decomposition level remained a question. Third, the computation time of training procedure should be reduced.

6. CONCLUSION AND FUTURE WORK

In this study, we employed WE to replace DWT method for feature extraction, proposed a novel optimization method HBP, and further proposed a novel classification method of HBP-FNN, with the aim of developing a CAD tool for detecting pathological brains. The experiments over three datasets of different sizes demonstrated that the HBP outperformed existing FNN training methods and that the HBP-FNN outperformed state-of-the-art MR brain classification methods in terms of classification accuracy.

Future work should focus on the following four aspects: (i) We shall include other advanced imaging modalities, such as DTI and MRSI [56]; (ii) The classification performance may be increased by using other advanced wavelet families, varying decomposition level, and using other advanced entropy forms. (iii) Additional advanced DWT variants may be tested, such as wavelet packet analysis [57], dual tree complex wavelet transform, second generation wavelet transform, etc. (iv) Novel classification methods will be tested, such as deep leaning, and others.

7. CONFLICT OF INTEREST

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

APPENDIX A.

Table A1. Abbreviation list.

Abbreviation	Definition
(A)(BP)(F)NN	(Artificial) (Back propagation) (Feed-forward) Neural Network
(B)PSO	(Biogeography) Particle swarm optimization
(LS)(k)SVM	(Least square) (kernel) Support vector machine
(M)BP	(Momentum) back-propagation
(SC)ABC	(Scaled chaotic) Artificial bee colony
AD	Alzheimer's disease

Abbreviation	Definition
BBO	Biogeography-based optimization
CAD	Computer-aided diagnosis
DW(P)T	Discrete wavelet (packet) transform
FP-ANN	Feed-forward back propagation ANN
GEPSVM	Generalized eigenvalue proximal SVM
GRB	Gaussian radial basis
H(I)POL	Homogeneous (Inhomogeneous) polynomial
HBP	Hybridization of BBO and PSO
HPA	Hybridization of PSO and ABC
HSI	Habitat suitability index
KNN	K-nearest neighbors
MR(I)	Magnetic resonance (imaging)
MSE	Mean squared error
NBC	Naive Bayes classifier
PNN	Probabilistic neural network
RT	Ripplet transform
SCG	Scaled conjugate gradient
SIV	Suitability index variable
SOM	Self-organizing map
SWP	Spider-web plot
SWT	Stationary wavelet transform

REFERENCES

1. Goh, S., et al., "Mitochondrial dysfunction as a neurobiological subtype of autism spectrum disorder: Evidence from brain imaging," *JAMA Psychiatry*, Vol. 71, No. 6, 665–671, 2014.

2. Zhang, Y., et al., "An MR brain images classifier system via particle swarm optimization and kernel support vector machine," *The Scientific World Journal*, Vol. 2013, 9, 2013.

3. Maji, P., M. K. Kundu, and B. Chanda, "Second order fuzzy measure and weighted co-occurrence matrix for segmentation of brain MR images," *Fundamenta Informaticae*, Vol. 88, Nos. 1–2, 161–176, 2008.

4. Wu, W., et al., "Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features," *International Journal of Computer Assisted Radiology and Surgery*, Vol. 9, No. 2, 241–253, Mar. 2014.

5. Chaplot, S., L. M. Patnaik, and N. R. Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network," *Biomedical Signal Processing and Control*, Vol. 1, No. 1, 86–92, 2006.

6. Maitra, M. and A. Chatterjee, "A Slantlet transform based intelligent system for magnetic resonance brain image classification," *Biomedical Signal Processing and Control*, Vol. 1, No. 4, 299–306, Oct. 2006.

7. El-Dahshan, E. S. A., T. Hosny, and A. B. M. Salem, "Hybrid intelligent techniques for MRI brain images classification," *Digital Signal Processing*, Vol. 20, No. 2, 433–441, Mar. 2010.

8. Zhang, Y., L. Wu, and S. Wang, "Magnetic resonance brain image classification by an improved artificial bee colony algorithm," *Progress In Electromagnetics Research*, Vol. 116, 65–79, 2011.

9. Zhang, Y., et al., "A hybrid method for MRI brain image classification," *Expert Systems with Applications*, Vol. 38, No. 8, 10049–10053, 2011.

10. Ramasamy, R. and P. Anandhakumar, "Brain tissue classification of MR images using fast fourier transform based expectation-maximization Gaussian mixture model," *Advances in Computing and Information Technology*, 387–398, Springer, 2011.

11. Zhang, Y. and L. Wu, "An MR brain images classifier via principal component analysis and kernel support vector machine," *Progress In Electromagnetics Research*, Vol. 130, 369–388, 2012.

12. Saritha, M., K. P. Joseph, and A. T. Mathew, "Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network," *Pattern Recognition Letters*, Vol. 34, No. 16, 2151–2156, Dec. 2013.

