

# Automated classification of brain images using wavelet-energy and biogeography-based optimization

Gelan Yang<sup>1</sup> · Yudong Zhang<sup>2,3</sup> · Jiquan Yang<sup>3</sup> ·  
Genlin Ji<sup>2</sup> · Zhengchao Dong<sup>4</sup> · Shuihua Wang<sup>2</sup> ·  
Chunmei Feng<sup>3</sup> · Qiong Wang<sup>3</sup>

Received: 15 March 2015 / Revised: 7 April 2015 / Accepted: 20 April 2015 /

Published online: 1 May 2015

© Springer Science+Business Media New York 2015

**Abstract** It is very important to early detect abnormal brains, in order to save social and hospital resources. The wavelet-energy was a successful feature descriptor that achieved excellent performances in various applications; hence, we proposed a novel wavelet-energy based approach for automated classification of MR brain images as normal or abnormal. SVM was used as the classifier, and biogeography-based optimization (BBO) was introduced to optimize the weights of the SVM. The results based on a 5×5-fold cross validation showed the performance of the proposed BBO-KSVM was superior to BP-NN, KSVM, and PSO-KSVM in terms of sensitivity and accuracy. The study offered a new means to detect abnormal brains with excellent performance.

**Keywords** Classification · Pattern recognition · Support vector machine · Magnetic resonance imaging · Biogeography-based optimization

## 1 Introduction

The problem of automatic classification of normal/pathological subjects is of great importance in clinical medicine [33]. Magnetic resonance imaging (MRI) is concerned with soft tissue

---

✉ Yudong Zhang  
zhangyudong@njnu.edu.cn

<sup>1</sup> School of Information Science and Engineering, Hunan City University, Yiyang 413000, China

<sup>2</sup> School of Computer Science and Technology, Nanjing Normal University, Nanjing, Jiangsu 210023, China

<sup>3</sup> Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing, Nanjing, Jiangsu 210042, China

<sup>4</sup> Translational Imaging Division & MRI Unit, Columbia University and New York State Psychiatric Institute, New York, NY 10032, USA

anatomy and generates a large information set and details about the subject's brain condition. There exists a large body of work on using brain MR images for automatic diagnosis [12, 34].

A promising approach for analyzing the obtained images is the wavelet transform which offers the capability of simultaneous feature localization in the time and frequency domains [8, 16]. By applying a single-level wavelet transform to an image, four subbands are usually defined as the low-low (LL), low-high (LH), high-low (HL), and high-high (HH) subbands. The LL provides an approximation to the image, and the other three provide high-frequency coefficients for finer-scale image details representation. The wavelet transform have already been applied successfully in many areas related to computer vision.

In the last decade, scholars have applied wavelet transform and other techniques to detect abnormal brain images. Chaplot, et al. [3] 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 [17] 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. [9] 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 neighbor (KNN) classifiers. Zhang, et al. [29] 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. [30] suggested to replace SCABC with scaled conjugate gradient (SCG) method. Ramasamy and Anandhakumar [23] used fast-Fourier-transform based expectation-maximization Gaussian mixture model for brain tissue classification of MR images. Zhang and Wu [28] proposed to use kernel SVM (KSVM), and they suggested three new kernels as homogeneous polynomial, inhomogeneous polynomial, and Gaussian radial basis. Saritha, et al. [24] 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). Das, et al. [7] proposed to use Ripplet transform (RT) + PCA + least square SVM (LS-SVM), and the  $5 \times 5$  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. Zhang, et al. [31] suggested to use particle swarm optimization (PSO) to train the KSVM, and their result on a 90 image database achieved 97.11 % accuracy. Padma and Sukanesh [21] used combined wavelet statistical texture features, to segment and classify AD benign and malignant tumor slices. El-Dahshan, et al. [10] 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. Zhang, et al. [35] used discrete wavelet packet transform (DWPT), and harnessed Tsallis entropy to obtain features from DWPT coefficients. Then, they used a generalized eigenvalue proximal SVM (GEPSVM) with RBF kernel.

All abovementioned methods achieved promising results; however, there are two problems. (1) Most of them used DWT, the coefficients of which require a large storage memory and (2) The classifier training is not stable, and the accuracy can be improved.

To address these two problems, we gave two potential contributions in this study. First we introduced the wavelet energy (WE), which is a rather novel feature-extraction technique that calculates the energy from the wavelet coefficients. It has already been applied to many academic

and industrial fields [1, 6, 22]. Next, we proposed to use the biogeography-based optimization (BBO) method to train the classifier, with the aim of improving its classification performance.

