

Article

Dual-Tree Complex Wavelet Transform and Twin Support Vector Machine for Pathological Brain Detection

Shuihua Wang^{1,2,†}, Siyuan Lu^{3,4,†}, Zhengchao Dong^{5,6}, Jiquan Yang³, Ming Yang^{7,†}
and Yudong Zhang^{1,8,9,*}

¹ School of Computer Science and Technology & School of Psychology, Nanjing Normal University, Nanjing 210023, China; wangshuihua@njnu.edu.cn (S.W.)

² Key Laboratory of Statistical information technology and data mining, State Statistics Bureau, Chengdu 610225, China

³ Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing, Nanjing 210042, China; siyuan_lu@foxmail.com (S.L.); jiquany@126.com (J.Y.)

⁴ Key laboratory of symbolic computation and knowledge engineering of ministry of education, Jilin University, Changchun 130012, China

⁵ Translational Imaging Division, Columbia University, New York, NY 10032, USA; zhengchaodongcu@gmail.com (Z.D.)

⁶ State Key Lab of CAD & CG, Zhejiang University, Hangzhou 310027, China

⁷ Department of Radiology, Nanjing Children's Hospital, Nanjing Medical University, Nanjing 210008, China; yangming19710217@163.com (M.Y.)

⁸ Guangxi Key Laboratory of Manufacturing System & Advanced Manufacturing Technology, Guilin 541004, China

⁹ Department of Neurology, First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China

* Correspondence: zhangyudong@njnu.edu.cn; Tel.: +86-159-0518-3664

† These authors contributed equally to this work.

Academic Editor: Christian Dawson

Received: 28 March 2016; Accepted: 30 May 2016; Published: 3 June 2016

Abstract: (Aim) Classification of brain images as pathological or healthy case is a key pre-clinical step for potential patients. Manual classification is irreproducible and unreliable. In this study, we aim to develop an automatic classification system of brain images in magnetic resonance imaging (MRI). (Method) Three datasets were downloaded from the Internet. Those images are of T2-weighted along axial plane with size of 256×256 . We utilized an s -level decomposition on the basis of dual-tree complex wavelet transform (DTCWT), in order to obtain 12s “variance and entropy (VE)” features from each subband. Afterwards, we used support vector machine (SVM) and its two variants: the generalized eigenvalue proximal SVM (GEPSVM) and the twin SVM (TSVM), as the classifiers. In all, we proposed three novel approaches: DTCWT + VE + SVM, DTCWT + VE + GEPSVM, and DTCWT + VE + TSVM. (Results) The results showed that our “DTCWT + VE + TSVM” obtained an average accuracy of 99.57%, which was not only better than the two other proposed methods, but also superior to 12 state-of-the-art approaches. In addition, parameter estimation showed the classification accuracy achieved the largest when the decomposition level s was assigned with a value of 1. Further, we used 100 slices from real subjects, and we found our proposed method was superior to human reports from neuroradiologists. (Conclusions) This proposed system is effective and feasible.

Keywords: magnetic resonance imaging; parameter estimation; support vector machine; dual-tree complex wavelet transform; twin support vector machine; variance; entropy

1. Introduction

Stroke, brain tumors, neurodegenerative diseases, and inflammatory/infectious diseases are the four main types of brain diseases. Stroke is also called vascular disease of cerebral circulation. Brain tumors occur when abnormal cells form inside the brain. Neurodegenerative diseases occur when neurons lose structure or function progressively. Inflammatory/infectious disease suffers from inflammation or infection in or around the brain tissues. All diseases cause serious problems for both patients and the society. Hence, it is important to make early diagnosis system, with the aim of providing more opportunities for better clinical trials. This type of task is commonly named as “pathological brain detection (PBD)”.

Magnetic resonance imaging (MRI) offers the best diagnostic information in the brain; however, to make a diagnosis usually needs human manual interpretation. Existing manual methods are costly, tedious, lengthy, and irreproducible, because of the huge volume data of MRI. Those shortcomings lead to the necessity to develop automatic tools such as computer-aided diagnosis (CAD) systems [1,2]. Due to the better performance provided by magnetic resonance (MR) images, many CAD systems are based on MR images [3].

