Multimed Tools Appl (2018) 77:10393–10417
DOI 10.1007/s11042-016-4222-4

Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization

Shui-Hua Wang[1] · **Yin Zhang**[2] · **Yu-Jie Li**[3] · **Wen-Juan Jia**[1] · **Fang-Yuan Liu**[1] · **Meng-Meng Yang**[1] · **Yu-Dong Zhang**[1,4]

Received: 26 September 2016 / Revised: 25 October 2016 / Accepted: 28 November 2016 /
Published online: 9 December 2016

Abstract Detection of Alzheimer's disease (AD) from magnetic resonance images can help neuroradiologists to make decision rapidly and avoid missing slight lesions in the brain. Currently, scholars have proposed several approaches to automatically detect AD. In this study, we aimed to develop a novel AD detection system with better performance than existing systems. 28 ADs and 98 HCs were selected from OASIS dataset. We used inter-class variance criterion to select single slice from the 3D volumetric data. Our classification system is based on three successful components: wavelet entropy, multilayer perceptron, and biogeography-base optimization. The statistical results of our method obtained an accuracy of 92.40 ± 0.83%, a sensitivity of 92.14 ± 4.39%, a specificity of 92.47 ± 1.23%. After comparison, we observed that our pathological brain detection system is superior to latest 6 other approaches.

Keywords Pathological brain detection · Inter-class variance · Alzheimer's disease · Wavelet entropy · Multilayer perceptron · Biogeography-based optimization

✉ Yu-Dong Zhang
zhangyudong@njnu.edu.cn

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

[2] School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan, Hubei 430073, China

[3] School of Information Engineering, Yangzhou University, Yangzhou, China

[4] Translational Imaging Division & MRI Unit, Columbia University and New York State Psychiatric Institute, New York, NY 10032, USA

1 Background

As is known, Alzheimer's disease (AD) is a chronic neurodegenerative disease aging progress [44]. It accounts for 60% and 70% of dementia that causes severe thinking, memory, and behavior problems [66]. In the past, nearly 30 million elders were reported as suffering from AD. In the coming 2050, Alzheimer's disease (AD) shall affect 1 in 85 people in the worldwide [63].

As the world steps into an aging society, people with AD bring heavy burdens and negative impacts to both their families and the society. In US, the cost for the healthcare on people with AS is about $100 billion every year, and shall increase to $1 trillion annually after thirty years [48].

In recent decades, several neuroimaging techniques have been widely applied in clinical diagnosis and AD detection. Those techniques are composed of computed tomography (CT) [11], single-photon emission computed tomography (SPECT) [37], positron emission tomography (PET), magnetic resonance imaging (MRI) [9, 19, 49, 69], magnetic resonance spectral imaging (MRSI) [12], functional magnetic resonance imaging (fMRI) [62], etc.

Automatic detection for AD is extremely important for patients, so they can have enough time to get early treatment. Two types of detection systems exist in past researches, one is whole brain based detection (WBD) [47], and the other is single slice based detection (SSD) [78]. In hospitals, the latter was widely used due to its inexpensiveness (only 300–500 RMB per scan) and rapidness (only two or three minutes per scan). In this study, we focus on the latter one.

How to detect important features from brain images? Physicians and computer experts have different opinions on this problem. Physicians like to extract local features, they either measure the volume of segmented region of interest (ROI), or use voxel-based morphometry (VBM) to measure the atrophy, or measure cortical thickness and other features related to brain tissues [46]. On the other hand, computer experts like to use image processing [35, 36, 46] techniques and artificial intelligence methods [5, 75], and therefore directly extract image global features, like Hu's moment invariants [60], Zernike moment [21], wavelet energy [72], wavelet transform [34], etc.

Scholars have proposed various methods to detect AD, which are listed in Section 2. Our contribution aims to propose a novel SSD system for AD with higher accuracy than state-of-the-art approaches, on the basis of wavelet entropy, multilayer perceptron, and an improved biogeography-based optimization method. Besides, our contribution is to compare our proposed method with widely used methods by strict statistical experiments.

The structure of the paper is organized as below: Section 2 discusses the background and latest methods. Section 3 presents the feature extraction methods. Section 4 offers the classification methods. Section 5 reports the data, the results and gives corresponding discussions. Final Section 6 presents the concluding remarks.

2 State-of-the-art

Currently, there are many novel AD detection methods: Dong (2014) [14] employed the undersampling (US) technique. They used principal component analysis (PCA) and singular value decomposition (SVD) to select features. Finally, they combined decision tree (DT) with support vector machine (SVM). Plant (2010) [51] employed brain region cluster (BRC) and information gain (IG). Savio (2013) [55] offered a new deformation-based morphometry (DBM) method.

They found modulated gray matter (MGM) performed well. Furthermore, they utilized Pearson's correlation (PC) to select important features. Yuan (2015) [73] employed the eigenbrain (EB) to extract features. Afterwards, they employed Welch's t-Test (WTT) to reduce the feature dimensionality. Gray (2013) [23] put forward a voxel-based morphometry (VBM) method and employed random forest (RF) technique. Zhang (2015) [76] proposed a novel displacement field (DF) to detect AD.

This paper proposed a novel AD detection system, which is based on wavelet entropy (WE) and multi-layer perceptron (MLP). WE has been successfully applied in various medical applications. For example, Shiyang (2007) [57] applied WE to analyze the heart rate fluctuation. Bakhshi (2013) [4] applied continuous-time WE to detect cardiac repolarization alternans. Frantzidis (2014) [18] used relative WE and electroencephalographic (EEG) to detect AD. Candra (2015) [10] used WE to classify EEG-emotion signal.

On the other hand, MLP is also a prevalent classification tool in medical fields. For instance, Sonawane (2014) [58] applied MLP to predict heart diseases. Behera (2015) [6] used bird mating optimization method and MLP to classify diseases. Ibrahim (2015) [26] used MLP to diagnose breast cancer. Meng (2015) [41] analyzed the meteorological factors related to emergency admission of elder stroke patients in Shanghai, with the tool of MLP.

From above, we see WE and MLP are efficient tools to analyze medical signal and images, and they have achieved success in recent studies. This gives solid support for our study.

3 Feature extraction

3.1 Single slice selection

The single slice was selected via our past proposed inter-class variance (ICV) criterion [76] with important modifications. Reference [76] selected 10 most important slices from each 3D volumetric brain image. In this paper, we only choose one important slice, hence, the slice with the highest ICV was picked up from all slices, and it was used for following processing. The slice direction may be sagittal, coronal, or axial. In this study, we chose axial direction by experience.

3.2 Wavelet transform

In this study, Wavelet transform (WT) was firstly analyzed. Torrents-Barrena (2015) [61] selected complex wavelet transform to handle Alzheimer's electroencephalography signals. Aggarwal (2015) [2] used 3d discrete wavelet transform on T1-weighted brain magnetic resonance images for the diagnosis of AD. However, wavelet transform introduces additional parameters such as wavelet families and decomposition scales [79]. Previous studies commonly selected wavelet families and decomposition scales by experience or arbitrarily.

When the wavelet analysis method is employed for AD detection, the ultimate aim is to obtain better identification rate of AD subjects. Another problem raised is wavelet transform will generate the same size of coefficients as original 3D brain image, which will cause a burden to the consequent analysis.

Entropy can be employed to measure the information content over the decomposed wavelet coefficients. The wavelet entropy (WE) have been proposed to calculate the entropy of t wavelet subband coefficients distribution.

The discrete wavelet transform (abbreviated as DWT) [39] implements continuous wavelet transform (CWT) using the dyadic scales and positions [43]. Suppose t represents the time and $r(t)$ is a given signal in the time domain (can be extended to spatial domain easily), the CWT is defined below:

$$S(a,\tau) = \int_{-\infty}^{\infty} r(t)\frac{1}{\sqrt{a}}\psi^*\left(\frac{t-u}{a}\right)\mathrm{d}t \tag{1}$$

where Ψ is a real-valued wavelet function. S denotes the wavelet coefficients, a the dilation factor, u the translation factor.

We discretize formula (1) by limiting a and u to a discrete lattice

$$a = 2^j \tag{2}$$

$$u = 2^j k \tag{3}$$

Then, we have the DWT form of

$$\begin{aligned} L_{j,k}(n) &= \Omega\left[\sum_n r(n) l_j^*\left(n-2^j k\right)\right] \\ H_{j,k}(n) &= \Omega\left[\sum_n r(n) h_j^*\left(n-2^j k\right)\right] \end{aligned} \tag{4}$$

Here Ω denotes the downsampling [16]. n is the discrete counterpart of variable t. L represents the approximation coefficients through a low-pass filter $l(n)$. H represents the detail coefficients through a high-pass filter $h(n)$. j and k denotes the scale and translation factor for wavelet function, respectively.