## 2 Feature extraction

### 2.1 2D-DWT

The two-dimensional DWT (2D-DWT) decomposes an image into several sub-bands according to a recursive process [20, 32]. The 1-level decomposition obtains two kinds of coefficients. One contains LH1, HL1, and HH1, which represent details of the original images. The other is LL1 that corresponds to the approximation of original image [11]. The approximation LL1 is then decomposed into second-level approximation and detail coefficients, and the process is repeated to achieve the desired level of resolution.

The obtained coefficients for the approximation and detail subbands are useful features for texture categorization. The 2D-DWT decomposes an image into spatial frequency components that describe the image texture. In practical, the coefficients are obtained by convolving the image with a bank of filters. Afterwards, selected features are extracted from the coefficients for further processing.

### 2.2 Wavelet-energy

After wavelet transformation is applied on the image, wavelet coefficients from the detail subbands of all the decomposition levels are used to formulate the local wavelet energy feature [16]. The wavelet energy in LL, HL, LH, and HH subbands can be, respectively, defined as:

$$E(LL) = \sum_x \sum_y LL(x,y)^2 \quad (1)$$

$$E(LH) = \sum_x \sum_y LH(x,y)^2 \quad (2)$$

$$E(HL) = \sum_x \sum_y HL(x,y)^2 \quad (3)$$

$$E(HH) = \sum_x \sum_y HH(x,y)^2 \quad (4)$$

where  $E$  represents the wavelet-energy. These energies reflect the strength of the images' details in different subbands. Suppose a 3-level decomposition, the final extracted features are the energies from the 10 subbands including (LL3, LH3, HL3, HH3, LH2, HL2, HH2, LH1, HL1, HH1).

## 3 Classifier

The introduction of support vector machine (SVM) is a landmark of the field of machine learning. The advantages of SVMs include high accuracy, elegant mathematical tractability, and direct geometric interpretation. Recently, multiple improved SVMs have grown rapidly,

among which the kernel SVMs are the most popular and effective. Kernel SVMs have the following advantages [27]: (1) they work very well in practice and have been remarkably successful in such diverse fields as natural language categorization, bioinformatics and computer vision; (2) they have few tunable parameters; and (3) their training often involves convex quadratic optimization. Hence, solutions are global and usually unique, thus avoiding the convergence to local minima exhibited by other statistical learning systems, such as neural networks.

### 3.1 SVM

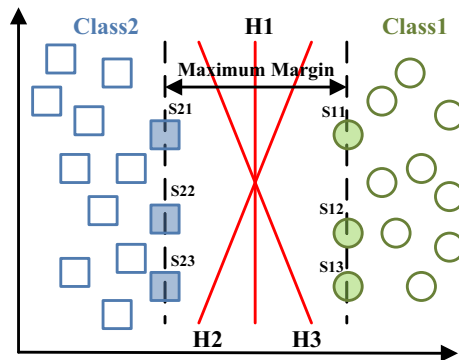
Suppose some prescribed data points each belong to one of two classes, and the goal is to classify which class a new data point will be located in. Here a data point is viewed as a  $p$ -dimensional vector, and our task is to create a  $(p-1)$ -dimensional hyperplane. Many possible hyperplanes might classify the data successfully [2]. One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes, since we could expect better behavior in response to unseen data during training, i.e., better generalization performance [26]. Therefore, we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized. Figure 1 shows the geometric interpolation of linear SVMs, here H1, H2, H3 are three hyperplanes which can classify the two classes successfully, however, H2 and H3 does not have the largest margin, so they will not perform well to new test data. The H1 has the maximum margin to the support vectors (S11, S12, S13, S21, S22, and S23), so it is chosen as the best classification hyperplane.

Given a  $p$ -dimensional  $N$ -size training dataset of the form

$$\left\{ (x_n, y_n) \mid x_n \in \mathbb{R}^p, y_n \in \{-1, +1\} \right\}, n = 1, \dots, N \tag{5}$$

where  $y_n$  is either  $-1$  or  $+1$  corresponds to the class 1 or 2. Each  $x_n$  is a  $p$ -dimensional vector. The maximum-margin hyperplane that divides class 1 from class 2 is the desired SVM. Considering that any hyperplane can be written in the form of

$$\mathbf{w}\mathbf{x} - b = 0 \tag{6}$$



**Fig. 1** The geometric interpolation of linear SVMs (H denotes for the hyperplane, S denotes for the support vector)

We choose the  $\mathbf{w}$  and  $b$  to maximize the margin between the two parallel hyperplanes as large as possible while still separating the data. Hence, we define the two parallel hyperplanes by the equations as

$$\mathbf{w}\mathbf{x}-b = \pm 1 \tag{7}$$

Therefore, the task can be transformed to an optimization problem. That is, we want to maximize the distance between the two parallel hyperplanes, subject to prevent data falling into the margin. Using simple mathematical knowledge, the problem can be finalized as

$$\begin{aligned} &\min_{\mathbf{w},b} \|\mathbf{w}\| \\ &s.t. y_n(\mathbf{w}x_n-b) \geq 1, n = 1, \dots, N \end{aligned} \tag{8}$$