Existing methods over brain CAD systems could be divided into two types according to the data dimension. One type is for three-dimensional (3D) image, but it needs to scan the whole brain. The other type is based on a single-slice that contains the disease related areas, which is cheap and commonly used in Chinese hospitals. El-Dahshan *et al.* [4] employed a 3-level discrete wavelet transform (DWT), followed by principal component analysis (PCA) to reduce features. Finally, they used K-nearest neighbor (KNN) for classification. Patnaik *et al.* [5] utilized DWT to extract the approximation coefficients. Then, they employed support vector machine (SVM) for classification. Dong *et al.* [6] further suggested to train feedforward neural network (FNN) with a novel scaled conjugate gradient (SCG) approach. Their proposed classification method achieved good results in MRI classification. Wu [7] proposed to use kernel SVM (KSVM), and suggested three new kernels: Gaussian radial basis, homogeneous polynomial, and inhomogeneous polynomial. Das *et al.* [8] proposed to combine Ripplet transform (RT) and PCA and least square SVM (LS-SVM), and the 5×5 cross validation test showed high classification accuracies. El-Dahshan *et al.* [9] used the feedback pulse-coupled neural network to preprocess the MR images, the DWT and PCA for features extraction and reduction, and the FBPNN to detect pathological brains from normal brains. Dong *et al.* [10] combined discrete wavelet packet transform (DWPT) and Tsallis entropy (TE). In order to segment and classify malignant and benign brain tumor slices in Alzheimer’s disease (AD), Wang *et al.* [11] employed stationary wavelet transform (abbreviated as SWT) to replace the common used DWT. Besides, they proposed a hybridization of Particle swarm optimization and Artificial bee colony (HPA) algorithm to obtain the optimal weights and biases of FNN. Nazir *et al.* [12] implemented denoising at first. Their method achieved an overall accuracy of 91.8%. Sun *et al.* [13] combined wavelet entropy (WE) and Hu moment invariant (HMI) as features.

After analyzing the above methods, we found all literatures treated the PBD problem as a classification problem and their studies aimed at improving the classification accuracy. As we know, a standard classification task is composed of two steps: feature extraction and classifier training.

The former step “feature extraction” in PBD usually employed discrete wavelet transform (DWT) [5]. The reason is DWT can provide multiresolution analysis at any desired scale for a particular brain image. Besides, the abundant texture features in various brain regions are coherent with wavelet analysis [14,15]. However, DWT is shift variant, *i.e.*, a slight shift in the image degrades the performance of DWT-based classification. Our study chose a variant of DWT, *viz.*, the dual-tree complex wavelet transform (DTCWT). Scholars have proven that the DTCWT offers “*more directional selectivity*” than canonical DWT, with merely 2^n redundancy for data of n -dimensional [16–18].

Although stationary wavelet transform (SWT) can also deal with the shift variance problem, it will lead to more redundancy than DTCWT. Then, we need to extract features from the DTCWT results. In this paper, we proposed to use variance and entropy (VE) [19] from all the subbands of DTCWT.

Although energy is also a common feature extracted from the wavelet subbands, scholars have proven it is not as efficient as entropy in MR image classification [20,21]. Besides, variance with the form of $E[(x - \mu)^2]$ (E , expectation; x , random variable; μ , the mean) indicates how data points be close to the mean value, while energy with the form of $E[x^2]$ does not consider the mean value. Hence, it is self-explanatory that variance will perform better than energy even if the expected mean value has a slight shift.

The latter step “classifier training” usually employed support vector machines (SVMs) or artificial neural network. Compared with other conventional classification methods, SVMs have significant advantages of elegant mathematical tractability [22], direct geometric interpretation [23], high accuracy [24], *etc.* Hence, we continued to choose SVM. Besides, two variants of SVM were introduced in this study: generalized eigenvalue proximal SVM (GEPSVM) and twin SVM (TSVM), with the aim of augmenting the classification performance further.

The contribution of our study is three-fold: We applied DTCWT to pathological brain detection. We applied both variance and entropy to extract features. We applied TSVM for classification. The structure of the paper is organized as follows: Section 2 contains the materials used in this study. Section 3 presents the dual-tree complex wavelet transform, and offers the mathematical fundamental of SVM and its two variants. Section 4 designs the experiment and gives the evaluation measures. Section 5 contains the experimental results and offers the discussions. Section 6 is devoted to conclusions. The abbreviations used in this work are listed in the end of this paper.

