

METHODOLOGY

Open Access



Pathological brain detection in MRI scanning by wavelet packet Tsallis entropy and fuzzy support vector machine

Yu-Dong Zhang^{1,2*†}, Shui-Hua Wang^{1,2†}, Xiao-Jun Yang^{3†}, Zheng-Chao Dong⁴, Ge Liu⁵, Preetha Phillips⁶ and Ti-Fei Yuan^{7*}

*Correspondence:
zhangyudong@njnu.edu.cn;
ytf0707@126.com

[†]Yu-Dong Zhang, Shui-Hua Wang and Xiao-Jun Yang contributed equally

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

⁷ School of Psychology, Nanjing Normal University, Nanjing, Jiangsu 210008, China

Full list of author information is available at the end of the article

Abstract

An computer-aided diagnosis system of pathological brain detection (PBD) is important for help physicians interpret and analyze medical images. We proposed a novel automatic PBD to distinguish pathological brains from healthy brains in magnetic resonance imaging scanning in this paper. The proposed method simplified the PBD problem to a binary classification task. We extracted the wavelet packet Tsallis entropy (WPTE) from each brain image. The WPTE is the Tsallis entropy of the coefficients of the discrete wavelet packet transform. The, the features were submitted to the fuzzy support vector machine (FSVM). We tested the proposed diagnosis method on 3 benchmark datasets with different sizes. A ten runs of K-fold stratified cross validation was carried out. The results demonstrated that the proposed WPTE + FSVM method excelled 17 state-of-the-art methods w.r.t. classification accuracy. The WPTE is superior to discrete wavelet transform. The Tsallis entropy performs better than Shannon entropy. The FSVM excels standard SVM. In closing, the proposed method "WPTE + FSVM" is effective in PBD.

Keywords: Pathological brain detection (PBD), Tsallis entropy, Magnetic resonance imaging, Computer-aided diagnosis, Discrete wavelet packet transform, Fuzzy support vector machine, Pattern recognition

Background

Pathological brain detection (PBD) was of essential importance. It can help physicians make decisions, and to avoid wrong judgements on subjects. Magnetic resonance imaging (MRI) features in high-resolution on soft tissues in the subjects' brains, generating a mass dataset (Zhang et al. 2015a). At present, there are numerous works on using brain MR images for solving PBD problems (Goh et al. 2014; Yu et al. 2015b).

Recent computer-aided diagnosis (CAD) systems of PBD consisted of two types (LaViolette et al. 2014): to detect pathological from healthy brains, and to differentiate severity degrees. In this study, we research on the former one. A type of promising approach is to use discrete wavelet transform (DWT) that presents the solutions of simultaneous analysis in domains of both time and frequency (Lee et al. 2013; Dong et al. 2014; Zhang et al. 2015c; Yu et al. 2015c). DWT and its variants achieved good results; however, DWT are

translation-variant, hence, the coefficients behaved unpredictably if the input signal is translated slightly. In PBD problem, the subject's head usually have slightly move during the scan, which will cause the translation of MR images.

Another problems is the classifier. Current scholars tend to use either artificial neural network (ANN) or support vector machine (SVM). Nevertheless, both of them are sensitive to outliers and noises. That means, if the training set contains noises or outliers, the classifier will still treat it as important as normal data.

We suggested three improvements with the aim of solving above problems. First, we employed the discrete version of wavelet packet transform (WPT), which is an extension of standard discrete wavelet transform (DWT). Second, we introduced Tsallis entropy (TE), to replace with Shannon entropy (SE). (iii) We introduced the fuzzy support vector machine (FSVM) that combines the SVM with fuzzy logic approach (Ashkezari et al. 2013) and has the advantage of reducing the effect from outliers and noises.

The structure of the rest is organized as follows. "State-of-the-art" presents the state-of-the-art. "Materials" introduces the materials used in this study. "Feature extraction" discusses the features. "Classifier" gives the classifier. "Implementation and experiments" shows the implementation of the whole method, and designs the experiments. "Results and dicussion" contains the results and discussions. "Conclusion and future research" offers conclusion and future research. We explain the nomenclatures in Abbreviations at the end of the paper.

State-of-the-art

Chaplot et al. (2006) was the first to solve PBD problem. They used the approximation coefficients from DWT, and utilized the support vector machine (SVM) and self-organizing map (SOM). El-Dahshan et al. (2010) extracted all coefficients of all subbands of a three-level discrete wavelet transform (DWT). Then, they reduced the size of features by principal component analysis (PCA). Finally, two classifiers, K-nearest neighbors (KNN) and feed-forward back-propagation ANN (FP-ANN), were employed. Wu and Wang (2011) followed El-Dahshan's method, but suggest to use a feed-forward neural network (FNN) as the classifier, which was trained by scaled chaotic artificial bee colony (SCABC). Dong et al. (2011) proposed to employed scaled conjugate gradient (SCG) method to take place of SCABC. Zhang and Wu (2012) suggested to utilize kernel support vector machine (KSVM). 3 kernels were provided such as homogeneous and inhomogeneous polynomial, and radial basis function (RBF). Das et al. (2013) developed a novel method as Ripplet transform (RT) + principal component analysis (PCA) + least square support vector machine (LS-SVM). Their five-fold cross validation results showed promising classification accuracies. Saritha et al. (2013) proposed a novel feature of wavelet-entropy (WE), and employed spider-web plots (SWP) to further reduce features. Afterwards, they used the probabilistic neural network (PNN). Yu et al. (2015d) commented on Saritha's paper and stated that dropping the SWP can obtain the same results. Zhang et al. (2013) suggested to use particle swarm optimization to train KSVM. Padma and Sukanesh (2014) used combined wavelet statistical texture features, to segment and classify AD benign and malignant tumor slices. El-Dahshan et al. (2014) used the feedback pulse-coupled neural network for image segmentation, the DWT for features extraction, the PCA for reducing the dimensionality of the wavelet coefficients,