In practical situations the  $\|\mathbf{w}\|$  is usually be replace by

$$\begin{aligned} &\min_{\mathbf{w},b} \frac{1}{2} \|\mathbf{w}\|^2 \\ &s.t. y_n(\mathbf{w}x_n-b) \geq 1, n = 1, \dots, N \end{aligned} \tag{9}$$

The reason leans upon the fact that  $\|\mathbf{w}\|$  is involved to a square root calculation. After it is superseded with formula (9), the solution will not change, but the problem is altered into a quadratic programming optimization that is easy to solve by using Lagrange multipliers and standard quadratic programming techniques and programs.

### 3.2 Soft margin

In practical applications, there may is no hyperplane that splits the samples perfectly. In such case, the “soft margin” method will choose a hyperplane that splits the given samples as cleanly as possible, while still maximizing the distance to the nearest cleanly split samples.

Positive slack vector  $\xi=(\xi_1, \dots, \xi_n, \dots, \xi_N)$  are introduced to measure the misclassification degree of sample  $x_n$  (the distance between the margin and the vectors  $x_n$  that lying on the wrong side of the margin). Then, the optimal hyperplane separating the data can be obtained by the following optimization problem.

$$\begin{aligned} &\min_{\mathbf{w},\xi,b} \frac{1}{2} \|\mathbf{w}\|^2 + c\mathbf{e}^T \xi \\ &s.t. \begin{cases} y_n(\mathbf{w}^T x_n - \mathbf{b}) \geq 1 - \xi_n, n = 1, \dots, N \\ \xi_n \geq 0 \end{cases} \end{aligned} \tag{10}$$

where  $c$  is the error penalty (i.e., box constraint in some other literatures) and  $e$  is a vector of ones of  $N$ -dimension. Therefore, the optimization becomes a trade-off between a large margin and a small error penalty. The constraint optimization problem can be solved using “Lagrange multiplier” as

$$\min_{\mathbf{w},\xi,b} \max_{\alpha,\beta} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + c\mathbf{e}^T \xi - \sum_{n=1}^N \alpha_n [y_n(\mathbf{w}^T x_n - \mathbf{b}) - 1 + \xi_n] - \sum_{n=1}^N \beta_n \xi_n \right\} \tag{11}$$

The min-max problem is not easy to solve, so *dual form* technique is commonly proposed to solve it as

$$\max_{\alpha} \sum_{n=1}^N \alpha_n - \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_m \alpha_n y_m y_n x_m^T x_n \quad s.t. \begin{cases} 0 \leq \alpha_n \leq c \\ \sum_{n=1}^N \alpha_n y_n = 0, \quad n = 1, \dots, N \end{cases} \quad (12)$$

The key advantage of the dual form function is that the slack variables  $\xi_n$  vanish from the dual problem, with the constant  $c$  appearing only as an additional constraint on the Lagrange multipliers. Now, the optimization problem (12) becomes a *quadratic programming* (QP) problem, which is defined as the optimizing a quadratic function of several variables subject to linear constraints on these variables. Therefore, numerous methods can solve formula (10) within milliseconds, such as Sequential Minimal Optimization (SMO), least square, interior point method, augmented Lagrangian method, conjugate gradient method, simplex algorithm, etc.

### 3.3 Kernel SVM

Linear SVMs have the downside due to linear hyperplanes, which cannot separate complicated distributed practical data. In order to generalize it to nonlinear hyperplane, the kernel trick is applied to SVMs. The resulting algorithm is formally similar, except that every dot product is replaced by a nonlinear kernel function. In another point of view, the KSVMs allow to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space higher dimensional; thus though the classifier is a hyperplane in the higher-dimensional feature space, it may be nonlinear in the original input space. For each kernel, there should be at least one adjusting parameter so as to make the kernel flexible and tailor itself to practical data. RBF kernel is chosen due to its excellent performance with the form of

$$k(x_m, x_n) = \exp\left(-\frac{\|x_m - x_n\|}{2\sigma^2}\right) \quad (13)$$

where  $\sigma$  represents the scaling factor. Put formula (13) into formula (12), and we got the final SVM training function as

$$\max_{\alpha} \sum_{n=1}^N \alpha_n - \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_m \alpha_n y_m y_n \exp\left(-\frac{\|x_m - x_n\|}{2\sigma^2}\right) \quad s.t. \begin{cases} 0 \leq \alpha_n \leq C \\ \sum_{n=1}^N \alpha_n y_n = 0, \quad n = 1, \dots, N \end{cases} \quad (14)$$

It is still a quadratic programming problem, and we chose interior point method to solve the problem. However, there is still an outstanding issue, i.e., to determine the values of parameter  $C$  and  $\sigma$  in Eq. (14).