2. Materials

At present, there are three benchmark datasets as Dataset66, Dataset160, and Dataset255, of different sizes of 66-, 160-, and 255-images, respectively. All datasets contain T2-weighted MR brain images obtained along axial plane with size of 256×256 . We downloaded all the slices of subjects from the website of Medical School of Harvard University (Boston, MA, USA) [25]. Then, we selected five slices from each subject. The selection criterion is that for healthy subjects, these slices were selected at random. For pathological subjects, the slices should contain the lesions by confirmation of three radiologists with ten years of experiences.

The former two datasets (Dataset66 & Dataset160) consisted of 7 types of diseases (meningioma, AD, AD plus visual agnosia, sarcoma, Pick’s disease, Huntington’s disease, and glioma) along with normal brain images. The last dataset “Dataset255” contains all 7 types of diseases as mentioned before, and 4 new diseases (multiple sclerosis, chronic subdural hematoma, herpes encephalitis, and cerebral toxoplasmosis).

Figure 1 shows samples of brain MR images. Our method is for hospital other than research. In Chinese hospitals, we usually scanned one slice that is closest to the potential focus, other than the whole brain. Hence, one slice was obtained from one subject. Each slice in Figure 1 is selected from regions related to the foci of diseases (in total 26 axial slices).

Note that we treated all different diseased brains as pathological, so our task is a two-class classification problem, that is, to detect pathological brains from healthy brains. The whole image is treated as the input. We did not choose local characteristics like point and edge, and we extract global image characteristics that are further learned by the CAD system. Note that our method is different from the way neuroradiologists do. They usually select local features and compare with standard template to check whether focuses exist, such as shrink, expansion, bleeding, inflammation, *etc.* While our method is like AlphaGo [26], the computer scientists give the machine enough data, and then the machine can learn how to classify automatically.

Including subjects’ information (age, gender, handedness, memory test, *etc.*) can add more information, and thus may help us to improve the classification performance. Nevertheless, this CAD system in our study is only based on the imaging data. Besides, the imaging data from the website does not contain the subjects’ information.

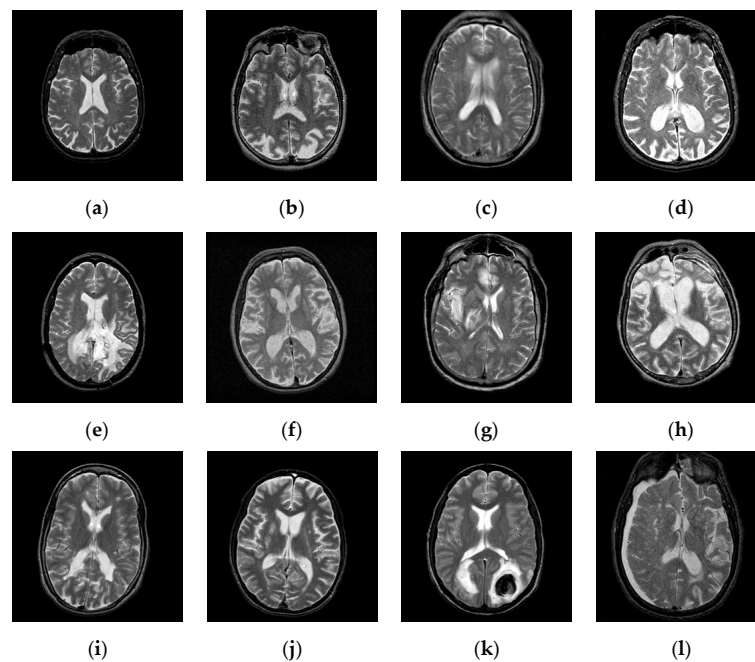


Figure 1. Sample of pathological brains. (a) Normal brain; (b) Alzheimer's disease (AD) with visual agnosia; (c) Meningioma; (d) AD; (e) Glioma; (f) Huntington's disease; (g) Herpes encephalitis; (h) Pick's disease; (i) Multiple sclerosis; (j) Cerebral toxoplasmosis; (k) Sarcoma; (l) Subdural hematoma.