3.3 Brain image oriented wavelet

What kind of wavelet is suitable for brain image? To answer the question, we selected three row-lines and three column-lines from a randomly selected brain image. Figure 1(a) shows the randomly selected brain image. Figure 1(b-d) presents three row-lines with indexes of 30, 40, and 50, respectively. Figure 1(e-g) presents three column-lines with indexes of 60, 70, and 80, respectively.

After comparing the row-lines and column-lines with different wavelets, we finally selected to use bior4.4 wavelet, since the wavelet forms of bior4.4 are similar to the sharp changes of gray-level values in the lines of brain images. Figure 2 shows the decomposition functions of bior4.4.

Compared to orthogonal wavelets, biorthogonal wavelet has more freedom degrees, and its wavelet transform is invertible but not necessarily orthogonal. Another advantage of biorthogonal wavelet is to generate symmetric wavelet functions.

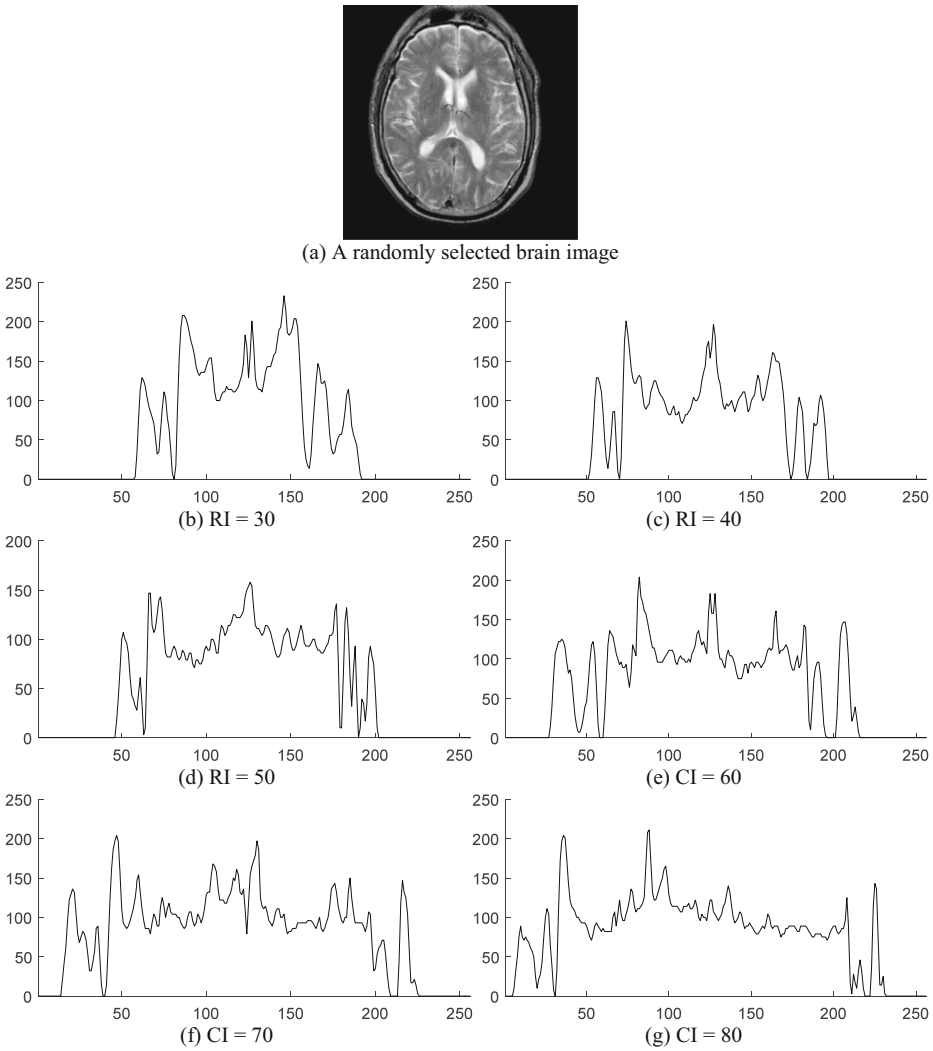


Fig. 1 Randomly selected row-line and column-line of brain images (RI = Row Index, CI = Column Index)

3.4 Entropy and wavelet entropy

In statistics, entropy is defined for a stochastic system to measure its randomness quantitatively. Suppose we have a continuous random variable $X \in R^n$, then the entropy S can be calculated as:

$$S = \int_0^\infty -\eta(x)\log\eta(x)\mathrm{d}x \tag{5}$$

where $\eta(x)$ denotes the probability density function (PDF) of variable X. The value range of entropy sit between zero and one. The less the entropy is, the less uncertainty degree the system is, and vice versa.

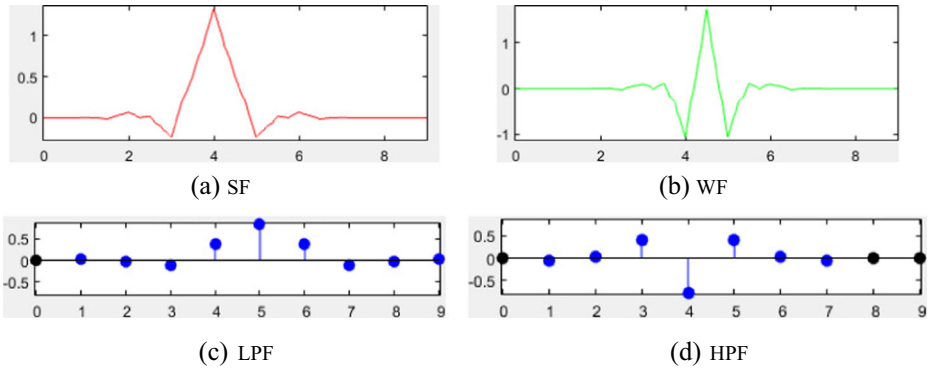


Fig. 2 Some important functions of bior4.4 decomposition (SF = scaling function; WF = wavelet function; LPF = low-pass filter; HPF = high-pass filter)

Wavelet entropy (WE) [25] calculates the entropy value of the PDF of the energy distribution of wavelet subband coefficients in the wavelet domain. It combines wavelet transform and Shannon entropy, so as to estimate the disorder/order degree of a particular image with specified spatial-frequency resolution. Suppose we have a brain image with size of 256 × 256, and take a 2-level WE as an example.

Figure 3 shows the diagram of calculating WE. Here a brain image was submitted, after firstly taking 1-level DWT, we have four subbands in total (HL1, LL1, HH1, and LH1). Then, a 2-level DWT decomposes the LL1 subband, which is then transformed to four other subbands (HL2, LL2, HH2, and LH2). The LL1 subband usually contains more image information, thus, it is also called approximation subband. The other three bands contain only detail information, and therefore they are called detail subband. In a word, the next decomposition is always performed over the LL subband. In total, there are 7 subbands (LL2, HL2, LH2, HH2, HL1, LH1, and LL1). Entropy is then implemented over each subband, and finally a 7-element entropy vector is output.

4 Classifier

4.1 Multilayer perceptron

In the field of artificial intelligence [5], a multilayer perceptron (MLP) is a feed-forward neural network structure that maps input training points to target labels. The universal approximation theorem [74] guarantees the MLP can approximate to the model required for our task. Figure 4 shows that an MLP is composed of multiple layers (usually 3) of nodes in a directed graph.