### 4 Optimization method

To determine the best parameters of  $C$  and  $\sigma$ , traditional method used either trial-and-error or grid-searching methods. They will cause heavy computation burden, and cannot guarantee to find the optimal or even near-optimal solutions. In this study, we used BBO.

### 4.1 Biogeography

Biogeography-based optimization (BBO) was inspired by biogeography, which describes speciation (i.e., the formation of new and distinct species in the course of evolution) and migration of species between isolated habitats (such as islands), and the extinction of species [5]. Habitats friendly to life are termed to have a high habitat suitability index (HSI), and vice versa. Features that correlate with HSI include land area, rainfall, topographic diversity, temperature, vegetative diversity, etc. Those features are called suitability index variables (SIV). Like other bio-inspired algorithms, the SIV and HSI are considered as search space and objective function, respectively [14].

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 immigration rate and a low emigration rate; the reason is not because species want to immigrate, but because there is a lot of resources for additional species [25].

To illustrate, Fig. 2 shows the relationship of immigration and emigration probabilities, where the  $\lambda$  and  $\mu$  represents the immigration and emigration probability, respectively.  $I$  and  $E$  represents the maximum immigration and emigration rate, respectively.  $S_{max}$  is the maximum number of species the habitat can support, and  $S_0$  the equilibrium species count. 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:

$$\lambda_S = I \left( 1 - \frac{S}{S_{max}} \right) \tag{15}$$

$$\mu_S = \frac{ES}{S_{max}} \tag{16}$$

The emigration and immigration rates are used to share information between habitats. Consider the special case  $E=I$ , we have  $\lambda_S + \mu_S = E$ .

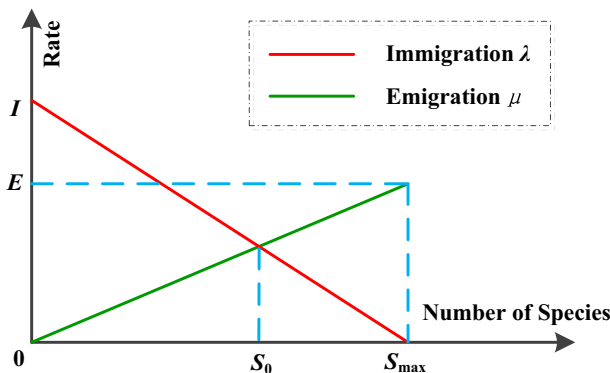


Fig. 2 Model of immigration  $\lambda$  and emigration  $\mu$  probabilities

## 4.2 Biogeography-based optimization

Emigration and immigration rates of each habitat are used to share information among the ecosystem. With modification probability  $P_{\text{mod}}$ , 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  $P_S$ .

$$m(H) = m_{\text{max}} \left( \frac{1 - P_S}{P_{\text{max}}} \right) \quad (17)$$

where  $m_{\text{max}}$  is a user-defined parameter, representing the maximum mutation rate.  $P_{\text{max}}$  is the maximum value of  $P(\infty)$ .

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 pseudocodes of BBO were listed as follows.

- Step 1 Initialize BBO parameters, which include a problem-dependent method of mapping problem solutions to SIVs and habitats, the modification probability  $P_{\text{mod}}$ , the maximum species count  $S_{\text{max}}$ , the maximum migration rates  $E$  and  $I$ , the maximum mutation rate  $m_{\text{max}}$ , and elite number  $p$ .
- Step 2 Initialize a random set of habitats.
- Step 3 Compute HSI for each habitat.
- Step 4 Computer  $S$ ,  $\lambda$ , and  $\mu$  for each habitat.
- Step 5 Modify the whole ecosystem by migration based on  $P_{\text{mod}}$ ,  $\lambda$  and  $\mu$ .
- 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.

## 5 Experiments

The experiments were carried out on the platform of P4 IBM with 3.2GHz processor and 8GB RAM, running under Windows 7 operating system. The algorithm was in-house developed based on the wavelet toolbox of 64bit Matlab 2014a (The Mathworks ©). The programs can be run or tested on any computer platforms where Matlab is available.