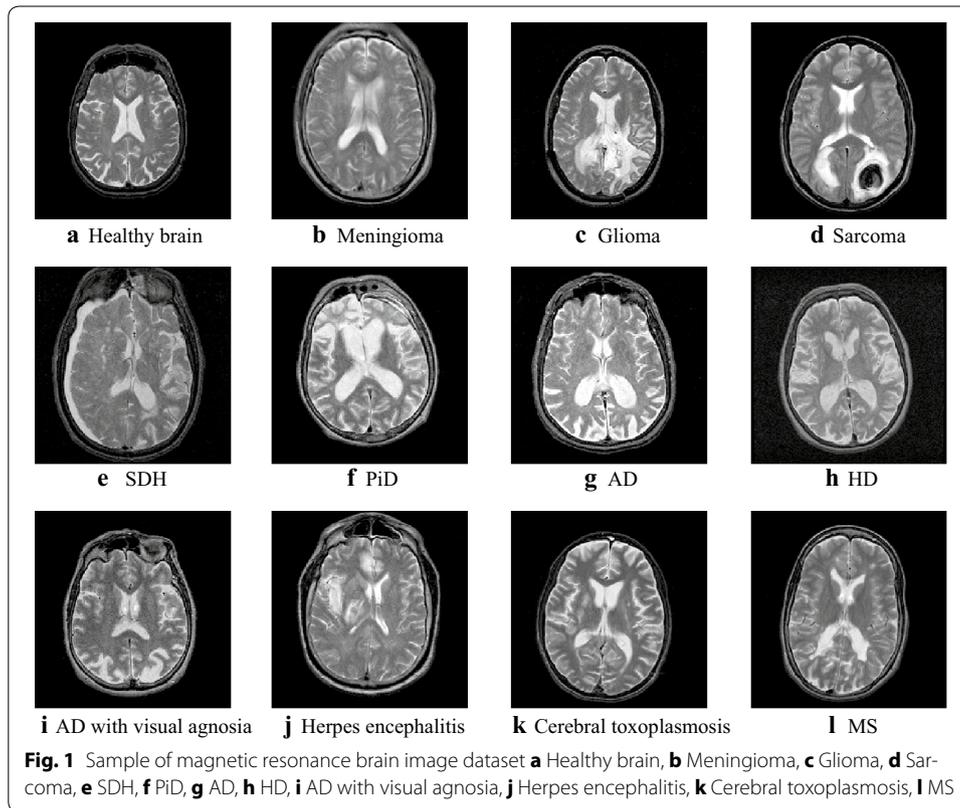
and the FBPNN to classify inputs into normal or abnormal. Wang et al. (2014) used kernel support vector machine decision tree. Zhou et al. (2015) used wavelet-entropy as the feature space, then they employed a Naive Bayes classifier (NBC) classification method. Their results over 64 images showed that the sensitivity of the classifier was 94.50 %, the specificity 91.70 %, the overall accuracy 92.60 %. Damodharan and Raghavan (2015) combined tissue segmentation and neural network for brain tumor detection. Yang et al. (2015) selected wavelet-energy as the features, and introduced biogeography-based optimization (BBO) to train the SVM. Their method reached 97.78 % accuracy on 90 T2-weighted MR brain images. Nazir et al. (2015) suggested to use filters for the removal of noises, and extracted color moments as mean features. Finally, they achieved an overall accuracy of 91.8 %. Dong et al. (2015) suggested to use a 3D eigenbrain method to detect subjects and brain regions related to AD. The accuracy achieved 92.36 ± 0.94 . Harikumar and Kumar (2015) analyzed the performance of ANN, in terms of classification of medical images, using wavelets as feature extractor. Their classification accuracy achieved 96 %. Wang et al. (2015a) suggested to use stationary wavelet transform (SWT) to replace DWT, and then they proposed a Hybridization of Particle swarm optimization and Artificial bee colony (HPA) algorithm to train the classifier. Farzan et al. (2015) used longitudinal percentage of brain volume changes (PBVC) in two-year follow up and its intermediate counterparts in early 6-month and late 18-month as features. Their experiment results obtained accuracy of 91.7 %. Munteanu et al. (2015) employed Proton Magnetic Resonance Spectroscopy (MRS) data, with the aim of detecting MCI and AD. They used a single-layer perceptron with only two spectroscopic voxel volumes obtained in the left hippocampus, with an AUROC value of 0.866. Zhang et al. (2015d) combined wavelet entropy with Hu moment invariants (HMI). The feature number is in total 14. They also used GEPSVM as the classifier.

Materials

Magnetic resonance brain image dataset

Three benchmark magnetic resonance brain image datasets with various image numbers: D-66, D-160, and D-255, were downloaded from the website of Harvard University. Those data contain T2-weighted images obtained along axial plane. Their sizes are all 256×256 . Those three datasets are commonly used in PBD test. Except healthy brain images, D-66 and D-160 consisted of 7 types of brain diseases: AD, AD plus visual agnosia, glioma, meningioma, sarcoma, Huntington's disease (HD), and Pick's disease (PiD). D-255 introduced four other diseases as cerebral toxoplasmosis, subdural hematoma (SDH), multiple sclerosis (MS), and herpes encephalitis. Figure 1 shows samples of brain MR images.

The costs of two kinds of misclassifications are different. The cost of predicting a pathological brain to a healthy one is very serious. It will defer the necessary treatment, whereas the misprediction of a healthy brain to a pathological one can be second-checked by other techniques. Hence, we intentionally create the three imbalanced datasets, which covers more pathological brains than usual, so the PBD system is biased to detect pathological ones, with the aim of addressing this cost-sensitive task.



Statistical setting

Cross validation (CV) is commonly used for statistical test. Stratification is embedded to CV so that each fold contains nearly the same class distributions. In this work, six-fold stratified CV (SCV) was utilized for the smallest dataset (D-66), and five-fold SCV for the other datasets (D-160 and D-255). Table 1 lists the SCV setting of all datasets.

Feature extraction

Co-registration was unnecessary since many publications about PBD did not use it with excellent classification results, comparative with the results that employed coregistration (Ribbens et al. 2014; Schwarz and Kasperek 2014).

Table 1 SCV setting of our datasets

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

P Pathological, H Healthy

Wavelet packet transform

Compared to standard discrete wavelet transform (DWT), the wavelet packet transform (WPT) is an extension where the signal is passed through more filters than DWT. The DWT calculate each level by passing only the previous approximation coefficients to quadrature mirror filters (QMF). Nevertheless, the WPT passes all coefficients (both approximation and detail) through QMF to create a full binary tree. Therefore, more features can be generated by WPT at different levels to obtain more information. The mathematical equation of WPT is given below

$$S_p^{m,d} = \int_{-\infty}^{\infty} x(t)\psi_m(2^{-d}t - p)dt \tag{1}$$

where m represents the index of channel, p the position parameter, d the decomposition level, ψ the wavelet function, and S the decomposition coefficients. 2^d sequences will be yielded at the d level. The decomposition equations of next level is provided as

$$S_k^{2m,d+1} = \sum_{p \in Z} h(p - 2k)S_p^{m,d} \tag{2}$$

$$S_k^{2m+1,d+1} = \sum_{p \in Z} l(p - 2k)S_p^{m,d} \tag{3}$$

Suppose a d -level decomposition, DWT produces $(3d + 1)$ coefficient sets, while the WPT produces 2^d different coefficients sets. Note that the number of coefficients of WPT is still the same of DWT, because of the downsampling process (Fig. 2).

Shannon and Tsallis entropy

Shannon entropy (SE) is defined as a measure of uncertainty regarding the information content (IC):

$$E = - \sum_{k=1}^Z p_k \log_2(p_k) \tag{4}$$

here E represents the entropy, Z the total number of greylevels, k the greylevel, and p_k the probability of k . Shannon entropy can merely describes scenarios with simple effective microscopic interactions and short-ranged microscopic memory (Campos 2010). Assume a physical system can be broken down into two independent subsystems X and Y , then the Shannon entropy (SE) exists the additivity property as

$$E(X + Y) = E(X) + E(Y) \tag{5}$$

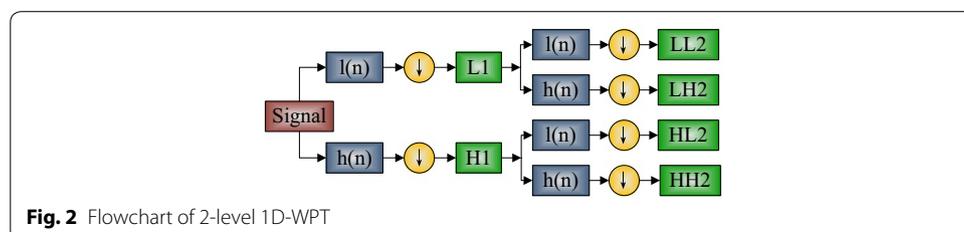


Fig. 2 Flowchart of 2-level 1D-WPT

Nevertheless, realistic scenarios are usually usually involved with long-time memory and long-range interactions, therefore, Tsallis (2009) proposed a generalization of SE. He termed it as Tsallis entropy (TE) with following form

$$E_q = \frac{\sum_{k=1}^q (p_k)^q - 1}{1 - q} \tag{6}$$

here q is a real number, representing the nonextensivity degree. For a statistical dependent system, the Tsallis entropy (TE) is defined as (Zhang and Wu 2011)

$$E_q(X + Y) = E_q(X) + E_q(Y) + (1 - q) \times E_q(X) \times E_q(Y) \tag{7}$$

This equation obeys the pseudo additivity rule. Further, three different entropies can be deduced and listed in Table 2, when q is assigned with different values (Tsallis 2011). In this study, TE was employed to extract features from 16 subbands of WPT coefficients of MR brain images.