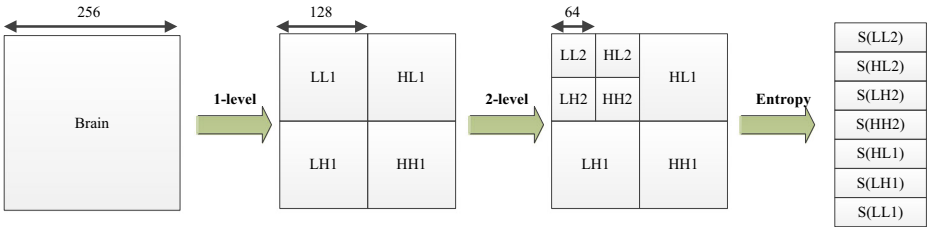


Fig. 3 Pipeline of Calculating WE (L = Low; H = High; S = Entropy)

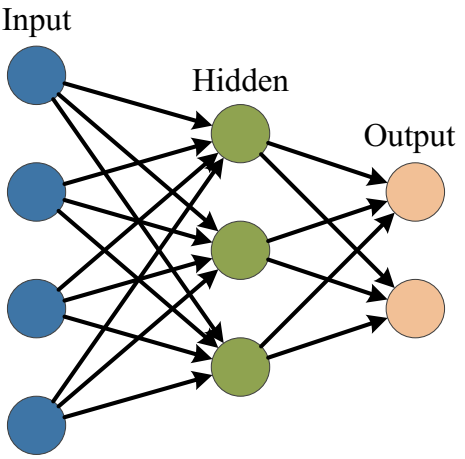


Fig. 4 Diagram of multilayer perceptron, in which each layer connecting fully to the next layer

4.2 Biogeography-based optimization

Traditionally, MLP is trained by backpropagation (BP) algorithm. The BP method uses gradient of the loss function to train the weights and biases in the MLP. Before training, the weights and biases were generated at random. Then, we measure the mean-squared error (MSE) between realistic output R and target output T

$$E = \frac{1}{2}\|T-R\|^2 \tag{6}$$

Here E represents the value of MSE. The target here was interpreted as the clinical dementia rating (CDR) as shown in Table 2.

Nevertheless, the loss function contains many local minimal points, and gradient descent method may converge to one of the local minimal points. To improve the performance, scholars have suggested to use swarm intelligence that can converge to global minimal point at very high probability. For instance, Dil (2016) [13] used genetic algorithm (GA) to train the artificial neural network. Saghatforoush (2016) [54] combined ant colony optimization (ACO) with neural network. Mashhadban (2016) [40] applied particle swarm optimization (PSO) in training ANN. Shamshirband (2016) [56] combined cuckoo search (CS) with ANN. Two years ago, Mirjalili (2014) [42] proposed to use a novel training method, named biogeography-based optimization (BBO), to train MLP, and they reported the superior performances of BBO to other swarm intelligence based training algorithms. Therefore, we chose the BBO in this study.

Biogeography-based optimization (BBO) iteratively approximates to the global optimal point of an optimization problem [28, 65, 70], by mimicking the context of biogeography. First, we define the "habitat suitability index (HASI)" of the comfort measure of each habitat based on current living conditions. The HASI relies on numerous variables [17], such as temperature, rainfall, area, humidity, vegetation, etc. Those variables are defined as "suitability index variables (SUIV)" [52].

Three important components are covered in BBO algorithm: migration, elitism, and mutation. In below texts, we shall discuss them in sequence.

Habitats with higher and lower HASI values tend to emigrate and immigrate, respectively, since emigration is caused by intense competitions among existed species, and immigration is yielded due to abundant resources left for extra species [15, 71]. Therefore, based on the relationship of emigration rate x and the immigration rate y, we can model the migration of species as:

$$x(z) = X \times \frac{z}{Z} \tag{7}$$

$$y(z) = Y \times \frac{1-z}{Z} \tag{8}$$

In the formula, z denotes the number of species, Z represents the maximum number of species. X and Y represents the highest values of emigration and immigration rates, respectively [24].

Suppose $a(z)$ represents the solution probability of species s, A is the maximum value of a. Hence, mutation u is defined as

$$u(z) = \frac{1-a(z)}{A} \times U \tag{9}$$

where U is the maximum mutation rate [8]. The mutation operation is carried out by:

$$F_{i,\mathrm{m}} = F_{i,k} + \theta \times \left(F_{i,\max} - F_{i,\min}\right) \tag{10}$$

Where θ is a random number in the range of [0, 1]. $F_{i,k}$ represents the SUIV value at k-th step. $F_{i,\mathrm{m}}$ is the mutated value, and will be assigned to $F_{i,k}$ if it can provide better HASI value. $F_{i,\max}$ and $F_{i,\min}$ represents the lower and upper bounds of $F_{i,k}$, respectively. Remember that mutation was carried out independently on each SUIV.

$$\begin{aligned} &F_{i,k} \leftarrow F_{i,\mathrm{m}}, \\ &\text{if } v\left(\left[F_{1,k}, F_{2,k}, \ldots, F_{i-1,k}, F_{i,\mathrm{m}}, F_{i+1,k}, \ldots\right]\right) < v(F_k) \end{aligned} \tag{11}$$

where v represents the HASI objective function. On the other hand, elitism keeps the best solutions within the ecosystem [50], to counteract the effect of mutation operation. Assume the number of elitism is ξ, then the algorithm performs elitism by assigning $y = 0$ for the best ξ elites.

Table 1 Measurement of classification performance

Measure	Definition
Accuracy	(TN + TP)/(TP + FP + TN + FN)
Sensitivity	TP/(FN + TP)
Specificity	TN/(FP + TN)

4.3 Stratified cross validation

With the aim of statistical analysis, 10-fold stratified cross validation (SCV) was employed for fair comparison. The 10-fold SCV repeated 50 times, viz., a 50 × 10-fold SCV was implemented. For each run, we take the accuracy, sensitivity, and specificity, as the measurement of the performance (See Table 1).

A correctly recognized AD case was taken as a true positive. Based on the 50 runs, the final three measures of were reported, in the form of both the mean and standard deviation (SD).

5 Experiments, results, and discussions

Data used in the simulation experiments in this paper, were downloaded from "Open Access Series of Imaging Studies (OASIS)" [3]. The OASIS is a project, compiling and freely distributing MRI data sets, in order to make MRI data sets of the brain freely available to the scientific community [53].

OASIS covers two types of data: cross-sectional MRI data and longitudinal MRI data. In this study, we used cross-sectional MRI data because our study aims at developing an automatic system to detect AD, which is not relevant to longitudinal data in which AD subjects were gathered over a long period of time.

The cross-sectional MRI data in OASIS include 416 subjects, who aged from 18 to 96. All subjects are right-handed, and include both men and women. In this study, we pick up 126 samples (28 ADs and 98 HCs). The exclusion criterion is subjects less than 60 years old or any of their records are missing. The demographic statuses are reported in Table 2. The imbalanced data may cause problem in future identification, we adjust cost matrix [29, 45] to solve this problem.

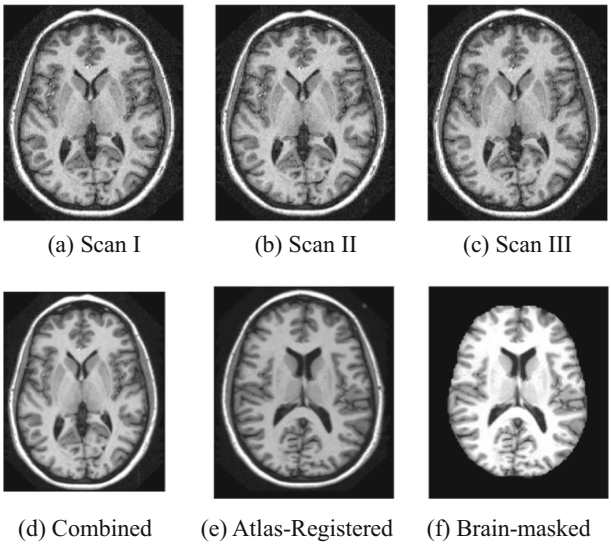
5.1 Image preprocessing

For each subject, each scanning session include three or four individual T1-weighted MRI scans. In order to increase the signal-to-noise ratio (SNR), all those MRI scans with the same protocol of the same person were motion-corrected, and spatially co-registered to the Talairach space to generate an averaged image, and then brain

Table 2 Demographic Status of subjects

Characteristic	AD	HC
PN	28	98
Age (Year)	77.75(6.99)	75.91(8.98)
MMSE	21.67(3.75)	28.95(1.20)
Ed.	2.57(1.31)	3.26(1.31)
SES	2.87(1.29)	2.51(1.09)
CDR	1	0
GD (M/F)	9/19	26/72

(PN = participant number; MMSE = mini-mental state examination; SES = Socioeconomic Status; Ed. = Education; CDR = clinical dementia rating; $x(y)$ represents x is the mean and y is the standard deviation; GD = gender; M = male; F = female)



(a) Scan I (b) Scan II (c) Scan III

(d) Combined (e) Atlas-Registered (f) Brain-masked

Fig. 5 Preprocessing of a specified subject (Axial View)

masked. The motion-correction registered the 3D images of all scans, and then generated an average 3D image in original acquisition space. The images were then resampled to 1 mm × 1 mm × 1 mm. The image was transformed from acquisition space to Talairach coordinate space. Finally, the brain extraction was implemented. The whole preprocessing can be viewed in Fig. 5.

All MR images were downloaded from OASIS and preprocessed. We only considered one slice of all MR images in axial view. Figure 6 shows exemplar instances of both HC and AD.