### 5.1 Materials

The datasets *brain* consists of 90 T2-weighted MR brain images in axial plane and 256x256 in-plane resolution, which were downloaded from the website of Harvard Medical School (URL: <http://www.med.harvard.edu/aanlib/home.html>). The abnormal brain MR images of the dataset consist of the following diseases: Glioma, Metastatic adenocarcinoma, Metastatic bronchogenic carcinoma, Meningioma, Sarcoma, Alzheimer, Huntington, Motor Neuron



disease, Cerebral Calcinosi, Pick's disease, Alzheimer plus visual agnosia, Multiple sclerosis, AIDS dementia, Lyme encephalopathy, Herpes encephalitis, Creutzfeld-Jakob disease, and Cerebral Toxoplasmosis. The samples of each disease are illustrated in Fig. 3. Note that all diseases are treated as abnormal brains, and our task is a binary classification problem, i.e., distinguish abnormal brains from normal brains.

## 5.2 Cross validation setting

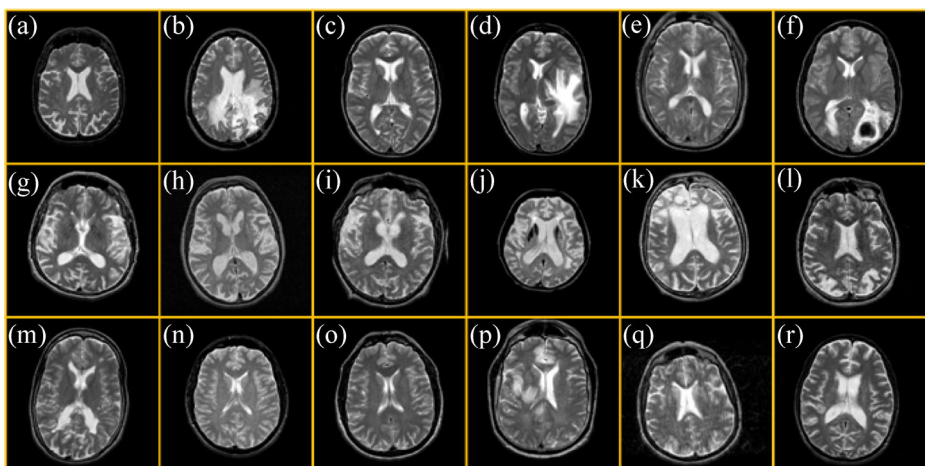
We randomly selected 5 images for each type of brain. Since there are 1 type of normal brain and 17 types of abnormal brain in the dataset,  $5 \times (1 + 17) = 90$  images was selected to construct the *brain* dataset, consisting of 5 normal and 85 abnormal brain images in total.

The setting of the 5-fold cross validation was shown in Fig. 4. In this study, we divided the dataset into 5 equally distributed subgroups, each subgroups contain one normal brain and 17 abnormal brains. We perform five experiments. In each experiment, four groups were used for training, and the left group was used for validation. Each group was used once for validation.

Further, the 5-fold CV was repeated 5 runs, i.e., a  $5 \times 5$ -fold CV was implemented, with the aim of reducing randomness. The confusion matrices of 5 runs were combined to form the final confusion matrix, with the ideal results of 425 abnormal instances and 25 normal instances perfectly recognized.


## 5.3 Feature extraction

The 3-level 2D-DWT of Haar wavelet decomposed the input image into 10 subbands as shown in Fig. 5. Symmetric padding method [18] was utilized to calculate the boundary value, with the aim of avoiding border distortion. After decomposition, 10 features were obtained from



**Fig. 3** Sample of brain MRIs: (a) Normal Brain; (b) Glioma, (c) Metastatic adenocarcinoma; (d) Metastatic bronchogenic carcinoma; (e) Meningioma; (f) Sarcoma; (g) Alzheimer; (h) Huntington; (i) Motor Neuron disease; (j) Cerebral Calcinosi; (k) Pick's disease; (l) Alzheimer plus visual agnosia; (m) Multiple sclerosis; (n) AIDS dementia; (o) Lyme encephalopathy; (p) Herpes encephalitis; (q) Creutzfeld-Jakob disease; and (r) Cerebral Toxoplasmosis

Experiment 1	1	1	1	1	1	Normal
	17	17	17	17	17	Abnormal
Experiment 2	1	1	1	1	1	Normal
	17	17	17	17	17	Abnormal
Experiment 3	1	1	1	1	1	Normal
	17	17	17	17	17	Abnormal
Experiment 4	1	1	1	1	1	Normal
	17	17	17	17	17	Abnormal
Experiment 5	1	1	1	1	1	Normal
	17	17	17	17	17	Abnormal