The cost of predicting pathological to healthy is severe; because the patients may be told that, she/he is healthy and thus ignore the mild symptoms displayed. The treatments of patients may be deferred. Nevertheless, the cost of misclassification of healthy to pathological is low, since correct remedy can be given by other diagnosis means.

This cost-sensitivity (CS) problem was solved by changing the class distribution at the beginning state, since original data was accessible. That means, we intentionally picked up more pathological brains than healthy ones into the dataset, with the aim of making the classifier biased to pathological brains, to solve the CS problem [27]. The overfitting problem would be monitored by cross validation technique.

3. Methodology

The proposed method consists of three decisive steps: wavelet analysis by dual-tree complex wavelet transform (DTCWT), feature extraction by "Variance & Entropy (VE)", and classification by three independent classifiers (SVM, GEPSVM, and TSVM). Figure 2 illustrates our modular framework. The output of DTCWT is wavelet subband coefficients, which are then submitted to VE block.

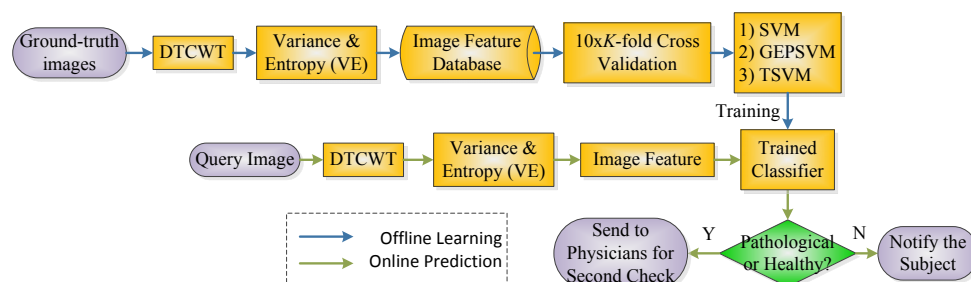


Figure 2. Modular framework of the proposed system for magnetic resonance (MR) brain classification (K may be 5 or 6 according to the dataset).

3.1. Discrete Wavelet Transform

The discrete wavelet transform (DWT) is an image processing method [28] that provides multi-scale representation of a given signal or image [29]. Standard DWT is vulnerable to shift variance problem, and only has horizontal and vertical directional selectivity [30]. Suppose s represents a particular signal, n represents the sampling point, h and g represents a high-pass filter and low-pass filter, respectively, H and L represents the coefficients of high-pass and low-pass subbands. We have

$$H(n) = \sum_m h(2n - m)s(m) \quad (1)$$

$$L(n) = \sum_m g(2n - m)s(m) \quad (2)$$

Figure 3 shows the directional selectivity of DWT. The LH denotes a low-pass filter along x -axis and high-pass filter along y -axis. HL denotes a high-pass filter along x -axis followed by a low-pass filter along y -axis. The LL denotes low-pass filters along both directions, and HH denotes high-pass filters along both directions.

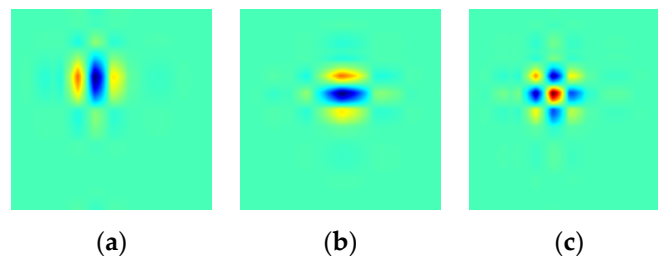


Figure 3. Directional selectivity of discrete wavelet transform (DWT). (L = Low, H = High). (a) LH ; (b) HL ; (c) HH .

Here the HL and LH have well-defined for both vertical and horizontal orientations. For the HH , it mixes directions of both -45 and $+45$ degrees together, which stems from the use of real-valued filters in DWT. This mixing also impedes the direction check [31].

3.2. Dual-Tree Complex Wavelet Transform

To help improve the directional selectivity impaired by DWT, we proposed a dual-tree DWT, which was implemented by two separate two-channel filter banks. Note that the scaling and wavelet filters in the dual-tree cannot be selected arbitrarily [32]. In one tree, the wavelet and scaling filters should produce a wavelet and scaling function, which are approximate Hilbert transforms of those generated by another tree [33]. In this way, the wavelet generated from both trees and the complexed-valued scaling function are approximately analytic, and are called dual-tree complex wavelet transform (DTCWT).

