


Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm

Yu-Dong Zhang¹  • Guihu Zhao² • Junding Sun¹ •
Xiaosheng Wu¹ • Zhi-Heng Wang¹ • Hong-Min Liu¹ •
Vishnu Varthanan Govindaraj³ • Tianmin Zhan⁴ •
Jianwu Li⁵

Received: 21 May 2017 / Revised: 20 June 2017 / Accepted: 7 July 2017
© Springer Science+Business Media, LLC 2017

Abstract Pathological brain detection is an automated computer-aided diagnosis for brain images. This study provides a novel method to achieve this goal. We first used synthetic minority oversampling to balance the dataset. Then, our system was based on three components: wavelet packet Tsallis entropy, extreme learning machine, and Jaya algorithm. The 10 repetitions of K-fold cross validation showed our method achieved perfect classification on two small datasets, and achieved a sensitivity of $99.64 \pm 0.52\%$, a specificity of $99.14 \pm 1.93\%$, and an accuracy of $99.57 \pm 0.57\%$ over a 255-image

Yu-Dong Zhang, Guihu Zhao and Junding Sun contributed equally to this work.

- ✉ Yu-Dong Zhang
yudongzhang@ieee.org
- ✉ Vishnu Varthanan Govindaraj
gvvarthanan@gmail.com
- ✉ Tianmin Zhan
ztm@ujs.edu.cn
- ✉ Jianwu Li
ljw@bit.edu.cn

¹ School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, People's Republic of China

² School of Information Science and Engineering, Central South University, Changsha 410083, China

³ Department of Instrumentation and Control Engineering, Kalasalingam University, Srivilliputtur, Virudhunagar, Tamil Nadu, India

⁴ School of Technology, Nanjing Audit University, Nanjing, Jiangsu 211815, China

⁵ School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

dataset. Our method performs better than six state-of-the-art approaches. Besides, Jaya algorithm performs better than genetic algorithm, particle swarm optimization, and bat algorithm as ELM training method.

Keywords Pathological brain detection · Synthetic minority oversampling · Extreme learning machine · Jaya algorithm

1 Introduction

Pathological brain detection (PBD) is an important and smart technique that can assist neuroradiologists to make diagnosis quickly and rapidly [36]. It is now applied into practical gradually. For example, many hospitals have used PBD systems to detect autism spectrum disorder [8], Alzheimer's disease [34], mild cognitive impairment [20], multiple sclerosis [37], microbleeding [1, 3], hearing loss [18], etc.

Currently, scholars in academics tend to develop an overall system that can detect all brain diseases based on latest computer techniques [13, 28, 33]. This overall system is composed of two stages. In the first stage, the system can identify all brain diseases from healthy brains. In the second stage, the system will identify particular brain diseases.

The first stage is now at thorough research. In the last decade, many literatures were published related over the first-stage research. For example, Sun (2016) [25] proposed a novel feature—fractional Fourier entropy, and used multilayer perceptron. They suggested to use biogeography-based optimization (BBO) to train the classifier. Yang (2016) [32] utilized dual-tree complex wavelet transform and support vector machine. Lu (2016) [15] employed radial basis function neural network. Chen (2016) [2] used Minkowski-Bouligand dimension to identify pathological brains. They used an improved particle swarm optimization (IPSO) to train the classifier. Chen and Du (2017) [4] employed wavelet packet Tsallis entropy. They used feedforward neural network (FNN) and an improved biogeography-based optimization (IBBO). Jiang and Zhu (2017) [12] employed pseudo Zernike moment and support vector machine. Wang and Lv (2016) [27] used seven Hu moment invariants and predator-prey particle swarm optimization.

There are two problems in above literature. First, their datasets were selected from the “The Whole Brain Atlas (TWBA)” dataset, which is unbalanced. Hence, the accuracy paradox may influence the classification performance. Second, they used complicated classifiers, which are not necessary in a PBD situation.

The contributions in this study aimed to solve above problems. On one hand, we used synthetic minority oversampling technique to balance the dataset. Undersampling method is not used, since it will reduce the size of dataset. On the other hand, we used a simple but effective classifier—extreme learning machine, and we proposed a new training algorithm.

The structure of the rest is organized as follows: Section 2 presents the materials and preprocessing method by synthetic minority oversampling technique. Section 3 offers the description of our methodology. Section 4 describes the results and provides the discussions. Finally, Section 5 concludes the paper.

2 Materials and preprocessing

2.1 Dataset

Three datasets were commonly used in PBD experiments and comparisons. They were all downloaded from TWBA datasets. Each of those axial slices was picked up by neuroradiologists with more than ten-year experiences. The first dataset (FD) includes 66 brain slices, the second dataset (SD) includes 160 brain slices, and the third dataset (TD) includes 255 brain slices. The descriptions over the three datasets are offered in reference [6]. Eleven pathological brain slices are illustrated in Fig. 1. These diseases include Alzheimer's disease (AD), Pick's disease (PD), Huntington's disease (HD), Multiple sclerosis (MS), Cerebral toxoplasmosis (CT), Subdural hematoma (SH), Herpes encephalitis (HE), AD with visual agnosia (ADVA), meningioma, glioma, and sarcoma.

2.2 Synthetic minority oversampling technique

As the number of pathological brain slices is much larger than the number of healthy brain slices in all three datasets, we used Synthetic Minority Oversampling Technique (SMOTE) [17] to increase the number of healthy brain slices. We first take a sample *A* from the minority