**Fig. 4** Illustration of 5-fold Cross Validation of *Brain* Dataset

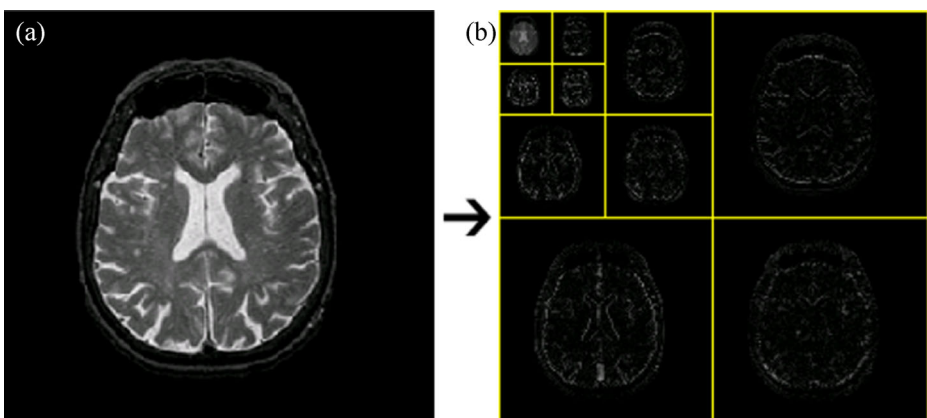
calculating the energy of those 10 subbands (LL3, HL3, LH3, HH3, HH2, HL2, LH2, HH1, HL1, and LH1).

The wavelet-energy can reduce the dimension of DWT coefficients. For a 256x256 image, the 3-level DWT did not reduce the dimension, but the 3-level wavelet-energy can reduce the feature dimension from 65,536 to only 10.

### 5.4 Algorithm comparison

The wavelet-energy features were fed into different classifiers. We compared the proposed BBO-KSVM method with Back Propagation Neural Network (BP-NN) [4], KSVM [19], and PSO-KSVM [31]. The results were shown in Table 1. The kernel function was set to RBF.

Two images of Lyme Encephalopathy are shown in Fig. 6. The left image was incorrectly classified as normal, and the right image was correctly predicted as abnormal. The reason is the left image did not fully capture the focus and deformed region. This gives us a hint that the proposed method may be improved by using multi-slices.



**Fig. 5** The results of a 3-level 2D DWT: (a) a normal brain MR image; (b) decomposition result

**Table 1** Classifier comparison based on WE features (WE=Wavelet-energy)

Method	Confusion matrix	Success cases	Sensitivity	Specificity	Precision	Accuracy
WE+BP-NN	[376 49 1114]	390	88.47 %	56.00 %	97.16 %	86.67 %
WE+KSVM	[393 32 718]	413	93.18 %	68.00 %	98.02 %	91.78 %
WE+PSO-KSVM	[415 10 322]	437	97.65 %	88.00 %	99.28 %	97.11 %
WE+BBO-KSVM (Proposed)	[417 8 223]	440	98.12 %	92.00 %	99.52 %	97.78 %

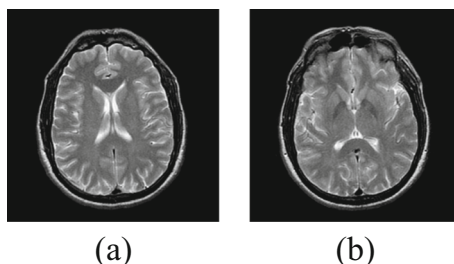
## 6 Discussion

Results in Table 1 are the main contribution of this study. It showed that BP-NN [4] correctly matched 390 cases with 86.67 % classification accuracy. KSVM [19] correctly matched 413 cases with accuracy of 91.78 %. The PSO-KSVM [31] correctly matched 437 brain images with accuracy of 97.11 % classification accuracy. Finally, the proposed BBO-KSVM correctly matched 440 cases with accuracy of 97.78 %.

Therefore, the proposed method had the most excellent classification performance. In addition, the proposed method achieved the sensitivity of 98.12 %, specificity of 92.00 %, and precision of 99.52 %. This validated the effectiveness of the proposed method. It may be applied in practical use.

Why the proposed BBO-KSVM performs the best? Because it is combined on two successful components: the KSVM and the BBO. KSVM is widely used for pattern-recognition problems, and it achieves excellent results. On the other hand, BBO is a rather novel method, which included mutation and elitism mechanisms. The importance of introducing BBO is to determine the optimal values of parameters  $C$  and  $\sigma$  in KSVM, otherwise it is hard to get the optimal weights for KSVM. Therefore, the BBO is an effective means to find the optimal values. Integrating BBO to KSVM enhances the classification capability of KSVM. The comparison in the experiment also shows BBO excels PSO.