5.2 WE results

The AD image was decomposed by bior4.4 wavelet. Figure 7(a) presents an original AD image. Clearly the ventricle is enlarged and the cortex is shrunk compared to healthy controls. Afterwards, Fig. 7(b) offers the 1-level DWT decomposition results, in which four subbands (LL1, LH1, HL1, and HH1) preserve different components from original image. Figure 7(c) shows the 2-level decomposition results. The LL1 was decomposed to LL2, LH2, HL2, and HH2. Figure 7(d) shows the 3-level decomposition results. The LL2 was further decomposed to LL3, LH3, HL3, and HH3.

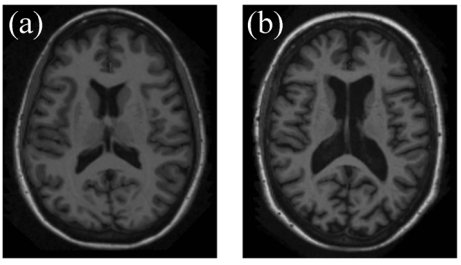
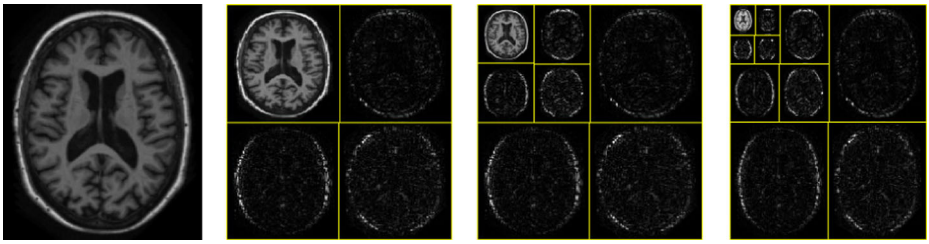


Fig. 6 Axial view of (**a**) HC and (**b**) AD



(a) AD (b) 1-level DC (c) 2-level DC (d) 3-level DC

Fig. 7 DWT Decomposition Results (DC = decomposition)

5.3 Optimal decomposition level

After 50 × 10-fold stratified cross validation, our "WE + MLP + BBO" approach achieved an accuracy of 92.40%, a sensitivity of 92.14%, a specificity of 92.47%, a precision of 77.76%, when 3-level decomposition was implemented. We also changed the decomposition level from 1 to 4, and plotted the corresponding performance change in Fig. 8, from which we can observe 3-level decomposition yields the best performance.

5.4 Statistical analysis

In the third experiment, we give the details result of each run over each fold of this proposed method. Appendix 1 gives the segmented results based stratified cross validation. Remember that we have 28 AD subjects and 98 HC healthy subjects. The 10-fold segmentation divides the dataset into ten folds. Each column in Appendix 1 represents a different fold. Each row in Appendix 1 represents a run. The same setting is for Appendix 2.

The sensitivities, specificities, and accuracies over the 50 runs of 10-fold SCV is shown below in Table 3. Here we observe that the sensitivity of our method is 92.14 ± 4.39%, the specificity is 92.47 ± 1.23%, the accuracy is 92.40 ± 0.83%.

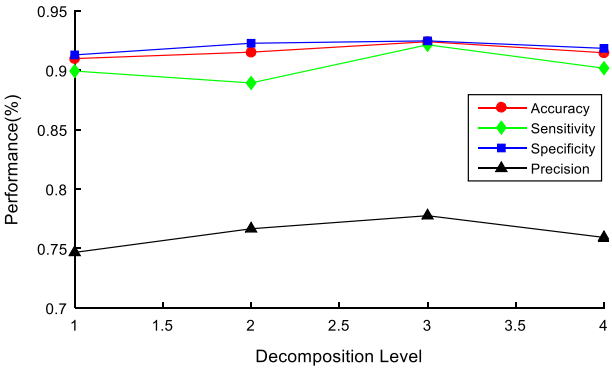


Fig. 8 Classification performance vary with decomposition level

Table 3 Measures over 50 runs (Unit: %)

	Sen.	Spc.	Acc.		Sen.	Spc.	Acc.		Sen.	Spc.	Accu.
R1	100.00	92.86	92.06	R18	92.86	92.86	92.86	R35	85.71	94.90	92.06
R2	96.43	93.88	92.86	R19	96.43	91.84	91.27	R36	96.43	92.86	92.86
R3	89.29	93.88	92.06	R20	96.43	94.90	92.06	R37	100.00	91.84	92.86
R4	96.43	89.80	91.27	R21	89.29	93.88	92.06	R38	100.00	92.86	94.44
R5	89.29	94.90	92.86	R22	96.43	90.82	92.06	R39	100.00	91.84	93.65
R6	89.29	94.90	91.27	R23	92.86	92.86	92.86	R40	82.14	93.88	91.27
R7	92.86	94.90	93.65	R24	96.43	91.84	92.86	R41	85.71	93.88	90.48
R8	92.86	93.88	93.65	R25	89.29	94.90	93.65	R42	92.86	91.84	92.06
R9	100.00	93.88	92.06	R26	96.43	92.86	93.65	R43	92.86	91.84	92.06
R10	89.29	93.88	92.06	R27	92.86	92.86	92.86	R44	92.86	91.84	92.06
R11	100.00	89.80	92.06	R28	89.29	93.88	91.27	R45	89.29	94.90	93.65
R12	96.43	92.86	93.65	R29	92.86	92.86	92.86	R46	92.86	93.88	91.27
R13	89.29	92.86	92.06	R30	89.29	91.84	91.27	R47	100.00	92.86	92.06
R14	92.86	91.84	91.27	R31	100.00	91.84	92.86	R48	100.00	92.86	92.86
R15	89.29	92.86	92.06	R32	92.86	91.84	92.06	R49	92.86	91.84	92.06
R16	92.86	92.86	92.86	R33	92.86	92.86	92.86	R50	96.43	92.86	92.06
R17	96.43	92.86	92.86	R34	89.29	92.86	92.06	Average	92.14 ± 4.39	92.47 ± 1.23	92.40 ± 0.83

(R = Run; $x \pm y$, x means the mean, y means the standard deviation)

5.5 Comparison with other approaches

To further demonstrate the effectiveness of this proposed "WE + MLP + BBO", we compared it to 6 state-of-the-art approaches in Table 4. Those methods include US + SVD-PCA + SVM-DT [14], BRC + IG + SVM [51], MGM + PEC + SVM [55], EB + WTT + SVM [73], VBM + RF [23], and DF + PCA + SVM [76]. The meaning of these abbreviations can be found in Table 5. For clear view, Fig. 9 presents the corresponding bar plot.

The results in Table 4 show that US + SVD-PCA + SVM-DT [14] and BRC + IG + SVM [51] did not report the standard deviation of three measures. The former obtained an accuracy of 90%, a sensitivity of 94%, and a specificity of 71%. The latter obtains an accuracy of 90.00%, a sensitivity of 96.88%, and a specificity of 77.78%. We can observe their specificities are too low compared to other approaches. Therefore, these two methods are not worthy to be studied.

Other five methods report both the average values and the standard deviation values. MGM + PC + SVM [55] obtained an accuracy of 92.07 ± 1.12%, and a sensitivity of

Table 4 Comparison with State-of-the-art Approaches

Approach	Accuracy	Sensitivity	Specificity
US + SVD-PCA + SVM-DT [14]	90	94	71
BRC + IG + SVM [51]	90.00	**96.88**	77.78
MGM + PC + SVM [55]	92.07 ± 1.12	86.67 ± 4.71	N/A
EB + WTT + SVM [73]	91.47 ± 1.02	90.17 ± 1.66	91.84 ± 1.09
VBM + RF [23]	89.0 ± 0.7	87.9 ± 1.2	90.0 ± 1.1
DF + PCA + SVM [76]	88.27 ± 1.89	84.93 ± 1.21	89.21 ± 1.63
WE + MLP + BBO (Our Method)	**92.40** ± 0.83	92.14 ± 4.39	**92.47** ± 1.23

(Bold means the best)

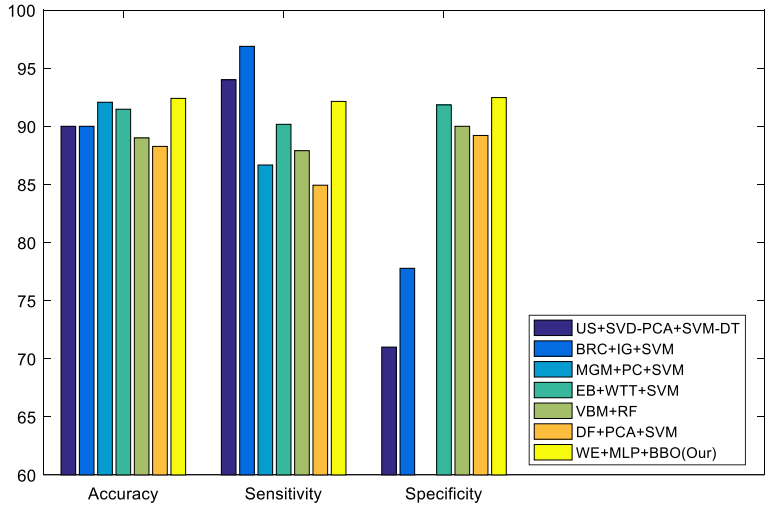


Fig. 9 Bar plot of algorithm comparison (MGM + PC + SVM [55] did not report its specificity)

86.67 ± 4.71. Nevertheless, they did not report anything about the specificity. Thus, it is not further considered in this study.