Wavelet packet Tsallis entropy

We employed both Shannon entropy (SE) and Tsallis entropy (TE) to extract wavelet-packet decomposition coefficients. The final extracted features were dubbed as Wavelet Packet Tsallis Entropy (WPTE), which degraded to Wavelet Packet Shannon Entropy (WPSE) when q equals to 1. The pseudocodes of feature extraction were listed in Table 3.

Classifier

Support vector machine

Let us suppose there is an N -size training samples of p -dimensional vector in two classes (-1 or $+1$), and the goal is to create a $(p - 1)$ -dimensional hyperplane. Assume the dataset takes the form of (Wang et al. 2014)

$$\{(x_n, y_n) | x_n \in \mathbb{R}^p, y_n \in \{+1, -1\}\}, n = 1, 2, 3, \dots, N \tag{8}$$

Table 2 Properties of TE change with q

Range of q	Type	Property
<1	Subextensive entropy	$E_q(A + B) < E_q(A) + E_q(B)$
$=1$	Standard extensive entropy (Shannon entropy)	$E_q(A + B) = E_q(A) + E_q(B)$
>1	Superextensive entropy	$E_q(A + B) > E_q(A) + E_q(B)$

Table 3 Pseudocode of WPTE

Algorithm: WPTE Extraction	
Step A	Import a brain image
Step B	Implement a two-level WPT decomposition
Step C	Extract the Tsallis entropy over each coefficient set
Step D	Output the 16-element WPTE vector

where y_n takes the value of -1 for class -1 , or $+1$ for class $+1$. The x_n denotes a training point that is a p -dimensional vector (Zhang et al. 2013). The maximum-margin hyperplane that separates the two classes is the desired SVM. Considering any hyperplane is in the form of $\mathbf{w}x - \mathbf{b} = 0$, we need to select the optimal \mathbf{b} and \mathbf{w} , with the aim of maximizing the distance between the two parallel hyperplanes, while it can yet separate the data of the two classes.

$$\begin{aligned} \min_{\mathbf{b}, \mathbf{w}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & y_n(\mathbf{w}x_n - \mathbf{b}) \geq 1, \quad n = 1, 2, 3, \dots, N \end{aligned} \tag{9}$$

Positive slack vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n, \dots, \xi_N)$ are utilized to measure the misclassification rate of sample x_n (the distance between the margin and the vectors x_n on the wrong side). The optimal hyperplane can be deduced by solving:

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

where C represents the error penalty and e a vector of ones of N -dimension. Therefore, the optimization turns to a trade-off between a large margin and a small error penalty. The constraint optimization problem can be solved using “Lagrange multiplier” as

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

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

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

The key advantage of the dual form function is that the slack variables ξ_n vanish from the dual problem, with the constant C appearing only as an additional constraint on the Lagrange multipliers.

Fuzzy SVM

Fuzzy SVM (FSVM) is more effective than standard SVM in predict or classify real-world data, in which a part of training points are less important than other points. We would like to force that the meaningful training points must be classified correctly and meaningless points like noises or outliers can be treated with less weight (Lin and Wang 2002).

FSVM applies a fuzzy membership function (FMF) s to each training data (Xian 2010), so that the training set is transformed into a fuzzy set, which can be expressed as

$$\{(x_n, s_n, y_n) | x_n \in \mathbb{R}^p, 0 < s_n \leq 1, y_n \in \{+1, -1\}, n = 1, \dots, N\} \tag{13}$$

where s_n is the altitude of the corresponding training point toward one class and $(1 - s_n)$ is the attitude of meaning less. The optimal hyperplane problem of FSVM is defined as:

$$\begin{aligned} \min_{\mathbf{w}, \xi, \mathbf{b}} & \frac{1}{2} \|\mathbf{w}\|^2 + C \mathbf{s}^T \xi \\ \text{s.t.} & \begin{cases} y_n (\mathbf{w}^T x_n - \mathbf{b}) \geq 1 - \xi_n, n = 1, \dots, N \\ \xi_n \geq 0 \end{cases} \end{aligned} \tag{14}$$

where $\mathbf{s} = (s_1, s_2, \dots, s_N)$ represents the fuzzy membership vector. A smaller s_n reduces the effect of the parameter ξ_n , such that the corresponding point x_n is treated less important. In a similar way, we construct the Lagrangian

$$\min_{\mathbf{w}, \xi, \mathbf{b}} \max_{\alpha, \beta} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \mathbf{s}^T \xi - \sum_{n=1}^N \beta_n \xi_n - \sum_{n=1}^N \alpha_n [y_n (\mathbf{w}^T x_n - \mathbf{b}) - 1 + \xi_n] \right\} \tag{15}$$

Again, the dual form is used to transform problem (15) to

$$\begin{aligned} \max_{\alpha} & \sum_{n=1}^N \alpha_n - \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_m \alpha_n y_m y_n x_m^T x_n \\ \text{s.t.} & \begin{cases} 0 \leq \alpha_n \leq s_n C \\ \sum_{n=1}^N \alpha_n y_n = 0, n = 1, \dots, N \end{cases} \end{aligned} \tag{16}$$

Fuzzy membership

We set the FMF as a distance function between the point and its class center. Suppose the mean of class +1 as x_+ and the mean of class -1 as x_- . Then we can get the radius of two classes as

$$r_- = \max_{\{x_n: y_n = -1\}} |x_- - x_n| \tag{17}$$

$$r_+ = \max_{\{x_n: y_n = +1\}} |x_+ - x_n| \tag{18}$$

The fuzzy membership s_n is defined as a function of the radius and mean of each class (Lin and Wang 2002)

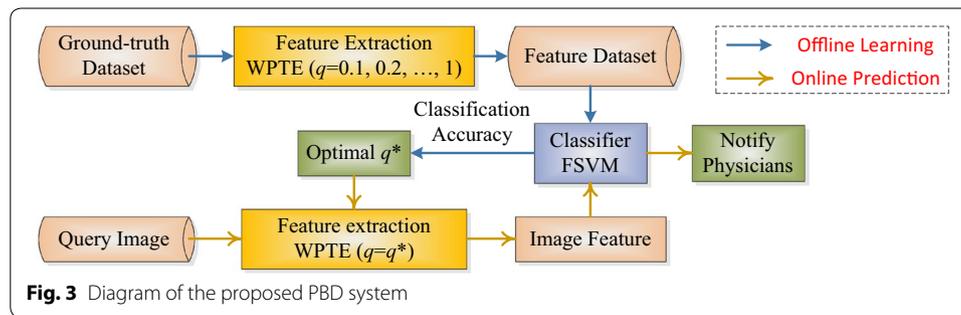
$$s_n = \begin{cases} 1 - |x_+ - x_n| / (r_+ + \delta) & y_n = +1 \\ 1 - |x_- - x_n| / (r_- + \delta) & y_n = -1 \end{cases} \tag{19}$$

where $\delta > 0$ is used to guarantee $s_n > 0$.