DTCWT obtained directional selectivity by using approximately analytic wavelets, *i.e.*, they have support on only one half of the whole frequency domain [34]. At each scale of a 2D DTCWT, it produces in total six directionally selective subbands ($\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$) for both real (\mathbb{R}) and imaginary (\mathbb{I}) parts [35]. Figure 4 shows the directional selectivity of DTCWT. The first row depicts the 6 directional wavelets of the real oriented DTCWT, and the second row shows the imaginary counterpart. The \mathbb{R} and \mathbb{I} parts are oriented in the same direction, and they together form the DTCWT as

$$\mathbb{M} = \sqrt{\mathbb{R}^2 + \mathbb{I}^2} \quad (3)$$

where \mathbb{M} represents the magnitude of the DTCWT coefficients.

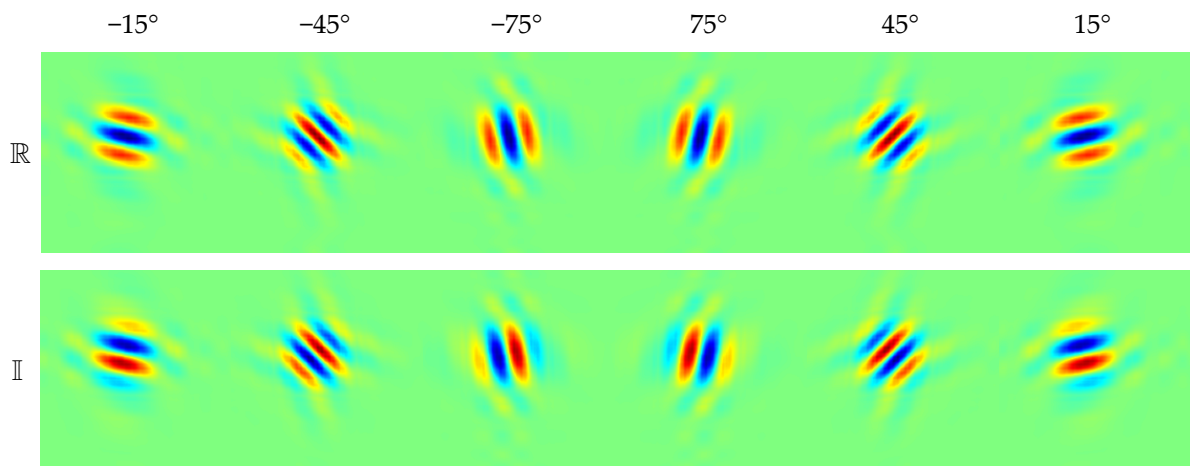


Figure 4. Directional Selectivity of dual-tree complex wavelet transform (DTCWT). (\mathbb{R} = Real, \mathbb{I} = Imaginary).

3.3. Comparison between DWT and DTCWT

To compare the directional selectivity ability between DWT and DTCWT, we performed a simulation experiment. Two simulation images (a pentagon and a heptagon) were generated. We decomposed both images to 4-level by DWT and DTCWT, respectively. Then, we reconstructed them to obtain an approximation to original images by a 4-th level detail subband. The results were shown in Figure 5.