All the rest algorithms achieved satisfying results. EB + WTT + SVM [73] obtained an accuracy of 91.47 ± 1.02%, a sensitivity of 90.17 ± 1.66%, and a specificity of 91.84 ± 1.09. Their excellent performance contributes to their proposed eigenbrain, which was inspired from the eigenface theory [7] widely used in face recognition.

VBM + RF [23] obtained an accuracy of 89.0 ± 0.7%, a sensitivity of 87.9 ± 1.2%, and a specificity of 90.0 ± 1.1. Their success contributes to the voxel based morphometry. Indeed, VBM has been commonly used to study brain changes. Maguire (2000) [38] showed taxi driver will have larger back part of posterior hippocampus in average. Good (2001) [20] showed global gray matter decreased linearly with age, but the global white matter did not. Nevertheless, it needs accurate spatial normalization, otherwise the classification performance may decrease significantly.

DF + PCA + SVM [76] obtained an accuracy of 88.27 ± 1.89%, a sensitivity of 84.93 ± 1.21%, a specificity of 89.21 ± 1.63%. This method relies on a novel method called displacement field (DF). This research measures and calculates the displace field of different slices between AD patients and HC subjects. Liu (2016) [33] extends DF to three-dimensional. DF method is promising, but it still needs further development to solve several problems: (i) DF is sensitive to noise, i.e., it will fail if the brain extraction result is not clear. (ii) The initial random solution candidate affects the final searched result. (iii) Fast algorithms are expected.

Finally, this proposed "WE + MLP + BBO" achieves the largest accuracy of 92.40% and the largest specificity of 92.47% among all methods. In addition, our method obtains a sensitivity of 92.14%, which is slightly worse than BRC + IG + SVM [51] of 96.88% and US + SVD-PCA + SVM-DT [14] of 94%. Considering all three measures, our method performs better than other six methods. Our method does not propose any new algorithm, and it is a simple combination of mature algorithms. Nevertheless, the result shows "simple is better than complex", as stated in Occam's razor [59]. This also gives us a hint to use

combination of simple methods in other medical applications, such as sensorineural hearing loss [22, 31], multiple sclerosis [77], breast cancer [32], etc.

6 Conclusions and future research

In this paper, our team proposed a new AD identification approach based on wavelet entropy (WE), multilayer perceptron (MLP), and biogeography-based optimization (BBO). This proposed "WE + MLP + BBO" approach yields an accuracy of 92.40%, a sensitivity of 92.14%, and a specificity of 92.47%.

In the future, we will test advanced variants of WE, such as relative wavelet entropy [30], wavelet singular entropy [27]. In addition, some advanced swarm intelligence [1] methods can be used for training MLP. Other image preprocessing methods will be tested to enhance the classification performance, such as image denoising [67], image enhancement [64], and image segmentation [68]. Morphological shared-weight neural network will be employed as an alternative to MLP.

7 Nomenclature

Table 5 Acronym list

Abbreviation	Definition
US	Under-sampling
SVD	Singular value decomposition
PCA	Principal component analysis
DT	decision tree
SVM	support vector machine
BRC	brain region cluster
ICV	inter-class variance
IG	information gain
DBM	deformation-based morphometry
MGM	modulated gray matter
PC	Pearson's correlation
EB	Eigenbrain
WTT	Welch's t-Test
VBM	voxel-based morphometry
RF	random forest
DF	displacement field

Acknowledgements This paper was supported by NSFC (61602250, 61503188), Natural Science Foundation of Jiangsu Province (BK20150983, BK20150982), Open Fund of Key Laboratory of Statistical Information Technology and Data Mining, State Statistics Bureau, (SDL201608), and Open Fund of Fujian Provincial Key Laboratory of Data Intensive Computing (BD201607). The authors express their gratitude to the OASIS dataset supported by NIH grants (P50 MH071616, P01 AG03991, P50AG05681, R01 AG021910, R01 MH56584, and U24 RR021382).

Compliance with ethical standards

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

Appendix 1

Table 6 Stratified cross validation segment over our dataset

3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	2 + 10 = 12	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13
2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13
3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	2 + 10 = 12	3 + 10 = 13
3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12
3 + 10 = 13	2 + 9 = 11	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13
3 + 9 = 12	3 + 9 = 12	2 + 10 = 12	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12
3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12
3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 10 = 13
2 + 10 = 12	3 + 10 = 13	2 + 9 = 11	2 + 10 = 12	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13
3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	2 + 10 = 12	3 + 10 = 13
3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12
3 + 10 = 13	2 + 9 = 11	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13
3 + 9 = 12	3 + 9 = 12	2 + 10 = 12	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12
3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12
3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13
2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13
3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13
3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12
3 + 10 = 13	2 + 9 = 11	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13
3 + 9 = 12	3 + 9 = 12	2 + 10 = 12	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12
3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12
3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	2 + 10 = 12	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13
2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13
3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	2 + 10 = 12	3 + 10 = 13
3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13

Table 6 (continued)

3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12
3 + 10 = 13	2 + 9 = 11	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13
3 + 9 = 12	3 + 9 = 12	2 + 10 = 12	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12
3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12
3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	2 + 10 = 12	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13
2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13
3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	2 + 10 = 12	3 + 10 = 13
3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12
3 + 10 = 13	2 + 9 = 11	3 + 10 = 13	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13
3 + 9 = 12	3 + 9 = 12	2 + 10 = 12	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12

($x + y = z$ represents x AD and y HC, in total z subjects, are segmented in each fold)