Implementation and experiments

Implementation

Figure 3 shows the diagram of the proposed PBD system. In the offline learning phase, the users expect to select the optimal q (to determine the value of q^*), and train the



classifier. In the online prediction phase, the users will get the prediction results for each query image.

Experiment design

In this study, we developed four different methods. “WPSE + SVM”, “WPSE + FSVM”, “WPTE + SVM”, and “WPTE + FSVM”. Theoretically, the last one will perform the best since WPSE in a special case of WPTE, and FSVM is an extension of SVM with additional ability to reduce influences from noises and outliers.

We need to prove it by experiments. In this work, we designed five tasks. (1) We gave a comparison between DWT and WPT. A healthy brain and a pathological brain were used. We use a 2-level Haar wavelet decomposition. (2) We compared the proposed WPSE and WPTE features with traditional DWT and “DWT + PCA”. All used SVM as classifiers (3) We compared the four proposed classifiers, to check whether FSVM is superior to SVM. (4) We selected the best of proposed methods, and compared it with state-of-the-art approaches. (5) We used grid searching to find the optimal parameter of q .

Results and discussions

The experiments were carried out on the platform of IBM machine with 3 GHz core i3 processor and 8 GB random access memory (RAM), running under Windows 7 operating system (OS). The algorithm was developed by ourselves based on the platform of Matlab 2014a (The Mathworks ©).

WPT versus DWT

In the first experiment, we compared DWT with WPT on a healthy brain and an Alzheimer’s disease brain, respectively. The second column shows the original image, the third column the DWT decomposition results, and the final column the WPT results. Pink colormap is employed for better view (Fig. 4).

Feature comparison

In the second experiment, we compared the proposed WPSE and WPTE (q is set to 0.8, please refer to “Optimal parameter q ”), with two types of traditional features: (i) DWT and (ii) DWT + PCA. (Note that Chaplot et al. (2006) proposed the DWT + SVM method, Zhang and Wu (2012) proposed DWT + PCA + SVM method). For fair comparison, we choose the same classifier—SVM.

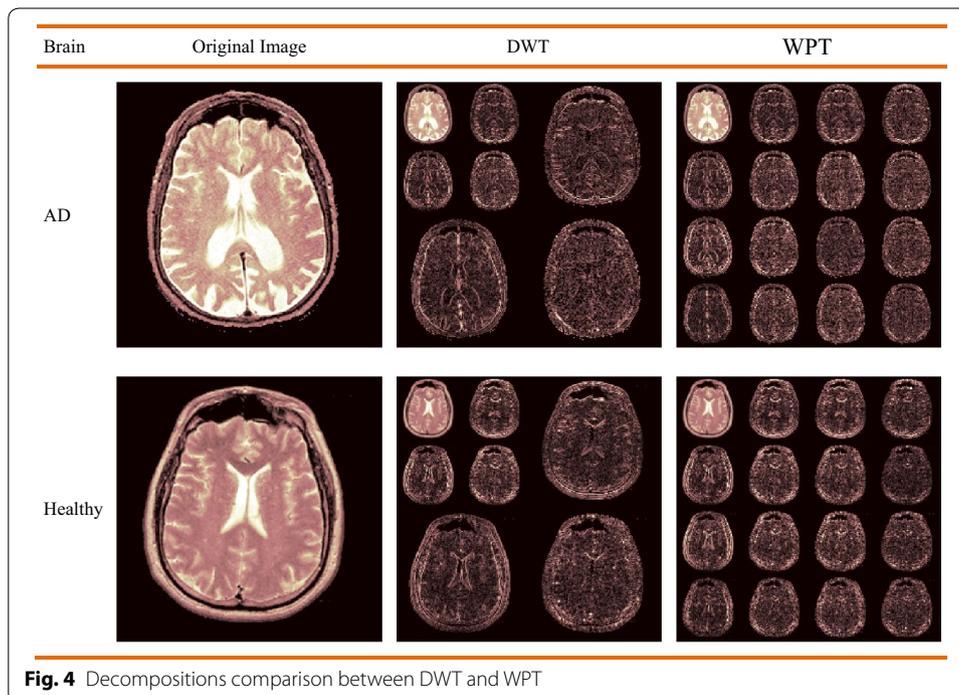


Table 4 shows that both “WPSE + SVM” achieves accuracies of 98.64, 97.12, and 97.02 % on D-66, D-160, and D-255, respectively. The “WPTE + SVM” achieves accuracies of 99.09, 98.94, and 98.39 % over three datasets. The results are better than those obtained either by “DWT + SVM (Chaplot et al. 2006)” or by “DWT + PCA + SVM (Zhang and Wu 2012)”. Therefore, we can conclude that WPSE and WPTE excel traditional feature extraction methods of “DWT” and “DWT + PCA”. Particularly, WPTE is better than WPSE. The reason is three-fold: (1) TE is a generalization of traditional SE (Tsallis 2014), and TE had been successfully applied in brain images (Amaral-Silva et al. 2014; Venkatesan and Parthiban 2014; Khader and Ben Hamza 2011). (2) The combination of TE and wavelet transform had proven to perform better than either TE or DWT in other applications (Hussain 2014; Liu et al. 2014; Chen and Li 2014). (3) Brain images entail long-range interaction and fractal-type structure, because of the self-similarity observed brain structures imaged with a finite resolution, which can be easily extracted by the corresponding wavelet packet coefficients. In summary, there are similarities at different spatial scales in brain images, which makes WPTE more suitable than WPSE in describing brains.

Table 4 Feature comparison with SVM as classifier (K-fold SCV)

Approaches	Feature #	Run #	D-66	D-160	D-255
DWT+SVM (Chaplot et al. 2006)	4761	5	96.15	95.38	94.05
DWT + PCA + SVM (Zhang and Wu 2012)	19	5	96.01	95.00	94.29
WPSE + SVM (proposed)	16	10	98.64	97.12	97.02
WPTE + SVM (proposed)	16	10	99.09	98.94	98.39

Classifier comparison

To compare the classification performance between SVM and FSVM. We set the features as WPSE and WPTE ($q = 0.8$). Then, we applied both SVM and FSVM for classification. The 10 runs of K-fold SCV results are listed below in Table 5.

Results in Table 5 shows that “WPSE + FSVM” obtains accuracies of 99.85, 99.69, 98.94 % over three datasets, which are higher than those obtained by “WPSE + SVM”. The similar results occur between “WPTE + FSVM” and “WPTE + SVM” in the way that the classification accuracy increases after SVM is replaced with FSVM. The reason is FSVM applies a FMF to each training data, so FSVM can reduce the influence of noises and outliers. In addition, the “WPTE + FSVM” performs the best among all four proposed approaches. It will be used as the default proposed method in following text.