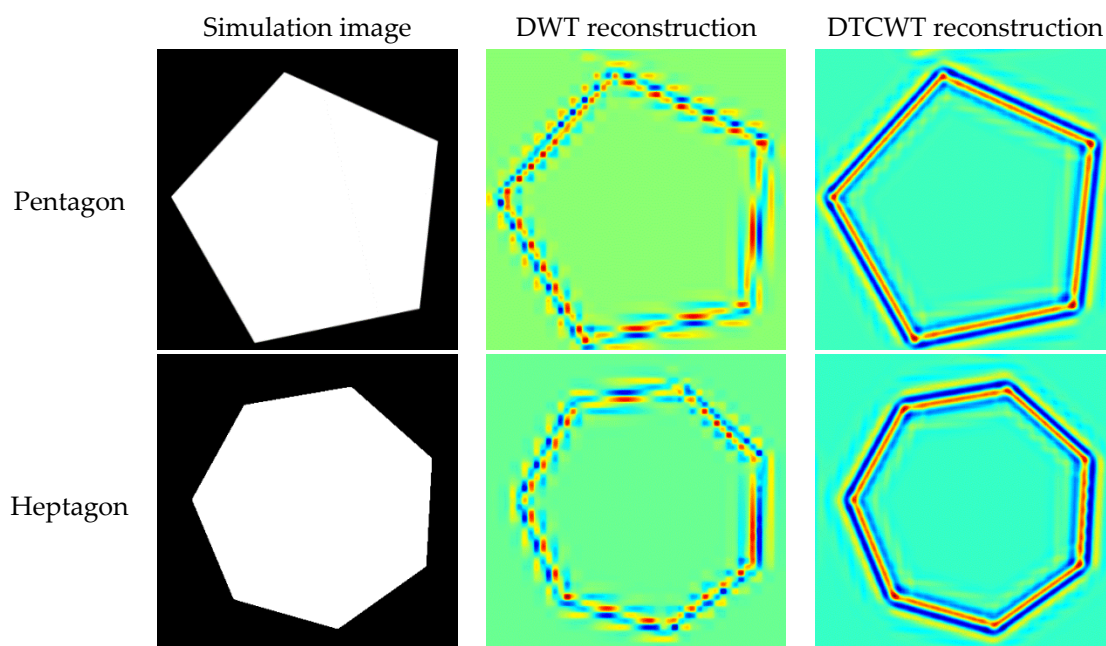


Figure 5. The reconstruction comparison between DWT and DTCWT.

The first column in Figure 5 shows the simulation images, the second column shows the DWT reconstruction results, and the last column shows the DTCWT reconstruction results. Both DWT and DTCWT can extract edges from detail subbands, which are abundant in brain tissues.

Those edges are discriminant features that are different in pathological brains and healthy brains. The reason is all the focus-related areas contain either shrink, or expand, or bleed, or become inflamed. Those that will yield structural alternations that are associated with edges. We find from the last

column in Figure 5 that the edges detected by DTCWT have a clear contour, so DTCWT can detect nearly all directions clearly and perfectly. Nevertheless, the edges detected by DWT (See Figure 5) are discontinued, stemming from that DWT can only detect horizontal and vertical edges. The results fall in line with Figure 3.

3.4. Variance and Entropy (VE)

Based on the coefficients of DTCWT, we extract variance and entropy (VE) features for each decomposition level s and each direction d . Suppose (x, y) is the spatial coordinate of the corresponding subband, and (L, W) is the length and width of the corresponding subband, we can define the variance $V_{(s,d)}$ as

$$V_{(s,d)} = \frac{1}{LW} \sum_{x=1}^L \sum_{y=1}^W \left(\mathbb{M}_{(s,d)}(x, y) - \mu_{(s,d)} \right)^2 \quad (4)$$

$$\mu_{(s,d)} = \frac{1}{LW} \sum_{x=1}^L \sum_{y=1}^W \mathbb{M}_{(s,d)}(x, y)$$

here μ denotes the mathematical expectation of \mathbb{M} . The variance V measures the spread of grey-level distribution of the subbands. The larger the value of V is, the more widely the gray-levels of the image vary. V also reflect the contrast of the texture.

Another indicator is the entropy E , which measures the randomness of the gray-level distribution [36]. The larger the value of V is, the more randomly the gray-level distribution spreads [37]. The entropy E is defined in following form:

$$E_{(s,d)} = -\frac{1}{LW} \sum_{x=1}^L \sum_{y=1}^W \mathbb{P} \left(\mathbb{M}_{(s,d)}(x, y) \right) \log \mathbb{P} \left(\mathbb{M}_{(s,d)}(x, y) \right) \quad (5)$$