Appendix 2

Table 7 Successful identified result over each fold

3 + 10 = 13	2 + 9 = 11	2 + 8 = 10	2 + 9 = 11	3 + 10 = 13	2 + 9 = 11	2 + 9 = 11	3 + 9 = 12	3 + 10 = 13	3 + 7 = 10
2 + 9 = 11	3 + 9 = 12	3 + 10 = 13	1 + 9 = 10	3 + 8 = 11	3 + 8 = 11	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13
3 + 8 = 11	2 + 9 = 11	3 + 8 = 11	3 + 10 = 13	3 + 9 = 12	1 + 8 = 9	3 + 10 = 13	3 + 10 = 13	2 + 8 = 10	1 + 10 = 11
3 + 9 = 12	3 + 10 = 13	2 + 9 = 11	3 + 8 = 11	3 + 9 = 12	2 + 10 = 12	3 + 8 = 11	2 + 9 = 11	1 + 9 = 10	3 + 7 = 10
3 + 10 = 13	1 + 10 = 11	3 + 7 = 10	1 + 10 = 11	3 + 9 = 12	3 + 9 = 12	3 + 9 = 12	3 + 10 = 13	2 + 8 = 10	3 + 9 = 12
2 + 10 = 12	3 + 9 = 12	3 + 8 = 11	3 + 10 = 13	3 + 9 = 12	3 + 9 = 12	2 + 10 = 12	2 + 8 = 10	3 + 10 = 13	1 + 10 = 11
3 + 10 = 13	2 + 9 = 11	3 + 10 = 13	2 + 10 = 12	3 + 7 = 10	1 + 9 = 10	2 + 9 = 11	2 + 9 = 11	2 + 10 = 12	3 + 8 = 11
2 + 10 = 12	3 + 10 = 13	3 + 8 = 11	3 + 7 = 10	3 + 10 = 13	1 + 10 = 11	2 + 8 = 10	3 + 9 = 12	2 + 9 = 11	2 + 10 = 12
1 + 8 = 9	3 + 8 = 11	2 + 9 = 11	2 + 9 = 11	3 + 9 = 12	3 + 10 = 13	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	2 + 8 = 10	0 + 9 = 9	3 + 8 = 11	3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	1 + 9 = 10	2 + 10 = 12	3 + 9 = 12
3 + 8 = 11	3 + 7 = 10	3 + 9 = 12	2 + 9 = 11	3 + 9 = 12	2 + 9 = 11	2 + 9 = 11	3 + 9 = 12	1 + 9 = 10	2 + 10 = 12
2 + 9 = 11	2 + 8 = 10	3 + 9 = 12	2 + 10 = 12	3 + 8 = 11	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	3 + 7 = 10
3 + 10 = 13	2 + 10 = 12	2 + 7 = 9	2 + 10 = 12	3 + 10 = 13	3 + 8 = 11	2 + 8 = 10	3 + 9 = 12	1 + 9 = 10	3 + 10 = 13
1 + 7 = 8	2 + 10 = 12	3 + 9 = 12	1 + 9 = 10	3 + 10 = 13	3 + 10 = 13	1 + 10 = 11	3 + 8 = 11	3 + 9 = 12	3 + 7 = 10
3 + 9 = 12	2 + 9 = 11	3 + 8 = 11	2 + 10 = 12	3 + 9 = 12	2 + 10 = 12	3 + 10 = 13	2 + 9 = 11	2 + 9 = 11	2 + 8 = 10
2 + 10 = 12	3 + 10 = 13	0 + 7 = 7	3 + 9 = 12	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	3 + 9 = 12	3 + 7 = 10	1 + 10 = 11
3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	3 + 8 = 11	1 + 8 = 9	2 + 9 = 11	3 + 10 = 13	2 + 9 = 11	2 + 10 = 12	2 + 9 = 11
3 + 9 = 12	3 + 10 = 13	1 + 9 = 10	2 + 9 = 11	3 + 7 = 10	3 + 9 = 12	3 + 9 = 12	2 + 9 = 11	3 + 10 = 13	2 + 10 = 12
3 + 10 = 13	2 + 9 = 11	3 + 10 = 13	2 + 10 = 12	3 + 8 = 11	3 + 9 = 12	1 + 8 = 9	3 + 10 = 13	3 + 8 = 11	1 + 8 = 9
2 + 10 = 12	3 + 10 = 13	3 + 10 = 13	2 + 10 = 12	2 + 10 = 12	3 + 9 = 12	2 + 9 = 11	3 + 9 = 12	3 + 8 = 11	2 + 8 = 10
3 + 9 = 12	1 + 8 = 9	3 + 10 = 13	2 + 9 = 11	0 + 10 = 10	2 + 9 = 11	3 + 10 = 13	3 + 9 = 12	2 + 9 = 11	3 + 8 = 11
1 + 9 = 10	3 + 9 = 12	2 + 8 = 10	2 + 9 = 11	3 + 9 = 12	3 + 9 = 12	2 + 9 = 11	2 + 9 = 11	3 + 8 = 11	3 + 10 = 13
2 + 6 = 8	3 + 10 = 13	1 + 10 = 11	2 + 9 = 11	1 + 10 = 11	3 + 10 = 13	2 + 10 = 12	2 + 8 = 10	3 + 10 = 13	3 + 7 = 10
3 + 9 = 12	2 + 9 = 11	3 + 9 = 12	1 + 10 = 11	3 + 10 = 13	3 + 9 = 12	2 + 9 = 11	3 + 7 = 10	2 + 9 = 11	3 + 9 = 12
1 + 9 = 10	3 + 9 = 12	2 + 9 = 11	2 + 10 = 12	2 + 9 = 11	2 + 10 = 12	1 + 9 = 10	3 + 10 = 13	3 + 8 = 11	3 + 9 = 12
3 + 10 = 13	3 + 9 = 12	2 + 8 = 10	3 + 9 = 12	0 + 10 = 10	3 + 8 = 11	2 + 7 = 9	3 + 10 = 13	3 + 10 = 13	3 + 10 = 13
3 + 10 = 13	3 + 9 = 12	1 + 9 = 10	2 + 10 = 12	2 + 9 = 11	2 + 9 = 11	3 + 9 = 12	3 + 9 = 12	3 + 8 = 11	3 + 9 = 12
3 + 10 = 13	3 + 9 = 12	2 + 10 = 12	1 + 6 = 7	3 + 9 = 12	3 + 9 = 12	2 + 9 = 11	2 + 10 = 12	2 + 10 = 12	3 + 9 = 12
2 + 9 = 11	2 + 10 = 12	3 + 9 = 12	2 + 9 = 11	3 + 10 = 13	2 + 7 = 9	2 + 10 = 12	3 + 9 = 12	3 + 8 = 11	3 + 8 = 11
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	0 + 8 = 8	2 + 8 = 10	3 + 9 = 12	2 + 8 = 10	3 + 9 = 12	3 + 9 = 12	2 + 10 = 12
3 + 10 = 13	3 + 9 = 12	3 + 10 = 13	3 + 8 = 11	3 + 10 = 13	2 + 10 = 12	1 + 8 = 9	3 + 10 = 13	3 + 9 = 12	3 + 6 = 9

Table 7 (continued)

1 + 9 = 10	2 + 10 = 12	2 + 9 = 11	3 + 9 = 12	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	3 + 7 = 10	1 + 8 = 9	3 + 10 = 13
3 + 9 = 12	2 + 9 = 11	2 + 10 = 12	3 + 9 = 12	2 + 8 = 10	3 + 10 = 13	3 + 8 = 11	2 + 9 = 11	3 + 9 = 12	2 + 10 = 12
3 + 9 = 12	0 + 9 = 9	3 + 10 = 13	2 + 9 = 11	1 + 6 = 7	3 + 9 = 12	3 + 9 = 12	2 + 10 = 12	3 + 9 = 12	3 + 10 = 13
2 + 10 = 12	1 + 10 = 11	3 + 8 = 11	2 + 9 = 11	2 + 9 = 11	3 + 8 = 11	2 + 10 = 12	2 + 8 = 10	3 + 10 = 13	3 + 9 = 12
3 + 9 = 12	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12	2 + 8 = 10	1 + 10 = 11	3 + 9 = 12	3 + 7 = 10	2 + 10 = 12	3 + 9 = 12
3 + 9 = 12	3 + 8 = 11	1 + 9 = 10	3 + 8 = 11	3 + 7 = 10	3 + 10 = 13	2 + 10 = 12	3 + 10 = 13	0 + 9 = 9	3 + 10 = 13
1 + 9 = 10	3 + 9 = 12	3 + 7 = 10	3 + 10 = 13	3 + 9 = 12	3 + 7 = 10	1 + 10 = 11	3 + 10 = 13	2 + 9 = 11	2 + 10 = 12
2 + 10 = 12	3 + 10 = 13	2 + 10 = 12	2 + 8 = 10	2 + 8 = 10	3 + 9 = 12	3 + 8 = 11	3 + 10 = 13	3 + 7 = 10	3 + 9 = 12
1 + 10 = 11	3 + 9 = 12	2 + 8 = 10	2 + 10 = 12	3 + 9 = 12	2 + 10 = 12	3 + 9 = 12	3 + 7 = 10	3 + 10 = 13	1 + 9 = 10
3 + 10 = 13	2 + 8 = 10	3 + 9 = 12	2 + 8 = 10	3 + 10 = 13	2 + 10 = 12	0 + 9 = 9	3 + 10 = 13	3 + 9 = 12	2 + 8 = 10
3 + 8 = 11	2 + 9 = 11	2 + 9 = 11	0 + 10 = 10	3 + 10 = 13	3 + 8 = 11	3 + 9 = 12	3 + 9 = 12	2 + 9 = 11	3 + 9 = 12
1 + 10 = 11	3 + 9 = 12	3 + 9 = 12	2 + 8 = 10	2 + 9 = 11	3 + 7 = 10	2 + 9 = 11	3 + 9 = 12	3 + 10 = 13	2 + 9 = 11
3 + 10 = 13	2 + 9 = 11	0 + 8 = 8	2 + 9 = 11	3 + 10 = 13	3 + 10 = 13	3 + 6 = 9	3 + 10 = 13	3 + 9 = 12	2 + 8 = 10
3 + 10 = 13	3 + 8 = 11	3 + 10 = 13	2 + 10 = 12	2 + 9 = 11	3 + 9 = 12	0 + 8 = 8	3 + 9 = 12	3 + 10 = 13	3 + 9 = 12
3 + 10 = 13	2 + 10 = 12	3 + 7 = 10	3 + 10 = 13	2 + 9 = 11	2 + 8 = 10	3 + 10 = 13	3 + 10 = 13	2 + 8 = 10	2 + 7 = 9
3 + 9 = 12	3 + 9 = 12	2 + 10 = 12	2 + 8 = 10	2 + 10 = 12	3 + 8 = 11	3 + 10 = 13	2 + 9 = 11	3 + 10 = 13	3 + 7 = 10
2 + 9 = 11	3 + 10 = 13	2 + 10 = 12	2 + 10 = 12	2 + 9 = 11	3 + 6 = 9	3 + 8 = 11	2 + 10 = 12	2 + 9 = 11	3 + 10 = 13
2 + 9 = 11	1 + 10 = 11	3 + 8 = 11	1 + 9 = 10	3 + 8 = 11	3 + 9 = 12	3 + 10 = 13	3 + 10 = 13	2 + 9 = 11	3 + 8 = 11
2 + 10 = 12	3 + 9 = 12	2 + 10 = 12	3 + 10 = 13	2 + 9 = 11	2 + 8 = 10	3 + 8 = 11	2 + 10 = 12	3 + 7 = 10	3 + 10 = 13