Comparison with state-of-the-art

We compared the best proposed method (WPTE + FSVM), with 17 recent proposed methods, which consist of DWT + SOM (Chaplot et al. 2006), DWT + SVM (Chaplot et al. 2006), DWT + SVM + RBF (Chaplot et al. 2006), DWT + SVM + POLY (Chaplot et al. 2006), DWT + PCA + KNN (El-Dahshan et al. 2010), DWT + PCA + FP-ANN (El-Dahshan et al. 2010), DWT + PCA + SCG-FNN (Dong et al. 2011), DWT + PCA + SVM (Zhang and Wu 2012), DWT + PCA + SVM + RBF (Zhang and Wu 2012), DWT + PCA + SVM + IPOL (Zhang and Wu 2012), DWT + PCA + SVM + HPOL (Zhang and Wu 2012), RT + PCA + LS-SVM (Das et al. 2013), DWT + SE + SWP + PNN (Saritha et al. 2013), PCNN + DWT + PCA + BPNN (El-Dahshan et al. 2014), SWT + PCA + IABAP-FNN (Wang et al. 2015a), SWT + PCA + ABC-SPSO-FNN (Wang et al. 2015a), and WE + HMI + GEPSVM (Zhang et al. 2015d).

We averaged the results of 10 runs of K-fold SCV. The comparison results are listed in Table 6, in which some old approaches ran five times in their papers with results extracted from literature (Das et al. 2013). This experiment ran ten times to get more robust results than a five-time run.

The value of q was again assigned with 0.8 (The reason can be found in “Optimal parameter q ”). The regularization constant C were obtained via grid-search method.

Table 6 shows the proposed “WPTE + FSVM” performed better than existing state-of-the-art methods, obtaining perfect classification for the first two datasets and an accuracy of 99.49 % for D-255. This demonstrated the effectiveness of FSVM, which can reduce the effect of noise and outliers in the training points, yielding a more reliable hyperplane than standard SVM. The second best classifier is “RT + PCA + LS-SVM” (Das et al. 2013) that achieved 99.39 % for D-255.

Table 5 SVM versus FSVM (10xK-fold SCV)

Proposed approaches	D-66	D-160	D-255
WPSE + SVM	98.64	97.12	97.02
WPSE + FSVM	99.85	99.69	98.94
WPTE + SVM	99.09	98.94	98.39
WPTE + FSVM	100.00	100.00	99.49

Table 6 Classification comparison

Existing approaches	Feature #	Run #	D-66	D-160	D-255
DWT + SOM (Chaplot et al. 2006)	4761	5	94.00	93.17	91.65
DWT + SVM (Chaplot et al. 2006)	4761	5	96.15	95.38	94.05
DWT + SVM + RBF (Chaplot et al. 2006)	4761	5	98.00	97.33	96.18
DWT + SVM + POLY (Chaplot et al. 2006)	4761	5	98.00	97.15	96.37
DWT + PCA + KNN (El-Dahshan et al. 2010)	7	5	98.00	97.54	96.79
DWT + PCA + FP-ANN (El-Dahshan et al. 2010)	7	5	97.00	96.98	95.29
DWT + PCA + SCG-FNN (Dong et al. 2011)	19	5	100.00	99.27	98.82
DWT + PCA + SVM (Zhang and Wu 2012)	19	5	96.01	95.00	94.29
DWT + PCA + SVM + RBF (Zhang and Wu 2012)	19	5	100.00	99.38	98.82
DWT + PCA + SVM + IPOL (Zhang and Wu 2012)	19	5	100.00	98.12	97.73
DWT + PCA + SVM + HPOL (Zhang and Wu 2012)	19	5	98.34	96.88	95.61
RT + PCA + LS-SVM (Das et al. 2013)	9	5	100.00	100.00	99.39
DWT + SE + SWP + PNN (Saritha et al. 2013)	3	5	100.00	99.88	98.90
PCNN + DWT + PCA + BPNN (El-Dahshan et al. 2014)	7	10	100.00	98.88	98.24
SWT + PCA + IABAP-FNN (Wang et al. 2015a)	7	10	100.00	99.44	99.18
SWT + PCA + ABC-SPSO-FNN (Wang et al. 2015a)	7	10	100.00	99.75	99.02
WE + HMI + GEPSVM (Zhang et al. 2015d)	14	10	100.00	99.56	98.63
Proposed approach	Feature #	Run #	D-66	D-160	D-255
WPTE + FSVM	16	10	100.00	100.00	99.49

The italic represents the highest accuracy among all algorithms

Finally, the average evaluations based on 10 runs of the proposed WPTE + FSVM method were listed in Table 7. For D-66 and D-160, the WPTE + FSVM yielded perfect classification. For the D-255, its performance slightly decreased with sensitivity of 99.50 %, specificity of 99.43 %, precision of 99.91 %, and accuracy of 99.49 %.

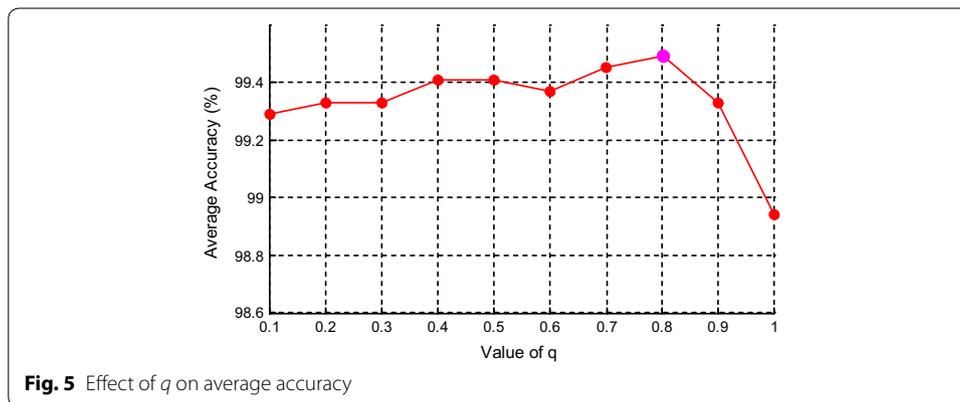
Optimal parameter q

The parameter q influences the extracted features, so it also influences classification performance. Its value should be no more than 1, since the brain image is subextensive, containing complicated regions. In this final experiment, we varied the value of q in the set of [0.1, 0.2, 0.3, ..., 0.1, 1] (Note $q = 1$ degrades WPTE to WPSE), and ran the offline training for each value. We recorded the average accuracy over 10 runs on the dataset D-255 by the proposed "WPTE + FSVM". The results are shown in Fig. 5 and Table 8.

Figure 5 demonstrates the value of q yields slight but discernible effect on average accuracy of 10 runs. As q increases to 0.8, the curve increases gradually till the highest. As q increases to 0.1, the average accuracy decreases sharply. The result again validates that WPTE ($q = 0.8$) is better than WPSE ($q = 1$).