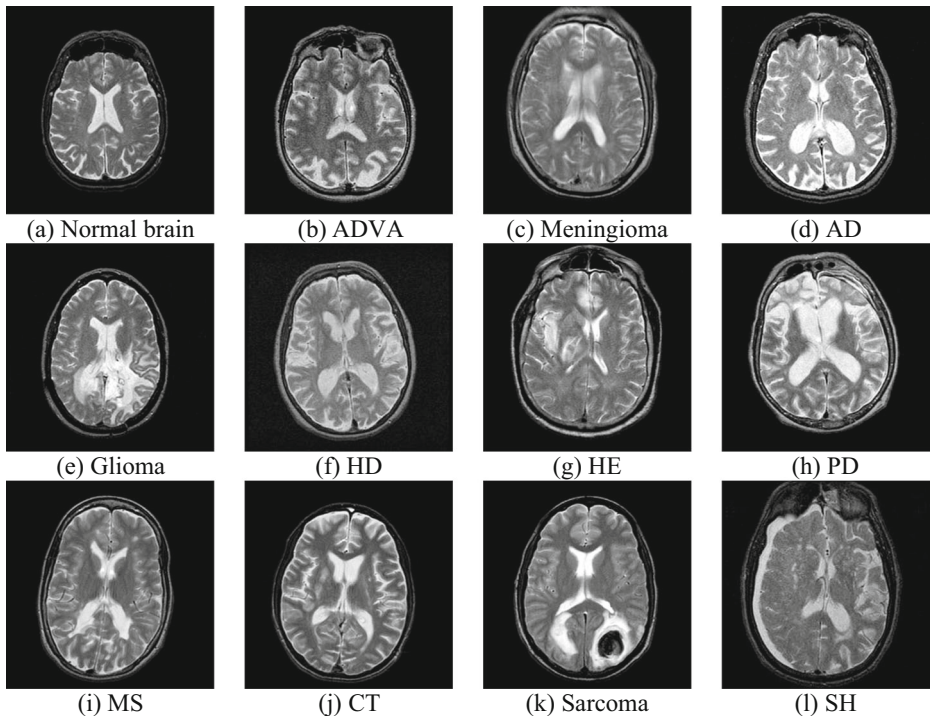


Fig. 1 Pathological brain slices

class, and select its k nearest neighbors. We select one neighbor B from the k neighbors. We then draw a vector V from A to B .

$$V = B - A \quad (1)$$

Finally, a new synthetic sample S is generated as

$$S = A + \beta \times V \quad (2)$$

where β is a random number in the range of $[0, 1]$, obeying the uniform distribution.

Figure 2 illustrates the idea of SMOTE. The red circle represents sample A , which has five samples around it with shape of square. We randomly select a sample B that is covered by yellow square. The β here is chosen as 0.6. Finally, the new sample S is obtained with shape of a blue diamond.

3 Methodology

3.1 Wavelet packet Tsallis entropy

Wavelet entropy [21, 24, 38] is widely used in brain slice identification. This study used an improved version—wavelet packet Tsallis entropy (WPTE), which is a combination of discrete wavelet packet transform (DWPT) and Tsallis entropy (TE).

Standard discrete wavelet transform (DWT) pass only the previous approximation subband (A) through quadratic mirror filters. Nevertheless, the DWPT pass both approximation subband, and three detail subband to the filters. Thus, for an n -level decomposition, DWT provides $(3n + 1)$ subbands, while DWPT provides 2^n subbands [16].

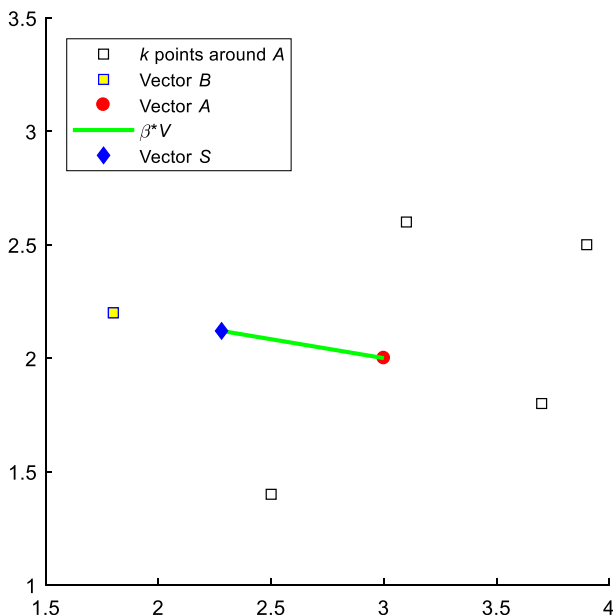


Fig. 2 Illustration of SMOTE

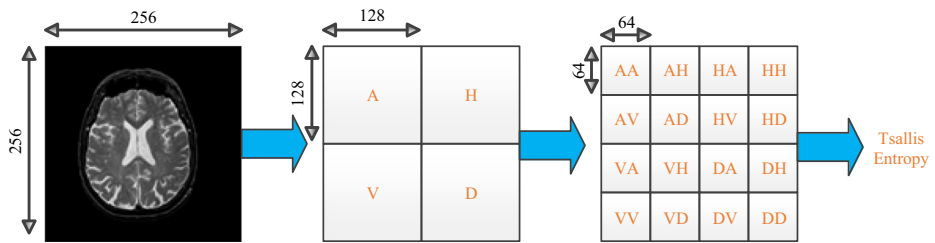


Fig. 3 Diagram of a 2-level wavelet packet Tsallis entropy

Tsallis entropy is an improved version of traditional Shannon entropy. Assume p represents the probability mass function, h the graylevel value of any subband coefficient, H the total number of graylevel values. We can define the Shannon entropy E_S as

$$E_S = - \sum_{h=1}^H p_h \log_2(p_h) \quad (3)$$

The Tsallis entropy [26] E_T introduced a new entropic-index parameter q , and its definition is written as

$$E_T = \frac{1 - \sum_{h=1}^H (p_h)^q}{q-1} \quad (4)$$

Parameter q measures the nonextensivity degree of the probability mass function. When q approximates to 1, the Tsallis entropy will degrade to Shannon entropy [14]. In this study, q is assigned with the value of 0.8, following the setting in reference [4].

Figure 3 shows the diagram of WPTE, where A means approximation, H horizontal, V vertical, and D diagonal. We first import a brain slice image, then we carry out a 3-level discrete wavelet packet transform (abbreviated as DWPT) [30]. We store the coefficients of all 64 subbands generated by the 3-level DWPT decomposition. We calculate the Tsallis entropy over each subband. Finally, we output a 64-element vector for each brain image.