13. Zhang, Y., et al., "Effect of spider-web-plot in MR brain image classification," *Pattern Recognition Letters*, Vol. 62, 14–16, Sep. 1, 2015.

14. Das, S., M. Chowdhury, and M. K. Kundu, "Brain MR image classification using multiscale geometric analysis of ripplet," *Progress In Electromagnetics Research*, Vol. 137, 1–17, 2013.

15. Kalbkhani, H., M. G. Shayesteh, and B. Zali-Vargahan, "Robust algorithm for brain magnetic resonance image (MRI) classification based on GARCH variances series," *Biomedical Signal Processing and Control*, Vol. 8, No. 6, 909–919, 2013.

16. Padma, A. and R. Sukanesh, "Segmentation and classification of brain CT images using combined wavelet statistical texture features," *Arabian Journal for Science and Engineering*, Vol. 39, No. 2, 767–776, Feb. 2014.

17. El-Dahshan, E. S. A., et al., "Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm," *Expert Systems with Applications*, Vol. 41, No. 11, 5526–5545, Sep. 2014.

18. Zhang, Y., et al., "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*, Vol. 17, No. 4, 1795–1813, 2015.

19. Zhou, X., et al., "Detection of pathological brain in MRI scanning based on wavelet-entropy and naive Bayes classifier," *Bioinformatics and Biomedical Engineering*, Vol. 9043, 201–209, F. Ortuño and I. Rojas, Eds., Springer International Publishing, Granada, Spain, 2015.

20. Damodharan, S. and D. Raghavan, "Combining tissue segmentation and neural network for brain tumor detection," *International Arab Journal of Information Technology*, Vol. 12, No. 1, 42–52, Jan. 2015.

21. Yang, G., et al., "Automated classification of brain images using wavelet-energy and biogeography-based optimization," *Multimedia Tools and Applications*, 1–17, May 1, 2015.

22. Zhang, G.-S., et al., "Automated classification of brain MR images using wavelet-energy and support vector machines," *Proceedings of the 2015 International Conference on Mechatronics, Electronic, Industrial and Control Engineering*, C. Liu, et al. (eds.), 683–686, Atlantis Press, USA, 2015.

23. Nazir, M., F. Wahid, and S. A. Khan, "A simple and intelligent approach for brain MRI classification," *Journal of Intelligent & Fuzzy Systems*, Vol. 28, No. 3, 1127–1135, 2015.

24. Zhang, Y., et al., "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," *Frontiers in Computational Neuroscience*, Vol. 66, No. 9, 1–15, 2015.

25. Harikumar, R. and B. V. Kumar, "Performance analysis of neural networks for classification of medical images with wavelets as a feature extractor," *International Journal of Imaging Systems and Technology*, Vol. 25, No. 1, 33–40, Mar. 2015.

26. Wang, S., et al., "Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection," *International Journal of Imaging Systems and Technology*, Vol. 25, No. 2, 153–164, 2015.

27. Zhang, Y., et al., "Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC," *Biomedical Signal Processing and Control*, Vol. 21, 58–73, Aug. 2015.

28. Yildiz, A., et al., "Application of adaptive neuro-fuzzy inference system for vigilance level estimation by using wavelet-entropy feature extraction," *Expert Systems with Applications*, Vol. 36, No. 4, 7390–7399, May 2009.

29. Zhang, Y.-D. and L. Wu, "Weights optimization of neural network via improved BCO approach," *Progress In Electromagnetics Research*, Vol. 83, 185–198, 2008.

30. Fang, L., L. Wu, and Y. Zhang, "A novel demodulation system based on continuous wavelet transform," *Mathematical Problems in Engineering*, Vol. 2015, 9, 2015.

31. Yang, Y. H., et al., "Wavelet kernel entropy component analysis with application to industrial process monitoring," *Neurocomputing*, Vol. 147, 395–402, Jan. 2015.

32. Aswathy, S. U., G. G. Deva Dhas, and S. S. Kumar, "A survey on detection of brain tumor from MRI brain images," *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, 871–877, 2014.

33. Guo, D., et al., "Improved radio frequency identification indoor localization method via radial basis function neural network," *Mathematical Problems in Engineering*, 2014.

34. Fuangkhon, P., "An incremental learning preprocessor for feed-forward neural network," *Artificial Intelligence Review*, Vol. 41, No. 2, 183–210, Feb. 2014.

35. Llave, Y. A., T. Hagiwara, and T. Sakiyama, "Artificial neural network model for prediction of cold spot temperature in retort sterilization of starch-based foods," *Journal of Food Engineering*, Vol. 109, No. 3, 553–560, 2012.