Table 7 Average evaluation of WPTE + FSVM method based on 10 runs

	Sensitivity	Specificity	Accuracy	Precision
D-66	Perfect			
D-160	Perfect			
D-255	99.50	99.43	99.49	99.91

**Table 8** The average accuracy changes with the value of q

q	Average accuracy
0.1	99.29
0.2	99.33
0.3	99.33
0.4	99.41
0.5	99.41
0.6	99.37
0.7	99.45
0.8	99.49
0.9	99.33
1.0	98.94

This optimal result ($q = 0.8$) in this work exactly identical to three recent literatures: Sturzbecher et al. (2009), Cabella et al. (2009), and Zhang et al. (2015b). Furthermore, Diniz et al. (2010) found the fact that $q = 1.5$ for gray matter (GM), 0.1 for white matter (WM), and 0.2 for cerebrospinal fluid (CSF). Here we treat the whole brain as a single, so we must assign a single value to q . The optimal q of 0.8 can be regarded as an average of best q of GM, WM, and CSF.

Discussion on the proposed method

There were three causes to use WPT, TE, and FSVM. (1) WPT yields more features than DWT does. (2) Entropy can efficiently represent the complexity of subband coefficients, and TE is a better feature descriptor for brain structures than SE. (3) FSVM applies a FMF to each training data, so it can reduce the influence of noises and outliers.

The contributions of this work centered in three points: (i) We employed WPTE that offered better information description than WPSE. (ii) We employed FSVM that can deal with noises and outliers compared to plain SVM; and (iv) We proved the proposed “WPTE + FSVM” approach obtained superior average accuracy to 17 state-of-the-art approaches.

Conclusion and future research

In this study, we treated the PBD as a binary classification problem as pathological and healthy. To solve it, we proposed a novel feature WPTE, which used WPT to replace traditional DWT method and used TE to replace traditional SE method, and fed WPTE into FSVM. The experiments showed the proposed “WPTE + FSVM” method yielded superior performance to state-of-the-art methods.

Future work should focus on the following four aspects: (i) we will include other imaging techniques, such as DTI, FMRI and MRSI; (ii) the classification performance may increase by using other advanced variants of SVMs, such as GEPSVM (Yu et al. 2015a) and Twin SVM (Jayadeva et al. 2007). (iii) we will check the effect produced by other wavelet family and other decomposition levels. (iv) We will try to develop fine-grid search to replace the coarse-grid search technique. (v) Swarm intelligence methods (Wang et al. 2015b) will be employed to train the weights of classifiers.

Abbreviations

(A)(BP)(F)NN: (artificial) (back-propagation) (feed-forward) neural network; (D)W(P)T: (discrete) wavelet (packet) transform; (k)(F)SVM: (kernel) (Fuzzy) support vector machine; (G)(W)M: (gray) (white) matter; (S)(T)E: (shannon) (tsallis) entropy; ABC: artificial bee colony; AD: Alzheimer's disease; BBO: biogeography-based optimization; CAD: computer-aided diagnosis; CSF: cerebrospinal fluid; FMF: fuzzy membership function; HD: huntington's disease; KNN: K-nearest neighbor; MR(I): magnetic resonance (imaging); MS: multiple sclerosis; PiD: pick's disease; PBD: pathological brain detection; PNN: probabilistic neural network; QMF: quadrature mirror filter; RT: ripple transform; SCV: stratified cross validation; SDH: subdural hematoma; SOM: self-organizing map; WP(S)(T)E: wavelet packet (Shannon) (Tsallis) entropy.

Authors' contributions

YDZ and SHW conceived the study. YDZ and XJY designed the model. SHW and ZCD acquired the data. YDZ, GL and TFY analyzed the data. GL and PP interpreted the data. YDZ and ZCD developed the program. YDZ, SHW, and TFY wrote the draft. All authors gave critical revisions and approved the submission. All authors read and approved the final manuscript.

Authors' information

Prof. Dr. YD Zhang is affiliated with School of Computer Science and Technology, Nanjing Normal University. He published over 60 papers on “Alzheimer's & Dementia (IF: 12.407)”, “JAMA Psychiatry (IF: 12.008)”, “Information Sciences (IF: 4.038)”, “NMR in Biomedicine (IF: 3.044)”, “Knowledge-Based Systems (IF: 2.947)”, “Food Research International (IF: 2.818)”, “Journal of Food Engineering (IF: 2.771)”, “Basic & Clinical Pharmacology & Toxicology (IF: 2.377)”, “Sensors (IF: 2.245)”, “Expert Systems with Applications (IF: 2.240)”, “Frontiers in Computational Neuroscience (IF: 2.201)”, etc. He was listed in “2014 Most Cited Chinese Researchers (Computer Science)” released by Elsevier. Two papers were included in “ESI Highly Cited Papers”. Dr. Zhang is now the editor of “Scientia Iranica (IF: 1.025)”, “Journal of Computer Information Systems (IF: 0.722)”, “Fundamenta Informaticae (IF: 0.717)”, and “Maejo International Journal of Science and Technology (IF: 0.367)”. He is the associate editor of “Neurocomputing (IF: 2.083)”. In the past, he served as the guest editor-in-chief of “Mathematical Problems in Engineering (IF: 0.762)”, “Sensors (IF: 2.245)”, and “SIMULATION: Transactions of The Society for Modeling and Simulation International (IF: 0.818)”. He now serves as the guest editor of “Neural Network World (IF: 0.479)”.

Prof. Dr. XJ Yang is affiliated with Department of Mathematics and Mechanics, China University of Mining and Technology. He is invited to serve as the editor of “Advances in Mathematical Physics (IF 1.100)”, “Maejo International Journal of Science and Technology (IF 0.367)”, “Thermal Science (IF 1.222)”, “Central European Journal of Physics (IF 1.085)”. He was invited to serve as the guest editor-in-chief of “Central European Journal of Mathematics (IF: 0.578)”, “Advances in Mechanical Engineering (IF: 0.575)”.

Prof. Dr. ZC Dong is Associate Professor in Division of Translational Imaging, Columbia University, USA. He published over 20 papers on “JAMA Psychiatry (IF: 12.008)”, “Progress in Nuclear Magnetic Resonance Spectroscopy (IF: 7.237)”, “Neuropsychopharmacology (IF: 7.048)”, “Neuroimage (IF: 6.357)”, “Human Brain Mapping (IF: 5.969)”, “Information Sciences (IF: 4.038)”, etc. He was invited as the guest editor-in-chief of “BioMed Research International (IF: 1.579)”.

Prof. Dr. TF Yuan is affiliated with Department of Psychology, Nanjing Normal University. He published over 60 papers on “Neuron (IF: 15.05)”, “Brain (IF: 9.196)”, “Journal of Neuroscience (IF: 6.344)”, “Science Signaling (IF: 6.279)”, “Brain Research Reviews (IF: 5.930)”, “Brain Structure & Function (IF: 5.618)”, “Scientific Reports (IF 5.578)”, “Molecular Neurobiology (IF 5.137)”, “Brain Stimulation (IF: 4.399)”, “Frontiers in Cellular Neuroscience (IF: 4.289)”, etc. He is invited to serve as the editor of “PeerJ (IF 2.112)”, “Journal of Molecular Neuroscience (IF 2.343)”, and “SpringerPlus”.