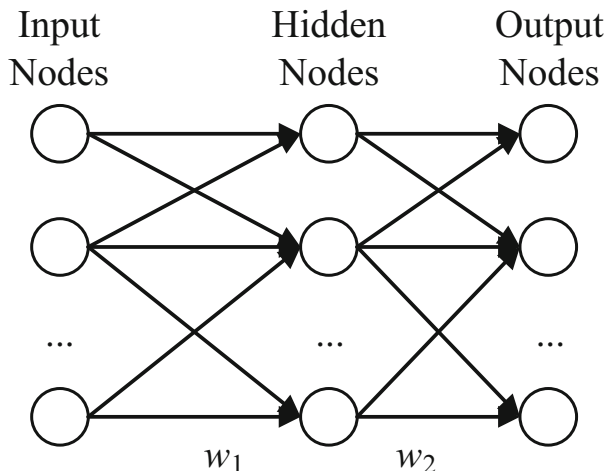


Fig. 4 Structure of ELM

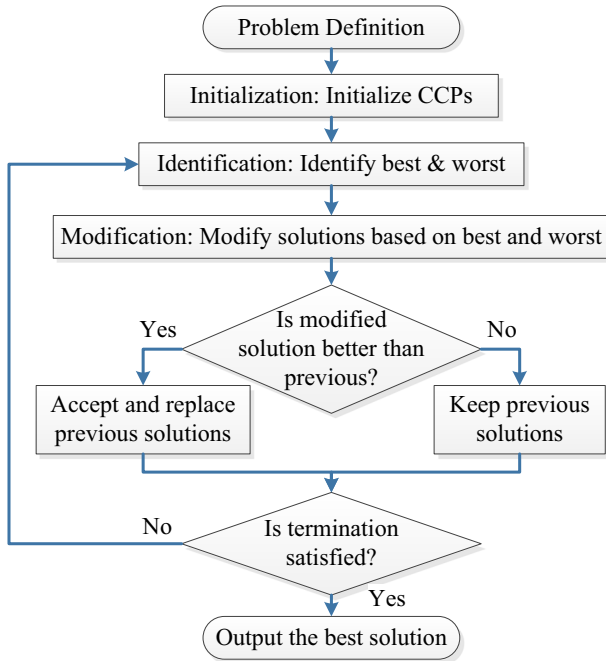


Fig. 5 Flowchart of Jaya algorithm

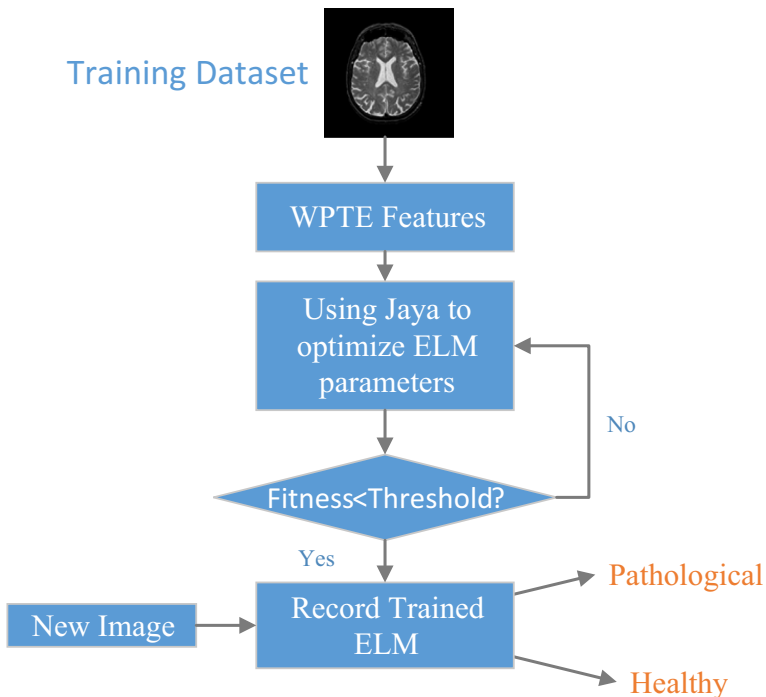


Fig. 6 Block diagram of the whole system, which identifies two classes: pathological brain or healthy brain

Table 1 Cross validation setting of the whole dataset (Without SMOTE)

Dataset	No. of Fold	No. of Repetition	Training			Validation			Total		
			H	P	T	H	P	T	H	P	T
FD	6	10	15	40	55	3	8	11	18	48	66
SD	5	10	16	112	128	4	28	32	20	140	160
TD	5	10	28	176	204	7	44	51	35	220	255

H Healthy, *P* Pathological, *T* Total

3.2 Extreme learning machine

Many methods can serve as the classifier. The support vector machine (SVM) used support vectors to create hyperplanes, but it is difficult to determine the kernel function and kernel parameters. The artificial neural network (ANN) exhibits powerful nonlinear mapping ability, but it may be easily overfitting.

The emerging extreme learning machine (ELM) techniques exhibit the advantages of fast learning, simple calculation, strong generalization ability. It has been validated to give better performance than both SVM and ANN in visual servoing [35], regression [5], etc.

The structure of ELM is a single hidden layer feedforward neural network. The weights w_1 connects input layer and hidden layer, and meanwhile w_1 is randomly assigned and never updated. The weights w_2 connects hidden nodes and output nodes, it is learned within a single iteration by:

$$w_2 = h(w_1 X)^+ Y \quad (5)$$

where h is the activation function, X is the design matrix, Y is the response variable, $()^+$ is the pseudoinverse operation. Figure 4 shows its structure.

3.3 Training of ELM—Jaya algorithm

Although ELM has a closed-form solution as in Eq. (5); in practical, scholars tend to use iterative methods in order to avoid overfitting. Bioinspired methods commonly served as the iterative methods. Rong, Zhang, Chang (2016) [23] used genetic algorithm (GA) to train ELM. Yadav, Ch, Mathur (2016) [29] employed particle swarm optimization (PSO) to train ELM. Doreswamy and Salma (2015) [7] utilized bat algorithm to update the weights of ELM.