36. Ludwig, Jr, O., et al., "Applications of information theory, genetic algorithms, and neural models to predict oil flow," *Communications in Nonlinear Science and Numerical Simulation*, Vol. 14, No. 7, 2870–2885, 2009.

37. Shojaee, S. A., et al., "Prediction of the binary density of the ionic liquids plus water using back-propagated feed forward artificial neural network," *Chemical Industry & Chemical Engineering Quarterly*, Vol. 20, No. 3, 325–338, Jul.–Sep. 2014.

38. Karmakar, S., G. Shrivastava, and M. K. Kowar, "Impact of learning rate and momentum factor in the performance of back-propagation neural network to identify internal dynamics of chaotic motion," *Kuwait Journal of Science*, Vol. 41, No. 2, 151–174, May 2014.

39. Chandwani, V., V. Agrawal, and R. Nagar, "Modeling slump of ready mix concrete using genetic algorithms assisted training of artificial neural networks," *Expert Systems with Applications*, Vol. 42, No. 2, 885–893, Feb. 2015.

40. Manoochehri, M. and F. Kolahan, "Integration of artificial neural network and simulated annealing algorithm to optimize deep drawing process," *International Journal of Advanced Manufacturing Technology*, Vol. 73, Nos. 1–4, 241–249, Jul. 2014.

41. Awan, S. M., et al., "An efficient model based on artificial bee colony optimization algorithm with neural networks for electric load forecasting," *Neural Computing & Applications*, Vol. 25, Nos. 7–8, 1967–1978, Dec. 2014.

42. Simon, D., "Biogeography-based optimization," *IEEE Transactions on Evolutionary Computation*, Vol. 12, No. 6, 702–713, Dec. 2008.

43. Momeni, E., et al., "Prediction of uniaxial compressive strength of rock samples using hybrid particle swarm optimization-based artificial neural networks," *Measurement*, Vol. 60, 50–63, Jan. 2015.

44. Christy, A. A. and P. Raj, "Adaptive biogeography based predator-prey optimization technique for optimal power flow," *International Journal of Electrical Power & Energy Systems*, Vol. 62, 344–352, Nov. 2014.

45. Guo, W. A., et al., "Biogeography-based particle swarm optimization with fuzzy elitism and its applications to constrained engineering problems," *Engineering Optimization*, Vol. 46, No. 11, 1465–1484, Nov. 2014.

46. Simon, D., "A probabilistic analysis of a simplified biogeography-based optimization algorithm," *Evolutionary Computation*, Vol. 19, No. 2, 167–188, Summer 2011.

47. Zhang, Y., S. Wang, and Z. Dong, "Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree," *Progress In Electromagnetics Research*, Vol. 144, 171–184, 2014.

48. Shahzad, F., S. Masood, and N. K. Khan, "Probabilistic opposition-based particle swarm optimization with velocity clamping," *Knowledge and Information Systems*, Vol. 39, No. 3, 703–737, Jun. 2014.

49. Clerc, M. and J. Kennedy, "The particle swarm — Explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 1, 58–73, Feb. 2002.

50. Mo, H. W. and L. F. Xu, "Research of biogeography particle swarm optimization for robot path planning," *Neurocomputing*, Vol. 148, 91–99, Jan. 2015.

51. Kim, S. S., et al., "Biogeography-based optimization for optimal job scheduling in cloud computing," *Applied Mathematics and Computation*, Vol. 247, 266–280, Nov. 2014.

52. Kiran, M. S. and M. Gunduz, "A recombination-based hybridization of particle swarm optimization and artificial bee colony algorithm for continuous optimization problems," *Applied Soft Computing*, Vol. 13, No. 4, 2188–2203, Apr. 2013.

53. Sait, S. M., A. T. Sheikh, and A. H. El-Maleh, "Cell assignment in hybrid CMOS/nanodevices architecture using a PSO/SA hybrid algorithm," *Journal of Applied Research and Technology*, Vol. 11, 653–664, Oct. 2013.

54. Zhang, Y., et al., "Remote-sensing image classification based on an improved probabilistic neural network," *Sensors*, Vol. 9, No. 9, 7516–7539, 2009.

55. Zhang, Y., et al., "Fruit classification using computer vision and feedforward neural-network," *Journal of Food Engineering*, Vol. 143, No. 0, 167–177, 2014.

56. Dong, Z., et al., "Improving the spectral resolution and spectral fitting of 1 H MRSI data from human calf muscle by the SPREAD technique," *NMR in Biomedicine*, Vol. 27, No. 11, 1325–1332, 2014.

57. Figlus, T. and M. Stanczyk, "Diagnosis of the wear of gears in the gearbox using the wavelet packet transform," *Metalurgija*, Vol. 53, No. 4, 673–676, Oct.–Dec. 2014.