Author details

¹ School of Computer Science and Technology, Nanjing Normal University, Nanjing, Jiangsu 210023, China. ² Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing, Nanjing, Jiangsu 210042, China. ³ Department of Mathematics and Mechanics, China University of Mining and Technology, Xuzhou, Jiangsu 221008, China. ⁴ Translational Imaging Division and MRI Unit, Columbia University and New York State Psychiatric Institute, New York, NY 10032, USA. ⁵ Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA. ⁶ School of Natural

Sciences and Mathematics, Shepherd University, Shepherdstown, WV 25443, USA. ⁷ School of Psychology, Nanjing Normal University, Nanjing, Jiangsu 210008, China.

Acknowledgements

This paper was supported by NSFC (610011024, 61273243, 51407095), Natural Science Foundation of Jiangsu Province (BK20150983), Program of Natural Science Research of Jiangsu Higher Education Institutions (13KJB460011, 14KJB520021), Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing (BM2013006), Key Supporting Science and Technology Program (Industry) of Jiangsu Province (BE2012201, BE2014009-3, BE2013012-2), Special Funds for Scientific and Technological Achievement Transformation Project in Jiangsu Province (BA2013058), Nanjing Normal University Research Foundation for Talented Scholars (2013119XGQ0061, 2014119XGQ0080).

Competing interests

The authors declare that they have no competing interests.

Received: 24 August 2015 Accepted: 10 November 2015

Published online: 24 November 2015

References

- Amaral-Silva H, Wichert-Ana L, Murta LO, Romualdo-Suzuki L, Itikawa E, Bussato GF, Azevedo-Marques P (2014) The superiority of Tsallis Entropy over traditional cost functions for brain MRI and SPECT registration. *Entropy* 16(3):1632–1651. doi:[10.3390/e16031632](https://doi.org/10.3390/e16031632)
- Ashkezari AD, Ma H, Saha TK, Ekanayake C (2013) Application of fuzzy support vector machine for determining the health index of the insulation system of in-service power transformers. *IEEE Trans Dielectr Electr Insul* 20(3):965–973
- Cabella BCT, Sturzbacher MJ, de Araujo DB, Neves UPC (2009) Generalized relative entropy in functional magnetic resonance imaging. *Phys A* 388(1):41–50. doi:[10.1016/j.physa.2008.09.029](https://doi.org/10.1016/j.physa.2008.09.029)
- Campos D (2010) Real and spurious contributions for the Shannon, Rényi and Tsallis entropies. *Physica A* 389(18):3761–3768
- Chaplot S, Patnaik LM, Jagannathan NR (2006) Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomed Signal Process Control* 1(1):86–92. doi:[10.1016/j.bspc.2006.05.002](https://doi.org/10.1016/j.bspc.2006.05.002)
- Chen JK, Li GQ (2014) Tsallis wavelet entropy and its application in power signal analysis. *Entropy* 16(6):3009–3025. doi:[10.3390/e16063009](https://doi.org/10.3390/e16063009)
- Damodharan S, Raghavan D (2015) Combining tissue segmentation and neural network for brain tumor detection. *Int Arab J Inf Technol* 12(1):42–52
- Das S, Chowdhury M, Kundu MK (2013) Brain MR image classification using multiscale geometric analysis of ripplet. *Prog Electromagn Res-Pier* 137:1–17. doi:[10.2528/pier13010105](https://doi.org/10.2528/pier13010105)
- Diniz PRB, Murta LO, Brum DG, de Araujo DB, Santos AC (2010) Brain tissue segmentation using q-entropy in multiple sclerosis magnetic resonance images. *Brazilian J Med Biol Res* 43(1):77–84. doi:[10.1590/s0100-879x2009007500019](https://doi.org/10.1590/s0100-879x2009007500019)
- Dong Z, Wu L, Wang S, Zhang Y (2011) A hybrid method for MRI brain image classification. *Expert Syst Appl* 38(8):10049–10053
- Dong Z, Zhang Y, Liu F, Duan Y, Kangaru A, Peterson BS (2014) Improving the spectral resolution and spectral fitting of 1H MRSI data from human calf muscle by the SPREAD technique. *NMR Biomed* 27(11):1325–1332
- Dong Z, Phillips P, Wang S, Ji G, Yang J, T-f Yuan (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* 66(9):1–15
- El-Dahshan ESA, Hosny T, Salem ABM (2010) Hybrid intelligent techniques for MRI brain images classification. *Digit Signal Proc* 20(2):433–441. doi:[10.1016/j.dsp.2009.07.002](https://doi.org/10.1016/j.dsp.2009.07.002)
- El-Dahshan ESA, Mohsen HM, Revett K, Salem ABM (2014) Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm. *Expert Syst Appl* 41(11):5526–5545. doi:[10.1016/j.eswa.2014.01.021](https://doi.org/10.1016/j.eswa.2014.01.021)
- Farzan A, Mashohor S, Ramli AR, Mahmud R (2015) Boosting diagnosis accuracy of Alzheimer's disease using high dimensional recognition of longitudinal brain atrophy patterns. *Behav Brain Res* 290:124–130. doi:[10.1016/j.bbr.2015.04.010](https://doi.org/10.1016/j.bbr.2015.04.010)
- Goh S, Dong Z, Zhang Y, DiMauro S, Peterson BS (2014) Mitochondrial dysfunction as a neurobiological subtype of autism spectrum disorder: evidence from brain imaging. *JAMA psychiatry* 71(6):665–671. doi:[10.1001/jamapsychiatry.2014.179](https://doi.org/10.1001/jamapsychiatry.2014.179)
- Harikumar R, Kumar BV (2015) Performance analysis of neural networks for classification of medical images with wavelets as a feature extractor. *Int J Imaging Syst Technol* 25(1):33–40. doi:[10.1002/ima.22118](https://doi.org/10.1002/ima.22118)
- Hussain M (2014) Mammogram enhancement using lifting dyadic wavelet transform and normalized Tsallis entropy. *J Comput Sci Technol* 29(6):1048–1057. doi:[10.1007/s11390-014-1489-7](https://doi.org/10.1007/s11390-014-1489-7)
- Jayadeva Khemchandani R, Chandra S (2007) Twin support vector machines for pattern classification. *IEEE Trans Pattern Anal Mach Intell* 29(5):905–910. doi:[10.1109/tpami.2007.1068](https://doi.org/10.1109/tpami.2007.1068)
- Khader M, Ben Hamza A (2011) Nonrigid image registration using an entropic similarity. *IEEE Trans Inf Technol Biomed* 15(5):681–690. doi:[10.1109/titb.2011.2159806](https://doi.org/10.1109/titb.2011.2159806)
- LaViolette PS, Daun MK, Paulson ES, Schmainda KM (2014) Effect of contrast leakage on the detection of abnormal brain tumor vasculature in high-grade glioma. *J Neurooncol* 116(3):543–549. doi:[10.1007/s11060-013-1318-9](https://doi.org/10.1007/s11060-013-1318-9)
- Lee SH, Lee CK, Park JB, Choi YH (2013) Diagnostic method for insulated power cables based on wavelet energy. *IEICE Electronics Express* 10(12):335–335. doi:[10.1587/elex.10.20130335](https://doi.org/10.1587/elex.10.20130335)
- Lin C-F, Wang S-D (2002) Fuzzy support vector machines. *Neural Netw IEEE Trans* 13(2):464–471. doi:[10.1109/72.991432](https://doi.org/10.1109/72.991432)