Table 2 Pseudocode of proposed method

```

for  $i = 1 : 10$ 
  Divide the dataset into  $K$  folds
  for  $j = 1 : K$ 
    Choose Fold  $j$  as the test set, the rest folds as the training set;
    Use SMOTE to balance the dataset;
    Extract 64 WPTE feature from each image;
    Use ELM as the classifier, and Jaya as the training algorithm;
    Report the confusion matrix  $C[i, j]$  over the test set.
  end
  Obtain the confusion matrix at  $i$ -th run as  $C[i] = C[i, 1] + C[i, 2] + \dots + C[i, K]$ 
  Report the sensitivity, specificity, and accuracy based on  $C[i]$ .
end
Report the mean and standard deviation of sensitivity, specificity, and accuracy.

```

Table 3 SMOTE setting of training dataset

Dataset	Before SMOTE		After SMOTE	
	Healthy (Minority Class)	Pathological (Majority Class)	Healthy	Pathological
FD	15	40	40	40
SD	16	112	112	112
TD	28	176	176	176

GA and PSO are powerful optimization algorithms, and they achieved great successes in training ELMs. Nevertheless, GA and PSO have a relatively slow convergence speed. Recently, Rao, Savsani and Vakharia (2011) [22] proposed a novel Jaya algorithm that does not require algorithm-specific parameters. Figure 5 offers its flowchart.

Assume i, j, k is the index of iteration, variable, and candidate. Then the update rule can be written as:

$$V(i+1, j, k) = V(i, j, k) + r(i, j, 1)(V(i, j, b) - |V(i, j, k)|) - r(i, j, 2)(V(i, j, w) - |V(i, j, k)|) \quad (6)$$

where $V(i, j, k)$ means the j -th variable of k -th candidate in i -th iteration. $V(i, j, b)$ and $V(i, j, w)$ represents the best and worst value of j -th variable in i -th iteration. The $r(i, j, 1)$ and $r(i, j, 2)$ are two positive numbers in the range of $[0, 1]$ generated at random.

3.4 The whole system

The block diagram of the whole system is illustrated in Fig. 6. We can see our method contains three important stages: the feature extraction via WPTE, the classification via ELM, and the training via Jaya. Finally, the output shows whether the input image is pathological or healthy.

Ten repetition of K -fold cross validation was used. Here the K is assigned with a value of 6 for FD, and with a value of 5 for SD and TD, due to the concept of stratification, which means each fold should contain the same distribution of the two classes. The cross validation setting is listed in Table 1. The pseudocode of our method is listed in Table 2.

3.5 SMOTE over training set

Here we should note the SMOTE technique is applied to the training set other than the whole dataset. For example, SMOTE shall increase 15 healthy brains to 40 healthy brains in FD, increase 16 healthy brains to 112 healthy brain in SD, and increase 28 healthy brains to 176 healthy brains in TD, respectively, as shown in Table 3.

In each cross validation trial, the SMOTE technique shall increase the sizes of the minority classes, i.e., the healthy brains.

Table 4 Definitions of TP, NP, TF, and NF

TP	Predict pathological brain as pathological brain
NP	Predict pathological brain as healthy brain
TF	Predict healthy brain as healthy brain
NF	Predict healthy brain as pathological brain

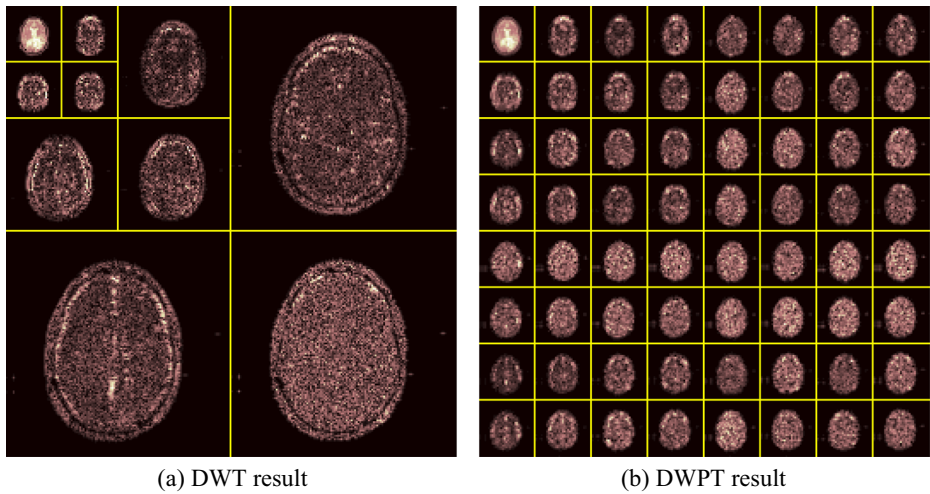


Fig. 7 DWT versus DWPT over a glioma brain image

3.6 Evaluation

Assume the abnormal brain is the positive class, the healthy brain is the negative class, then we can define true positive (TP), negative positive (NP), true false (TF), and negative false (NF) as in Table 4.

We measure the classification performance based on three indicators: sensitivity, specificity, and accuracy. They are defined as:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (8)$$