($x + y = z$ represents x AD and y HC, in total z subjects, are identified in each fold)

References

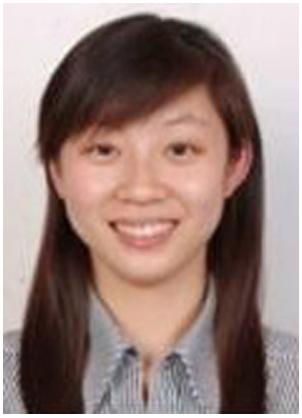
1. Agarwal P et al (2013) Swarm intelligence and its applications. Sci World J 2013:528069
2. Aggarwal N et al (2015) 3d discrete wavelet transform for computer aided diagnosis of Alzheimer's disease using t1-weighted brain MRI. Int J Imaging Syst Technol 25(2):179–190
3. Ardekani BA et al (2013) Sexual dimorphism in the human corpus callosum: an MRI study using the OASIS brain database. Cereb Cortex 23(10):2514–2520
4. Bakhshi AD et al (2013) Application of continuous-time wavelet entropy for detection of cardiac repolarisation alternans. IET Signal Processing 7(8):783–790
5. Balochian S (2014) Artificial intelligence and its applications. Mathematical problems in engineering Article ID: 840491
6. Behera NKS et al. (2015) Bird mating optimization based multilayer perceptron for diseases classification. In: 1st International Conference on Computational Intelligence in Data Mining (ICCIDM) Burla, India, Springer-Verlag Berlin, p 272–278
7. Bhuiyan MAA (2016) Towards face recognition using eigenface. Int J Adv Comput Sci Appl 7(5):25–31
8. Bozorg Haddad O et al (2016) Biogeography-based optimization algorithm for optimal operation of reservoir systems. J Water Resour Plan Manag 142(1):04015034
9. Bradley PS (2013) A support-based reconstruction for SENSE MRI. Sensors 13(4):4029–4040
10. Candra H et al. (2015) Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine. In: 37th Annual International Conference Of the Ieee Engineering In Medicine And Biology Society. Milan, Italy, IEEE p 7250–7253
11. Chen Y et al (2016) Curve-like structure extraction using minimal path propagation with back-tracing. IEEE Trans Image Process 25(2):988–1003
12. De Visschere P et al (2015) Prostate magnetic resonance spectroscopic imaging at 1.5 tesla with endorectal coil versus 3.0 tesla without endorectal coil: comparison of spectral quality. Clin Imaging 39(4):636–641
13. Dil EA et al (2016) Trace determination of safranin O dye using ultrasound assisted dispersive solid-phase micro extraction: artificial neural network-genetic algorithm and response surface methodology. Ultrason Sonochem 33:129–140
14. Dong Z (2014) Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. Prog Electromagn Res 144:171–184
15. Du S (2016) Multi-objective path finding in stochastic networks using a biogeography-based optimization method. Simulation 92(7):637–647
16. Fang L, Wu L (2015) A novel demodulation system based on continuous wavelet transform. Math Probl Eng 2015:513849
17. Farswan P et al. (2016) A modified biogeography based optimization. In: 2nd International Conference on Harmony Search Algorithm (ICHSA), Korea Univ, Seoul, South Korea: Springer-Verlag Berlin, p 227–238
18. Frantzidis CA et al (2014) Functional disorganization of small-world brain networks in mild Alzheimer's disease and amnestic mild cognitive impairment: an EEG study using relative wavelet entropy (RWE). Front Aging Neurosci 6:224
19. 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
20. Good CD et al (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage 14(1):21–36
21. Gorji HT, Haddadnia J (2015) A novel method for early diagnosis of Alzheimer's disease based on pseudo Zernike moment from structural MRI. Neuroscience 305:361–371
22. Gorriz JM, Ramírez J (2016) Wavelet entropy and directed acyclic graph support vector machine for detection of patients with unilateral hearing loss in MRI scanning. Front Comput Neurosci 2016(10):160
23. Gray KR et al (2013) Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. NeuroImage 65:167–175
24. Heidari AA et al. (2015) An effective hybrid support vector regression with chaos-embedded biogeography-based optimization strategy for prediction of earthquake-triggered slope deformations. In: International Conference on Sensors & Models In Remote Sensing & Photogrammetry, Kish Island, Iran, Copernicus Gesellschaft Mbh, p 301–305
25. Ibanez F et al (2015) Detection of damage in multiwire cables based on wavelet entropy evolution. Smart Mater Struct 24(8):14 Article ID: 085036
26. Ibrahim AO et al. (2015) Intelligent multi-objective classifier for breast cancer diagnosis based on multilayer perceptron neural network and differential evolution. In: International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE), Khartoum, Sudan, IEEE, p 422–427

27. Ji TY et al. (2014) Disturbance detection using hit-or-miss wavelet singular entropy for power quality monitoring. In: IEEE Power and Energy Society General Meeting PESGM, National Harbor, MD, IEEE, p 46–52
28. Jiang WJ et al (2016) Parameters identification of fluxgate magnetic Core adopting the biogeography-based optimization algorithm. Sensors 16(7):979
29. Krawczyk B et al (2015) A hybrid cost-sensitive ensemble for imbalanced breast thermogram classification. Artif Intell Med 65(3):219–227
30. Lee SG et al (2014) Reference-free damage detection for truss bridge structures by continuous relative wavelet entropy method. Structural Health Monitoring-an International Journal 13(3):307–320
31. Li J (2016) Detection of left-sided and right-sided hearing loss via fractional Fourier transform. Entropy 18(5):194
32. Liu G (2016) Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform. Adv Mech Eng 8(2):11
33. Liu G et al (2016) Detection of Alzheimer's disease by three-dimensional displacement field estimation in structural magnetic resonance imaging. J Alzheimers Dis 50(1):233–248
34. Lu HM et al (2012) Maximum local energy: an effective approach for multisensor image fusion in beyond wavelet transform domain. Computers & Mathematics with Applications 64(5):996–1003
35. Lu HM et al (2016a) Turbidity underwater image restoration using spectral properties and light compensation. IEICE Trans Inf Syst E99D(1):219–227
36. Lu HM et al (2016b) Underwater image enhancement method using weighted guided trigonometric filtering and artificial light correction. J Vis Commun Image Represent 38:504–516
37. Magnander T et al (2016) A novel statistical analysis method to improve the detection of hepatic foci of (111)In-octreotide in SPECT/CT imaging. EJNMMI Physics 3(1):1
38. Maguire EA et al (2000) Navigation-related structural change in the hippocampi of taxi drivers. Proc Natl Acad Sci U S A 97(8):4398–4403
39. Makbol NM et al (2016) Block-based discrete wavelet transform-singular value decomposition image watermarking scheme using human visual system characteristics. IET Image Process 10(1):34–52
40. Mashhadban H et al (2016) Prediction and modeling of mechanical properties in fiber reinforced self-compacting concrete using particle swarm optimization algorithm and artificial neural network. Constr Build Mater 119:277–287
41. Meng GL et al (2015) Meteorological factors related to emergency admission of elderly stroke patients in shanghai: analysis with a multilayer perceptron neural network. Med Sci Monit 21:3600–3607
42. Mirjalili S et al (2014) Let a biogeography-based optimizer train your multi-layer perceptron. Inf Sci 269:188–209
43. Mondal U et al (2016) Servomechanism for periodic reference input: discrete wavelet transform-based repetitive controller. Trans Inst Meas Control 38(1):14–22
44. Park JS, Ju I (2016) Prescription drug advertising, disease knowledge, and older adults' optimistic bias about the future risk of alzheimer's disease. Health Commun 31(3):346–354
45. Peng, I.B., et al. (2015) The cost of synchronizing imbalanced processes in message passing systems. In: International Conference on Cluster Computing, Chicago, IL, IEEE, p 408–417
46. Peng B et al (2016) Image processing methods to elucidate spatial characteristics of retinal microglia after optic nerve transection. Sci Rep 6:21816
47. Peters S et al (2013) Detection of irreversible changes in susceptibility-weighted images after whole-brain irradiation of children. Neuroradiology 55(7):853–859
48. Peterson BS (2011) A two-level iterative reconstruction method for compressed sensing MRI. Journal of Electromagnetic Waves and Applications 25(8–9):1081–1091
49. Peterson BS (2014) Energy preserved sampling for compressed sensing MRI. Comput Math Methods Med 2014:546814
50. Phillips P et al (2015) Pathological brain detection in magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimization. Prog Electromagn Res 152:41–58
51. Plant C et al (2010) Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. NeuroImage 50(1):162–174
52. Pu X, He W (2015) Chaotic biogeography-based optimization algorithm for job scheduler in cloud computing. In: International Conference on Mechanical Science and Mechanical Design, Destech Publications, Changsha, Peoples R China, p 223–229
53. Rajchl M et al (2016) Hierarchical max-flow segmentation framework for multi-atlas segmentation with Kohonen self-organizing map based Gaussian mixture modeling. Med Image Anal 27:45–56