We choose the Haar wavelet, although there are other outstanding wavelets. We will compare the performance of different families of wavelet in future work.



**Fig. 6** Two images of Lyme Encephalopathy: (a) Incorrectly classified; (b) Correctly classified

The **contributions** of this paper lie in following aspects: (i) A new approach for automatic classification of MR Images as normal or abnormal using Wavelet-Energy, SVM, and BBO was proposed. (ii) Our experiments demonstrated the proposed method performed better than BP-NN, KSVM, and PSO-KSVM. (iii) Wavelet-Energy was validated as effective feature for MR brain image classification.

## 7 Conclusions and future research

In this study, we proposed a new approach for automatic classification of MR Images as normal or abnormal using Wavelet-Energy, SVM, and BBO. The results showed that the proposed method gave better results than latest methods. This automated CAD system, could be further used for classification of image with different pathological condition, types and disease status.

In the future, we may focus on following regards: (i) we try to increase the decomposition level of 2D-DWT, in order to test whether higher level can lead to better classification performance. (ii) we may replace wavelet-energy with more efficient feature descriptors, such as scale-invariant features. (iii) Some advanced pattern recognition techniques may be used, such as deep learning and RBFNN [13]. (iv) Multiple slices can be used to improve the classification performance.

**Acknowledgments** This paper was supported by NSFC (610011024, 61273243, 51407095), Program of Natural Science Research of Jiangsu Higher Education Institutions (13KJB460011, 14KJB520021), Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing (BM2013006), Key Supporting Science and Technology Program (Industry) of Jiangsu Province (BE2012201, BE2014009-3, BE2013012-2), Special Funds for Scientific and Technological Achievement Transformation Project in Jiangsu Province (BA2013058), Nanjing Normal University Research Foundation for Talented Scholars (2013119XGQ0061, 2014119XGQ0080), and Science Research Foundation of Hunan Provincial Education Department (12B023).

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

## References

1. Bafroui HH, Ohadi A (2014) Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions. *Neurocomputing* 133:437–445
2. Cai ML et al (2014) Motion recognition for 3D human motion capture data using support vector machines with rejection determination. *Multimedia Tools Appl* 70(2):1333–1362
3. Chaplot S, Patnaik LM, Jagannathan NR (2006) Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomed Signal Process Control* 1(1):86–92
4. Choudhary R et al (2009) Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosyst Eng* 102(2):115–127
5. Christy AA, Raj P (2014) Adaptive biogeography based predator–prey optimization technique for optimal power flow. *Int J Electr Power Energy Syst* 62:344–352
6. Dai Y, Zhang JX, Xue Y (2013) Use of wavelet energy for spinal cord vibration analysis during spinal surgery. *Int J Med Robot Comput Assist Surg* 9(4):433–440
7. Das S, Chowdhury M, Kundu MK (2013) Brain MR image classification using multiscale geometric analysis of ripplelet. *Prog Electromagn Res Pier* 137:1–17