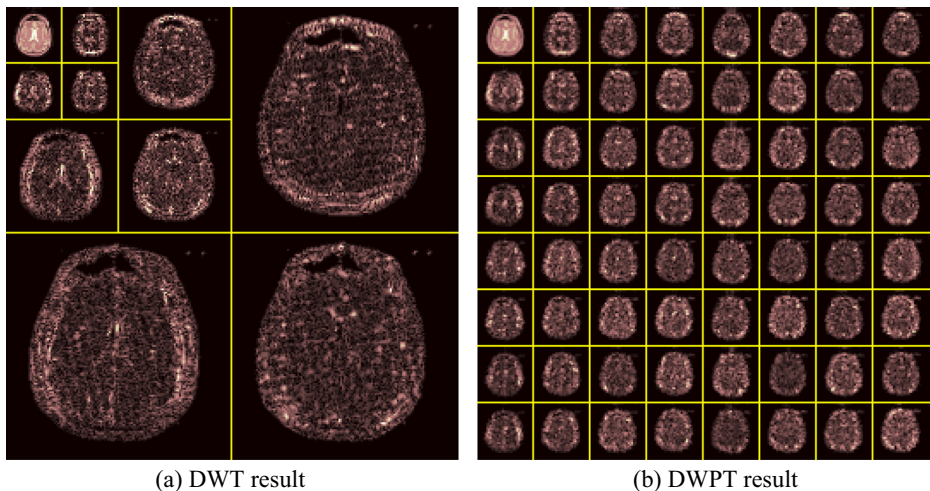


Fig. 8 DWT versus DWPT over a healthy brain image

Table 5 Correctly identified cases of our method over the third dataset

	F1	F2	F3	F4	F5	Total
R1	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(35 + 220)- > (255)
R2	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(35 + 220)- > (255)
R3	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 43)- > (50)	(7 + 44)- > (51)	(7 + 44)- > (51)	(35 + 219)- > (254)
R4	(7 + 44)- > (51)	(7 + 43)- > (50)	(7 + 43)- > (50)	(7 + 44)- > (51)	(7 + 44)- > (51)	(35 + 218)- > (253)
R5	(7 + 44)- > (51)	(6 + 44)- > (50)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(34 + 220)- > (254)
R6	(7 + 43)- > (50)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 42)- > (49)	(7 + 44)- > (51)	(35 + 217)- > (252)
R7	(6 + 43)- > (49)	(6 + 43)- > (49)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(33 + 218)- > (251)
R8	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(35 + 220)- > (255)
R9	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(35 + 220)- > (255)
R10	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(7 + 44)- > (51)	(35 + 220)- > (255)

F Fold, *R* Run

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

4 Results

4.1 DWT versus DWPT

The three-level DWT decomposition result of a glioma brain image is shown in Fig. 7a, while its DWPT decomposition result is shown in Fig. 7b. The same results of a healthy brain are shown in Fig. 8. We can observe that the DWPT extracts in total 64 subbands while DWT only provides 10 subbands. Generally, DWPT can give better subband information than DWT can, since DWPT decomposes not only approximation subband, but also detail subbands [31].

4.2 Statistical analysis

Due to page limit, we take third dataset (TD) as an example. The correct identified cases over each validation fold at each run are listed in Table 5. Note that SMOTE was only applied over

Table 6 Performance of our method over the third dataset

	Sensitivity	Specificity	Accuracy
R1	100.00	100.00	100.00
R2	100.00	100.00	100.00
R3	99.55	100.00	99.61
R4	99.09	100.00	99.22
R5	100.00	97.14	99.61
R6	98.64	100.00	98.82
R7	99.09	94.29	98.43
R8	100.00	100.00	100.00
R9	100.00	100.00	100.00
R10	100.00	100.00	100.00
Average	99.64 ± 0.52	99.14 ± 1.93	99.57 ± 0.57

Table 7 Performance of GA-ELM over the third dataset

	Sensitivity	Specificity	Accuracy
R1	97.73	97.14	97.65
R2	96.82	97.14	96.86
R3	98.64	97.14	98.43
R4	96.82	97.14	96.86
R5	97.73	97.14	97.65
R6	94.55	97.14	94.90
R7	96.82	97.14	96.86
R8	97.27	94.29	96.86
R9	96.82	97.14	96.86
R10	97.73	97.14	97.65
Average	97.09 \pm 1.08	96.86 \pm 0.90	97.06 \pm 0.93

the training set other than the validation set. Here $(a + b) \rightarrow c$ represents a healthy and b pathological brains, thus, in total c cases, were correctly identified. Revisit Table 1 that each fold contains seven healthy brains and 44 pathological brains.

The sensitivity, specificity, and accuracy of our method over 10 runs are listed in Table 6. Here the specificity is 99.14 ± 1.93 , the sensitivity is 99.64 ± 0.52 , and the accuracy is 99.57 ± 0.57 . All the values are in unit of percentage. For the other two datasets, our algorithm achieved perfect classification (i.e., 100%).

The performance of method may reach the upper limit. In the future, we shall try deep learning techniques, such as autoencoder, convolutional neural network, deep belief neural network [19], etc.

4.3 Training algorithm comparison

In this experiment, we compared proposed Jaya algorithm with genetic algorithm (GA) [23], particle swarm optimization (PSO) [29], and bat algorithm [7] in training the ELM. All the other settings are the same as previous experiments. The result of GA-ELM [23] is listed in Table 7, the result of PSO-ELM [29] is listed in Table 8, and the result of Bat-ELM [7] is listed in Table 9. For clew view, we picture the comparison results in Fig. 4. Obviously, the Jaya-ELM gives higher results than the GA-ELM [23], PSO-ELM [29], and Bat-ELM [7] in terms of all three measures: sensitivity, specificity, and accuracy (Fig. 9).

Table 8 Performance of PSO-ELM over the third dataset

	Sensitivity	Specificity	Accuracy
R1	97.73	100.00	98.04
R2	99.09	97.14	98.82
R3	99.09	97.14	98.82
R4	98.18	100.00	98.43
R5	98.64	94.29	98.04
R6	98.64	100.00	98.82
R7	98.18	100.00	98.43
R8	98.18	97.14	98.04
R9	99.09	97.14	98.82
R10	98.18	100.00	98.43
Average	98.50 \pm 0.48	98.29 \pm 2.00	98.47 \pm 0.34

Table 9 Performance of Bat-ELM over the third dataset

	Sensitivity	Specificity	Accuracy
R1	97.73	100.00	98.04
R2	97.73	100.00	98.04
R3	97.73	100.00	98.04
R4	97.27	97.14	97.25
R5	97.73	100.00	98.04
R6	98.18	94.29	97.65
R7	97.73	100.00	98.04
R8	98.18	100.00	98.43
R9	98.18	100.00	98.43
R10	96.82	100.00	97.25
Average	97.73 ± 0.43	99.14 ± 1.93	97.92 ± 0.42

4.4 Effect of SMOTE

If we did not use SMOTE, the performance result over the third dataset was presented in Table 10. Here the sensitivity is $99.50 \pm 0.45\%$, higher than the specificity of $95.43 \pm 2.76\%$. The reason is the positive class (pathological brains) is the majority; hence, the classifier will tend to learn from the samples of the positive class. Besides, the overall accuracy is $98.94 \pm 0.42\%$, less than the method with SMOTE.