54. Saghatforoush A et al (2016) Combination of neural network and ant colony optimization algorithms for prediction and optimization of flyrock and back-break induced by blasting. Eng Comput 32(2):255–266
55. Savio A, Grana M (2013) Deformation based feature selection for computer aided diagnosis of Alzheimer's disease. Expert Syst Appl 40(5):1619–1628
56. Shamshirband S et al (2016) Estimation of reference evapotranspiration using neural networks and cuckoo search algorithm. J Irrig Drain Eng 142(2):04015044
57. Shiyang L et al (2007) Analysis of heart rate fluctuation based on wavelet entropy. Fluctuation and Noise Letters 7(2):L135–L142
58. Sonawane JS, Patil DR (2014) Prediction of heart disease using multilayer perceptron neural network. In: IEEE International Conference on Information Communication and Embedded Systems, Chennai, India, p 5–11
59. Sterkenburg TF (2016) Solomonoff prediction and Occam's razor. Philos Sci 83(4):459–479
60. Sun P (2015) Pathological brain detection based on wavelet entropy and Hu moment invariants. Biomed Mater Eng 26(s1):1283–1290
61. Torrents-Barrena J et al (2015) Complex wavelet algorithm for computer-aided diagnosis of Alzheimer's disease. Electron Lett 51(20):1566–1567
62. Wang L et al (2014) The effect of APOE epsilon 4 allele on cholinesterase inhibitors in patients with Alzheimer disease evaluation of the feasibility of resting state functional connectivity magnetic resonance imaging. Alzheimer Dis Assoc Disord 28(2):122–127
63. Watamura N et al (2016) Colocalization of phosphorylated forms of WAVE1, CRMP2, and tau in Alzheimer's disease model mice: involvement of Cdk5 phosphorylation and the effect of ATRA treatment. J Neurosci Res 94(1):15–26
64. Wei G (2010) Color image enhancement based on HVS and PCNN. Science China Inf Sci 53(10):1963–1976
65. Wei L (2015) Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization. Entropy 17(8):5711–5728
66. Wilkins HM, Swerdlow RH (2016) Relationships between mitochondria and Neuroinflammation: implications for Alzheimer's disease. Curr Top Med Chem 16(8):849–857
67. Wu L (2008) Improved image filter based on SPCNN. Science in China Series F: Information Sciences 51(12):2115–2125
68. Wu L (2011) Optimal multi-level Thresholding based on maximum Tsallis entropy via an artificial bee Colony approach. Entropy 13(4):841–859
69. Wu L (2012) An MR brain images classifier via principal component analysis and kernel support vector machine. Prog Electromagn Res 130:369–388
70. Wu J (2016a) Fruit classification by biogeography-based optimization and feedforward neural network. Expert Syst 33(3):239–253
71. Wu X (2016b) Smart detection on abnormal breasts in digital mammography based on contrast-limited adaptive histogram equalization and chaotic adaptive real-coded biogeography-based optimization. Simulation 92(9):873–885
72. Yang G et al (2015) Automated classification of brain images using wavelet-energy and biogeography-based optimization. Multimedia Tools and Applications. doi:10.1007/s11042-015-2649-7
73. Yuan TF (2015) Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. Front Comput Neurosci 9:66
74. Zainuddin Z, Fard SP (2015) Approximation of multivariate 2 pi-periodic functions by multiple 2 pi-periodic approximate identity neural networks based on the universal approximation theorems. In: 11th International Conference on Natural Computation. Zhangjiajie, Peoples R China, IEEE, p 8–13
75. Zhan T (2016) Pathological brain detection by artificial intelligence in magnetic resonance imaging scanning. Prog Electromagn Res 156:105–133
76. Zhang Y (2015) Detection of Alzheimer's disease by displacement field and machine learning. PeerJ 3:e1251
77. Zhou X-X (2016) Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine. Simulation 92(9):861–871
78. Zhou X-X et al (2016) Detection of abnormal MR brains based on wavelet entropy and feature selection. IEEJ Trans Electr Electron Eng 11(3):1–10
79. Zou YC et al (2015) Wavelet entropy based analysis and forecasting of crude oil price dynamics. Entropy 17(10):7167–7184

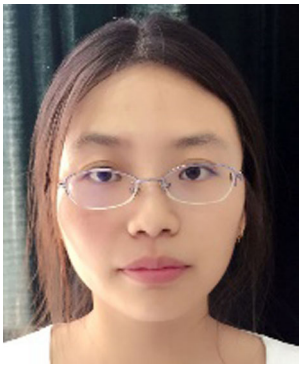
Dr. Shuihua Wang received a B.S. from Southeast University in 2008 and a M.S. from The City University of New York in 2012. She received her Ph.D. from Nanjing University in 2016. At present, she works as a lecturer in Nanjing Normal University. Her research interest is biomedical imaging.

Dr. Yin Zhang is an Assistant Professor of the School of Information and Safety Engineering, Zhongnan University of Economics and Law, China. He was a Poster-Doctoral Fellow in the School of Computer Science and Technology at Huazhong University of Science and Technology, China. He is Vice-chair of IEEE Computer Society Big Data STC. His research interests include data analysis, data mining, healthcare big data and social network. He has published more than 30 prestigious conference and journal papers. He serves as Guest Editor for IEEE Sensors Journal and New Review of Hypermedia and Multimedia. He also served as TPC Co-Chair of CloudComp 2015, and Local Chair of TRIDENTCOM 2014.

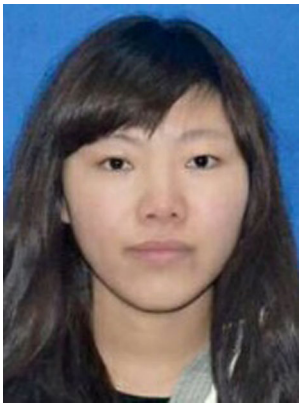


Dr. Yujie Li receivedthe B.S. degree in Computer Science and Technology from Yangzhou University in 2009. She received M.S. degrees in Electrical Engineering from Kyushu Institute of Technology and Yangzhou University in 2012, respectively. She received the Ph.D. degree from Kyushu Institute of Technology in 2015. Recently, she is a lecturer in Yangzhou University. Her research interests include computer vision, sensors, and image segmentation.

Wen-Juan Jia received B.S. in Yancheng Teachers University from the department of Mathematics & Statistics 2016. Now she is pursuing the M.S. in School of Computer Science & Technology, Nanjing Normal University. Her research interest is expert system.



Fang-Yuan Liu received B.S. from Jiangsu University at 2016. Now she is pursuing the M.S. at Nanjing Normal University in Computer Science & Technology. Her research interest is medical image processing.



Meng-Meng Yang received B.E. in Zhengzhou Normal University of Computer Science & Technology 2016. Now she is pursuing the Master of Engineering in Nanjing Normal University in the field of computer technology. Her research interest is image processing.

Prof. Dr. Yudong Zhang got his Ph.D. degree from Southeast University at 2010.He worked in Columbia University as a postdoc from 2010 to 2012. He worked in the same workplace from 2012 to 2013 as a research scientist. Now he is a professor in Nanjing Normal University. His research interest is pattern recognition.