8. Dong Z et al (2014) 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
9. El-Dahshan ESA, Hosny T, Salem ABM (2010) Hybrid intelligent techniques for MRI brain images classification. *Digital Signal Process* 20(2):433–441
10. El-Dahshan ESA et al (2014) Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm. *Expert Syst Appl* 41(11):5526–5545
11. Fang L, Wu L, Zhang Y (2015) A novel demodulation system based on continuous wavelet transform. *Math Probl Eng* 2015:9
12. Goh S et al (2014) Mitochondrial dysfunction as a neurobiological subtype of autism spectrum disorder: Evidence from brain imaging. *JAMA Psychiatry* 71(6):665–671
13. Guo DL et al. (2014) Improved Radio Frequency Identification Indoor Localization Method via Radial Basis Function Neural Network. *Math Probl Eng*. No. pp. 2014
14. Guo WA et al (2014) Biogeography-based particle swarm optimization with fuzzy elitism and its applications to constrained engineering problems. *Eng Optim* 46(11):1465–1484
15. Kalbkhani H, Shayesteh MG, Zali-Vargahan B (2013) Robust algorithm for brain magnetic resonance image (MRI) classification based on GARCH variances series. *Biomed Signal Process Control* 8(6):909–919
16. Lee SH et al (2013) Diagnostic method for insulated power cables based on wavelet energy. *Ieice Electron Express* 10(12):20130335
17. Maitra M, Chatterjee A (2006) A Slantlet transform based intelligent system for magnetic resonance brain image classification. *Biomed Signal Process Control* 1(4):299–306
18. Messina A (2008) Refinements of damage detection methods based on wavelet analysis of dynamical shapes. *Int J Solids Struct* 45(14–15):4068–4097
19. Mookiah MRK et al (2012) Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. *Knowl-Based Syst* 33:73–82
20. Nanthagopal AP, Rajamony RS (2013) Classification of benign and malignant brain tumor CT images using wavelet texture parameters and neural network classifier. *J Vis* 16(1):19–28
21. Padma A, Sukanesh R (2014) Segmentation and classification of brain CT images using combined wavelet statistical texture features. *Arab J Sci Eng* 39(2):767–776
22. Panchal MB, Mahapatra DR (2013) Influence of crack parameters on wavelet energy correlated damage index. *Adv Vib Eng* 12(3):311–318
23. Ramasamy R, Anandhakumar P (2011) Brain Tissue Classification of MR Images Using Fast Fourier Transform Based Expectation-Maximization Gaussian Mixture Model. in *Advances in Computing and Information Technology*, ed: Springer, pp. 387–398
24. Saritha M, Paul Joseph K, Mathew AT (2013) Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. *Pattern Recogn Lett* 34(16):2151–2156
25. Simon D (2011) A probabilistic analysis of a simplified biogeography-based optimization algorithm. *Evol Comput* 19(2):167–188
26. Wu YL et al (2014) Gaze direction estimation using support vector machine with active appearance model. *Multimedia Tools Appl* 70(3):2037–2062
27. Xu S et al (2014) Multi-task least-squares support vector machines. *Multimedia Tools Appl* 71(2):699–715
28. Zhang Y, Wu L (2012) An Mr brain images classifier via principal component analysis and kernel support vector machine. *Prog Electromagn Res* 130:369–388
29. Zhang Y, Wu L, Wang S (2011) Magnetic resonance brain image classification by an improved artificial Bee colony algorithm. *Prog Electromagn Res* 116:65–79
30. Zhang Y et al (2011) A hybrid method for MRI brain image classification. *Expert Syst Appl* 38(8):10049–10053
31. Zhang Y et al (2013) An MR brain images classifier system via particle swarm optimization and kernel support vector machine. *Sci World J* 2013:9
32. Zhang Y et al (2013) Genetic pattern search and its application to brain image classification. *Math Probl Eng* 2013:8
33. Zhang Y et al (2014) Energy preserved sampling for compressed sensing MRI. *Comput Math Methods in Med* 2014:12
34. Zhang Y et al (2015) Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging. *IEEJ Trans Electr Electron Eng* 10(1):116–117
35. Zhang Y et al (2015) Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with tsallis entropy and generalized eigenvalue proximate support vector machine (GEP SVM). *Entropy* 17(4):1795–1813



**Gelan Yang** Was born in Hunan, China, in October, 1975. He is a full professor of School of Information Science and Engineering at Hunan City University, China. His current fields of interests include: Objects detection and recognition, Action recognition and Video-sequence classification.



**Yudong Zhang** Is a full professor and doctoral advisor of the school of computer science and technology at Nanjing Normal University. He is the (guest) editors of seven SCI-indexed journals. He was included in “2014 Most Cited Chinese Researchers (elsevier)” in computer science. His research interests focus on knowledge engineering and medical image processing.



**Jiquan Yang** Is the Associate Dean of School of Electrical & Automation Engineering, NJNU. He is the Director of Jiangsu Key Laboratory of Three Dimensional Printing Equipment and Manufacturing. He is currently the President of Nanjing Institute of Three Dimensional Printing. His research direction consists of three Dimensional Printing, Digital Design and Manufacture.



**Genlin Ji** Is the Dean of School of Computer Science and Technology, NJNU. His research interests consist of data mining and machine learning.

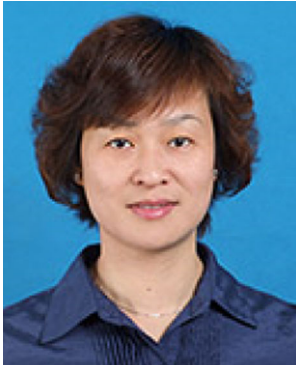


**Zhengchao Dong** Is the assistant professor of Columbia University, USA. He is also the Research Scientist (V) of New York State Psychiatric Institute.



**Shuihua Wang** Is the lecturer of Nanjing Normal University. She is currently pursuing the Ph.D. degree in Nanjing University. Her research interest is biomedical image processing.





**Chunmei Feng** Is the Associate Professor of School of Electrical & Automation Engineering, NJNU. Her main research direction consists of 3D Printing and electric control technology.



**Qiong Wang** Is the Associate Professor of School of Computer Science and Technology, NJNU. His main research direction is Computer Graphics.