4.5 Comparison to state-of-the-art approaches

We compared our method with six state-of-the-art approaches. Those methods are:

- 1) fractional Fourier entropy (FRFE) with multilayer perceptron (MLP) [25],
- 2) dual-tree complex wavelet transform (DTCWT) with support vector machine (SVM) [32],
- 3) Minkowski-Bouligand dimension (MBD) with improved particle swarm optimization (IPSO) [2],

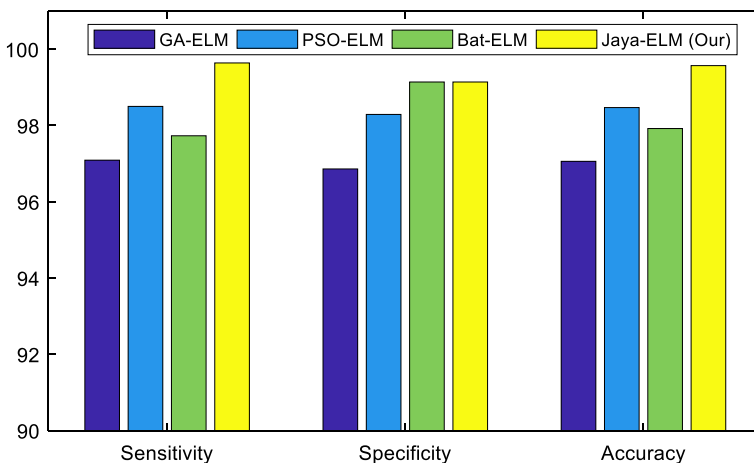
**Fig. 9** Training algorithm comparison

Table 10 Performance without SMOTE over the third dataset

	Sensitivity	Specificity	Accuracy
R1	99.55	97.14	99.22
R2	99.09	94.29	98.43
R3	98.64	97.14	98.43
R4	100.00	97.14	99.61
R5	99.55	97.14	99.22
R6	100.00	94.29	99.22
R7	99.09	97.14	98.82
R8	100.00	88.57	98.43
R9	99.55	97.14	99.22
R10	99.55	94.29	98.82
Average	99.50 ± 0.45	95.43 ± 2.76	98.94 ± 0.42

- 4) wavelet packet Tsallis entropy (WPTE) with feedforward neural network (FNN) trained by improved biogeography-based optimization (IBBO) [4],
- 5) pseudo Zernike moment (PZM) with support vector machine (SVM) [12],
- 6) Hu moment invariant (HMI) with predator-prey particle swarm optimization (PP-PSO) [27].

The results are offered in Table 11 and Fig. 10. We can observe that our proposed method achieved the highest accuracy over three datasets. Besides, our proposed gives better results than “WPTE + FNN + IBBO [16]”, which demonstrates that the ELM may perform better than feedforward neural network. Although we only increase the accuracy by less than 1%, but the result is obtained by strict cross validation. Hence, our algorithm indeed gives better performance than state-of-the-art approaches.

The preprocessing of brain segmentation [9–11] may give better results than our method. In the future, we shall test the effect of brain segmentation. The dual-tree complex wavelet transform is another excellent variant of standard wavelet transform (DTCWT). We shall compare DTCWT with our wavelet packet transform in the future. Further, there are several improved algorithms of Jaya, and we shall test their performance.

4.6 Optimal decomposition level

We used three-level decomposition in Section 4.1. In this section, we shall give an experiment why the three-level is optimal. We carried our one-level, two-level, three-level, four-level, five-

Table 11 Comparison to six state-of-the-art approaches

Existing approaches	Accuracy		
	FD	SD	TD
FRFE + MLP [25]	99.09	97.81	95.76
DTCWT + SVM [32]	100.00	99.69	98.43
MBD + IPSO [2]	100.00	98.19	98.08
WPTE + FNN + IBBO [4]	100.00	100.00	99.29
PZM + SVM [12]	99.09 ± 1.46	98.13 ± 0.83	98.75 ± 0.36
HMI + PP-PSO [27]	N/A	98.25	97.02
SMOTE + WPTE + ELM + Jaya (Proposed)	100.00	100.00	99.57 ± 0.57

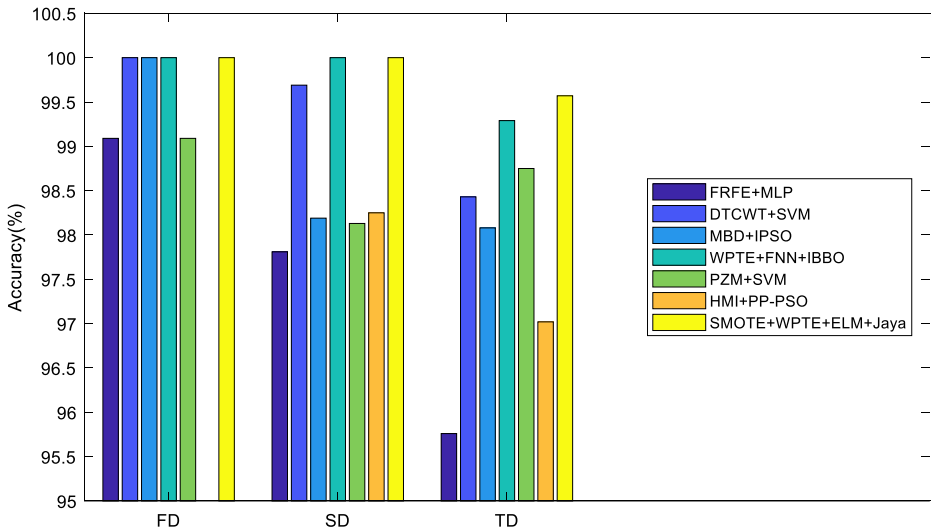


Fig. 10 Algorithm comparison

level, and six-level decomposition, respectively. The overall accuracy for the third dataset (TD) is shown in Fig. 11. We can observe the 3-level decomposition arrives at the highest accuracy.