- Liu ZG, Hu QL, Cui Y, Zhang QG (2014) A new detection approach of transient disturbances combining wavelet packet and Tsallis entropy. *Neurocomputing* 142:393–407. doi:[10.1016/j.neucom.2014.04.020](https://doi.org/10.1016/j.neucom.2014.04.020)
- Munteanu CR, Fernandez-Lozano C, Abad VM, Fernandez SP, Alvarez-Linera J, Hernandez-Tamames JA, Pazos A (2015) Classification of mild cognitive impairment and Alzheimer's Disease with machine-learning techniques using H-1 magnetic resonance spectroscopy data. *Expert Syst Appl* 42(15–16):6205–6214. doi:[10.1016/j.eswa.2015.03.011](https://doi.org/10.1016/j.eswa.2015.03.011)
- Nazir M, Wahid F, Khan SA (2015) A simple and intelligent approach for brain MRI classification. *J Intell Fuzzy Syst* 28(3):1127–1135. doi:[10.3233/ifs-141396](https://doi.org/10.3233/ifs-141396)
- Padma A, Sukanesh R (2014) Segmentation and classification of brain CT images using combined wavelet statistical texture features. *Arab J Sci Eng* 39(2):767–776. doi:[10.1007/s13369-013-0649-3](https://doi.org/10.1007/s13369-013-0649-3)
- Ribbens A, Hermans J, Maes F, Vandermeulen D, Suetens P, Alzheimers Dis N (2014) Unsupervised segmentation, clustering, and groupwise registration of heterogeneous populations of brain MR images. *IEEE Trans Med Imaging* 33(2):201–224. doi:[10.1109/tmi.2013.2270114](https://doi.org/10.1109/tmi.2013.2270114)
- Saritha M, Joseph KP, Mathew AT (2013) Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. *Pattern Recogn Lett* 34(16):2151–2156. doi:[10.1016/j.patrec.2013.08.017](https://doi.org/10.1016/j.patrec.2013.08.017)
- Schwarz D, Kasperek T (2014) Brain morphometry of MR images for automated classification of first-episode schizophrenia. *Inf Fusion* 19:97–102. doi:[10.1016/j.inffus.2013.02.002](https://doi.org/10.1016/j.inffus.2013.02.002)
- Sturzbecher MJ, Tedeschi W, Cabella BCT, Baffa O, Neves UPC, De Araujo DB (2009) Non-extensive entropy and the extraction of BOLD spatial information in event-related functional MRI. *Phys Med Biol* 54(1):161–174. doi:[10.1088/0031-9155/54/1/011](https://doi.org/10.1088/0031-9155/54/1/011)
- Tsallis C (2009) Nonadditive entropy: the concept and its use. *European Phys J A* 40(3):257–266. doi:[10.1140/epja/i2009-10799-0](https://doi.org/10.1140/epja/i2009-10799-0)
- Tsallis C (2011) The nonadditive entropy S_q and its applications in physics and elsewhere: some remarks. *Entropy* 13(10):1765–1804. doi:[10.3390/e13101765](https://doi.org/10.3390/e13101765)
- Tsallis C (2014) An introduction to nonadditive entropies and a thermostistical approach to inanimate and living matter. *Contemp Phys* 55(3):179–197. doi:[10.1080/00107514.2014.900977](https://doi.org/10.1080/00107514.2014.900977)
- Venkatesan AS, Parthiban L (2014) A novel nature inspired fuzzy tsallis entropy segmentation of magnetic resonance images. *Neuroquantology* 12(2):221–229
- Wang S, Dong Z, Ji G, Zhang Y (2014) Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Prog Electromagn Res* 144:171–184
- Wang S, Zhang Y, Dong Z, Du S, Ji G, Yan J, Yang J, Wang Q, Feng C, Phillips P (2015a) Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. *Int J Imaging Syst Technol* 25(2):153–164. doi:[10.1002/ima.22132](https://doi.org/10.1002/ima.22132)
- Wang S, Zhang Y, Ji G, Yang J, Wu J, Wei L (2015b) Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic abc and biogeography-based optimization. *Entropy* 17(8):5711–5728
- Wu L, Wang S (2011) Magnetic resonance brain image classification by an improved artificial bee colony algorithm. *Prog Electromagn Res* 116:65–79
- Xian G-m (2010) An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM. *Expert Syst Appl* 37(10):6737–6741
- Yang G, Zhang Y, Yang J, Ji G, Dong Z, Wang S, Feng C, Wang Q (2015) Automated classification of brain images using wavelet-energy and biogeography-based optimization. *Multimedia Tools Appl*. doi:[10.1007/s11042-015-2649-7](https://doi.org/10.1007/s11042-015-2649-7)
- Yu D-Z, Zheng C-D, Ai J-L, Shui H-W, Gen L-J, Zheng Z, Ji Q-Y (2015a) MR brain image classification via stationary wavelet transform and generalized eigenvalue proximal SVM. *J Med Imaging Health Inform* 5(7):1–9
- Yu D, Shui H, Gen L, Zheng C (2015b) Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging. *IEEJ Transact Electr Electron Eng* 10(1):116–117. doi:[10.1002/tee.22059](https://doi.org/10.1002/tee.22059)
- Yu D, Shui H, Zheng C, Phillip P, Ji G, Yang J (2015c) 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
- Yu D, Zheng C, Gen L, Shui H (2015d) Effect of spider-web-plot in MR brain image classification. *Pattern Recogn Lett* 62:14–16. doi:[10.1016/j.patrec.2015.04.016](https://doi.org/10.1016/j.patrec.2015.04.016)
- Zhang Y, Wu L (2011) Optimal multi-level thresholding based on maximum tsallis entropy via an artificial bee colony approach. *Entropy* 13(4):841–859
- Zhang Y, Wu L (2012) An Mr brain images classifier via principal component analysis and kernel support vector machine. *Prog Electromagn Res* 130:369–388
- Zhang Y, Wang S, Ji G, Dong Z (2013) An MR brain images classifier system via particle swarm optimization and kernel support vector machine. *Sci World J* 2013:9. doi:[10.1155/2013/130134](https://doi.org/10.1155/2013/130134)
- Zhang Y, Dong Z, Phillips P, Wang S, Ji G, Yang J (2015a) Exponential wavelet iterative shrinkage thresholding algorithm for compressed sensing magnetic resonance imaging. *Inf Sci* 322:115–132. doi:[10.1016/j.ins.2015.06.017](https://doi.org/10.1016/j.ins.2015.06.017)
- Zhang Y, Dong Z, Wang S, Ji G, Yang J (2015b) Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with tsallis entropy and generalized eigenvalue proximal support vector machine (GEP-SVM). *Entropy* 17(4):1795–1813
- Zhang Y, Wang S, Phillips P, Dong Z, Ji G, Yang J (2015c) 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
- Zhang Y, Wang S, Sun P, Phillips P (2015d) Pathological brain detection based on wavelet entropy and Hu moment invariants. *Bio-Med Mater Eng* 26(s1):1283–1290
- Zhou X, Wang S, Xu W, Ji G, Phillips P, Sun P, Zhang Y (2015) Detection of pathological brain in MRI scanning based on wavelet-entropy and naive bayes classifier. In: Ortuño F, Rojas I (eds) *Bioinformatics and Biomedical Engineering*, vol 9043. Lecture Notes in Computer Science. Springer International Publishing, Granada, pp 201–209. doi:[10.1007/978-3-319-16483-0_20](https://doi.org/10.1007/978-3-319-16483-0_20)