5 Conclusions

In this study, our team proposed a novel smart pathological brain detection system based on wavelet packet Tsallis entropy, extreme learning machine, and Jaya algorithm. The comparison results showed the superiority of our method to six state-of-the-art pathological brain detection

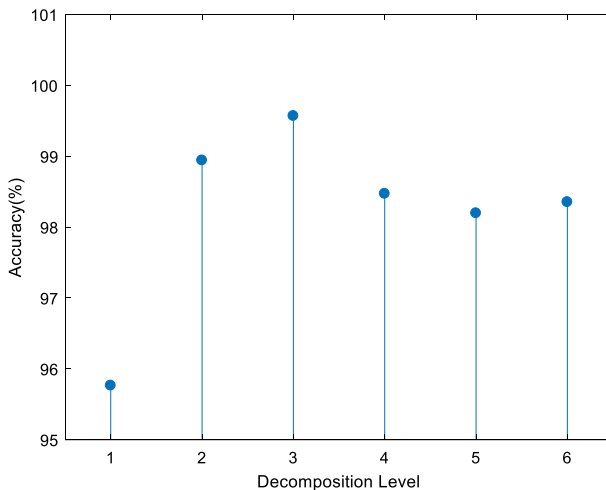


Fig. 11 Optimal decomposition level over third dataset

approaches. Besides, Jaya algorithm, as a training method, is superior to genetic algorithm, particle swarm optimization, and bat algorithm.

Acknowledgements The paper is supported by Natural Science Foundation of China (61602250), Natural Science Foundation of Jiangsu Province (BK20150983), Open fund of Key Laboratory of Guangxi High Schools Complex System and Computational Intelligence (2016CSCI01), Open fund for Jiangsu Key Laboratory of Advanced Manufacturing Technology (HGAMTL1601).

Author's contribution Y Zhang & G Zhao conceived the study. J Sun & T Zhan designed the model. Y Zhang & J Li acquired the data. Y Zhang, T Zhan, J Li analyzed the data. G Zhao interpreted the data. Y Zhang, X Wu, Z Wang, H Liu, V Govindaraj developed the programs. Y Zhang & T Zhan wrote the draft. All the authors gave critical revisions and approved the submission.

Compliance with ethical standards

Conflict of interest We declared there is no conflict of interest in terms with this submission.

References

1. Chen Y (2016) Voxelwise detection of cerebral microbleed in CADASIL patients by leaky rectified linear unit and early stopping: A class-imbalanced susceptibility-weighted imaging data study. *Multimed Tools Appl.* doi:10.1007/s11042-017-4383-9 (Online)
2. Chen X-Q (2016) Fractal dimension estimation for developing pathological brain detection system based on Minkowski-Bouligand method. *IEEE Access* 4:5937–5947
3. Chen H (2017) Seven-layer deep neural network based on sparse autoencoder for voxelwise detection of cerebral microbleed. *Multimed Tools Appl.* doi:10.1007/s11042-017-4554-8 (Online)
4. Chen P, Du S (2017) Pathological Brain Detection via Wavelet Packet Tsallis Entropy and Real-Coded Biogeography-based Optimization. *Fundamenta Informaticae* 151(1–4):275–291
5. Chen K, Lv Q, Lu Y et al (2017) Robust regularized extreme learning machine for regression using iteratively reweighted least squares. *Neurocomputing* 230:345–358
6. Das S, Chowdhury M, Kundu MK (2013) Brain MR image classification using multiscale geometric analysis of Ripplet. *Progress in Electromagnetics Research-Pier* 137:1–17
7. Doreswamy, Salma MU (2015) BAT-ELM: a bio inspired model for prediction of breast cancer data. In: *International Conference on Applied and Theoretical Computing And Communication Technology (Icatcct)*. Davangere, IEEE, pp. 501–506
8. Hazlett HC, Gu HB, Munsell BC et al (2017) Early brain development in infants at high risk for autism spectrum disorder. *Nature* 542(7641):348
9. Huo Y, Plassard AJ, Carass A et al (2016) Consistent cortical reconstruction and multi-atlas brain segmentation. *NeuroImage* 138:197–210
10. Huo Y, Carass A, Resnick SM et al (2016) Combining Multi-atlas Segmentation with Brain Surface Estimation. In: *Conference on Medical Imaging - Image Processing*. San Diego, Spie-Int Soc Optical Engineering, p 97840E
11. Huo YK, Asman AJ, Plassard AJ et al (2017) Simultaneous total intracranial volume and posterior fossa volume estimation using multi-atlas label fusion. *Hum Brain Mapp* 38(2):599–616
12. Jiang Y, Zhu W (2017) Exploring a smart pathological brain detection method on pseudo Zernike moment. *Multimed Tools Appl.* doi:10.1007/s11042-017-4703-0 (Online)
13. Li H, Bhowmick SS, Sun AX et al (2015) Conformity-aware influence maximization in online social networks. *VLDB J* 24(1):117–141
14. Liu G (2015) Pathological brain detection in MRI scanning by wavelet packet Tsallis entropy and fuzzy support vector machine. *SpringerPlus* 4(1):716
15. Lu Z (2016) A Pathological Brain Detection System Based on Radial Basis Function Neural Network. *Journal of Medical Imaging and Health Informatics* 6(5):1218–1222

16. Mavaddaty S, Ahadi SM, Seyedin S (2017) Speech enhancement using sparse dictionary learning in wavelet packet transform domain. *Computer Speech And Language* 44:22–47
17. Mustafa N, Memon RA, Li JP et al (2017) A Classification Model for Imbalanced Medical Data based on PCA and Farther Distance based Synthetic Minority Oversampling Technique. *Int J Adv Comput Sci Appl* 8(1):61–67
18. Nayak DR (2017) Detection of unilateral hearing loss by Stationary Wavelet Entropy. *CNS Neurol Disord Drug Targets* 16(2):122–128
19. Oyedotun O, Khashman A (2017) Iris nevus diagnosis: convolutional neural network and deep belief network. *Turk J Electr Eng Comput Sci* 25(2):1106–1115
20. Phillips P, Dong Z, Ji G et al (2015) Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC. *Biomed Signal Process Control* 21:58–73
21. Phillips P, Dong Z, Yang J (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
22. Rao RV, Savsani VJ, Vakharia DP (2011) Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems. *Comput Aided Des* 43(3):303–315
23. Rong YM, Zhang GJ, Chang Y et al (2016) Integrated optimization model of laser brazing by extreme learning machine and genetic algorithm. *Int J Adv Manuf Technol* 87(9–12):2943–2950
24. Sun P (2015) Pathological brain detection based on wavelet entropy and Hu moment invariants. *Biomed Mater Eng* 26(s1):1283–1290
25. Sun Y (2016) A Multilayer Perceptron Based Smart Pathological Brain Detection System by Fractional Fourier Entropy. *J Med Syst* 40(7):173
26. Tsallis C (2009) Nonadditive entropy: The concept and its use. *Eur Phys J A* 40(3):257–266
27. Wang H, Lv Y (2016) Smart pathological brain detection system by predator-prey particle swarm optimization and single-hidden layer neural-network. *Multimed Tools Appl*. doi:[10.1007/s11042-016-4242-0](https://doi.org/10.1007/s11042-016-4242-0) (Online)
28. Wang M, Li H, Cui JT et al (2016) PINOCCHIO: Probabilistic Influence-Based Location Selection over Moving Objects. *IEEE Trans Knowl Data Eng* 28(11):3068–3082
29. Yadav B, Ch S, Mathur S et al (2016) Estimation of in-situ bioremediation system cost using a hybrid Extreme Learning Machine (ELM)-particle swarm optimization approach. *J Hydrol* 543: 373–385
30. Yang J (2015) Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with Tsallis entropy and generalized eigenvalue proximal support vector machine (GEPSVM). *Entropy* 17(4):1795–1813
31. Yang J (2015) Identification of green, Oolong and black teas in China via wavelet packet entropy and fuzzy support vector machine. *Entropy* 17(10):6663–6682
32. Yang M (2016) Dual-Tree Complex Wavelet Transform and Twin Support Vector Machine for Pathological Brain Detection. *Appl Sci* 6(6):169
33. Ying ZB, Li H, Ma JF et al (2016) Adaptively secure ciphertext-policy attribute-based encryption with dynamic policy updating. *Science China-Information Sciences* 59(4):16, 042701
34. 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
35. Yuksel T (2017) Intelligent visual servoing with extreme learning machine and fuzzy logic. *Expert Syst Appl* 72:344–356
36. Zhan T (2016) Pathological brain detection by artificial intelligence in magnetic resonance imaging scanning. *Prog Electromagn Res* 156:105–133
37. 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
38. Zhou XX, Zhang GS (2016) Detection of abnormal MR brains based on wavelet entropy and feature selection. *IEEE Trans Electr Electron Eng* 11(3):364–373



Yu-Dong Zhang received his Ph.D. degree from Southeast University at 2010. He worked as postdoc from 2010 to 2012, and a research scientist from 2012 to 2013 at Columbia University. He is now a full professor in Nanjing Normal University, China. He is a guest professor in Henan Polytechnic University. He is included in “Most Cited Chinese researchers (Computer Science)” from 2015 to 2017.



Guihu Zhao received B.S. and M.S. degrees in computer science from Central South University, Changsha, Hunan, China, in 2007 and 2010, respectively. He is a Ph.D. Candidate at the Central South University. His major interests are fractals, complexity theory, brain MRI image processing, brain network, multimedia applications, mobile networks, and machine learning.



Junding Sun received B.S. degree in computer application and M.S. degree in control theory and control engineering from Henan Polytechnic University, Jiaozuo, Henan, China, in 1998 and 2001, respectively. He received his Ph.D. degree in computer application from Xidian University in 2005. His major interests are image processing, image retrieval, pattern recognition, and multimedia application.



Xiaosheng Wu is an associate professor in Henan Polytechnic University. She received her B.S. degree and M.S. in computer application technology from Henan Polytechnic University, Jiaozuo, Henan, China, in 1998 and 2010, respectively. Her research interests include image processing, image retrieval and pattern recognition.



Zhi-Heng Wang received his B.Sc degree in mechatronic engineering from Beijing Institute of Technology, China in 2004, and his Ph.D. from the Institute of Automation, Chinese Academy of Sciences, China in 2009.

Currently, he is an associate professor at School of Computer Science and Technique, Henan Polytechnic University, China. His research interests include computer vision, pattern recognition, and image processing.



Hong-Min Liu received her B.Sc degree in electrical & information engineering from Xi'dian University, Xi'an, China in 2004, and her Ph.D. from the Institute of Electronics, Chinese Academy of Sciences, China in 2009. Currently, she is an associate professor at School of Computer Science and Technique, Henan Polytechnic University, China. Her research interests include image processing, especially on feature detection and matching.



Vishnu Varthanan Govindaraj was born in Madurai district, Tamilnadu, India in the year 1986. He has completed his Bachelor of Engineering in Instrumentation and Control Engineering with first class in the year 2007 from Arulmigu Kalasalingam College of Engineering, Virudhunagar, Tamilnadu and Master of Technology with distinction in the year 2009 from Bharath University, Chennai, Tamilnadu. He has worked as a Lecturer in Electronics and Communication Engineering Department, Andal Alagar College of Engineering from May 2009 to June 2010. After gaining a year experience, he then joined as an Assistant Professor in Electrical and Electronics Engineering Department of Vels University, Pallavaram and he served there from June 2010 to May 2011. Later, he continued his academic service as an Assistant Professor in Electronics and Instrumentation Department of Kalasalingam University, since July 2011. He holds a PhD Degree awarded from the Department of Electronics and Communication Engineering and at present he is an Associate Professor in the Department of Instrumentation and Control Engineering of

Kalasalingam University, Tamilnadu, India. His areas of interest are Medical image processing and signal processing, and he has significant publications pertaining to these domains.



Tianmin Zhan received the Ph.D. degree in pattern recognition and intelligence system from Nanjing University of Science and Technology, Nanjing, China, in 2013. He is currently an associate professor in the school of technology, Nanjing Audit University. His research interests include in medical imaging processing and analysis.



Jianwu Li received the B.S., M.Eng. and Ph.D. degrees from Tianjin University, China, in 1997, 2000 and 2003, respectively. He is currently an Associate Professor with the School of Computer Science and Technology, Beijing Institute of Technology, China. He has authored more than 20 scientific papers in peer-reviewed journals and conferences. His research interests include image processing, pattern recognition and evolutionary computing.