Here, \mathbb{P} denotes the probability function. Both variance and entropy are sufficient to produce a good performance. All directional subbands of those two kinds of features are combined to form a new feature set \mathbf{V}_s and \mathbf{E}_s as

$$\mathbf{V}_s = \frac{[V_{(s,-15)}, V_{(s,15)}, V_{(s,-45)}, V_{(s,45)}, V_{(s,-75)}, V_{(s,75)}]}{\sqrt{V_{(s,-15)}^2 + V_{(s,15)}^2 + V_{(s,-45)}^2 + V_{(s,45)}^2 + V_{(s,-75)}^2 + V_{(s,75)}^2}} \quad (6)$$

$$\mathbf{E}_s = \frac{[E_{(s,-15)}, E_{(s,15)}, E_{(s,-45)}, E_{(s,45)}, E_{(s,-75)}, E_{(s,75)}]}{\sqrt{E_{(s,-15)}^2 + E_{(s,15)}^2 + E_{(s,-45)}^2 + E_{(s,45)}^2 + E_{(s,-75)}^2 + E_{(s,75)}^2}} \quad (7)$$

Hence, we extract 12 features for each scale, among which 6 is for \mathbf{V}_s and 6 for \mathbf{E}_s . For an s -level decomposition, we totally obtain $12s$ features \mathbf{VE} as

$$\mathbf{VE} = [\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_s, \mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_s] \quad (8)$$

The $12s$ \mathbf{VE} s were then submitted into classifiers. SVM is now probably treated as one of the most excellent classification approach in small-size (less than one-thousand samples) problem [7]. To further enhance the classification performance, two new variants of SVM were introduced:

3.5. Generalized Eigenvalue Proximal SVM

Mangasarian and Wild [38] proposed the generalized eigenvalue proximal SVM (GEPSVM). It drops the parallelism condition on the two hyperplanes (remember the parallelism is necessary in original SVM). Latest literatures showed that GEPSVM yielded superior classification performance to canonical support vector machines [39,40].

Suppose samples are from either class 1 (denote by symbol X_1) or class 2 (denoted by symbol X_2), respectively. The GEPSVM finds the two optimal nonparallel planes with the form of (\mathbf{w} and b denotes the weight and bias of the classifier, respectively)

$$\mathbf{w}_1^T x - b_1 = 0 \text{ and } \mathbf{w}_2^T x - b_2 = 0 \quad (9)$$

To obtain the first plane, we deduce from Equation (9) and get the following solution

$$(\mathbf{w}_1, b_1) = \arg \min_{(\mathbf{w}, b) \neq 0} \frac{\|\mathbf{w}^T X_1 - o^T b\|^2 / \|z\|^2}{\|\mathbf{w}^T X_2 - o^T b\|^2 / \|z\|^2} \quad (10)$$

$$z \leftarrow \begin{bmatrix} \mathbf{w} \\ b \end{bmatrix} \quad (11)$$

where o is a vector of ones of appropriate dimensions. Simplifying formula (10) gives

$$\min_{(\mathbf{w}, b) \neq 0} \frac{\|\mathbf{w}^T X_1 - o^T b\|^2}{\|\mathbf{w}^T X_2 - o^T b\|^2} \quad (12)$$

We include the Tikhonov regularization to decrease the norm of z , which corresponds to the first hyperplane. The new equation including Tikhonov regularization term is:

$$\min_{(\mathbf{w}, b) \neq 0} \frac{\|\mathbf{w}^T X_1 - o^T b\|^2 + t \|z\|^2}{\|\mathbf{w}^T X_2 - o^T b\|^2} \quad (13)$$

where t is a positive (or zero) Tikhonov factor. Formula (13) turns to the “Rayleigh Quotient (RQ)” in the following form of

$$z_1 = \arg \min_{z \neq 0} \frac{z^T P z}{z^T Q z} \quad (14)$$

where P and Q are symmetric matrices in $\mathbb{R}^{(p+1) \times (p+1)}$ in the forms of

$$P \leftarrow \begin{bmatrix} X_1 & -o \end{bmatrix}^T \begin{bmatrix} X_1 & -o \end{bmatrix} + tI \quad (15)$$

$$Q \leftarrow \begin{bmatrix} X_2 & -o \end{bmatrix}^T \begin{bmatrix} X_2 & -o \end{bmatrix} \quad (16)$$

Solution of (14) is deduced by solving a generalized eigenvalue problem, after using the stationarity and boundedness characteristics of RQ.

$$Pz = \lambda Qz, z \neq 0 \quad (17)$$

Here the optimal minimum of (14) is obtained at an eigenvector z_1 corresponding to the smallest eigenvalue λ_{\min} of formula (17). Therefore, \mathbf{w}_1 and b_1 can be obtained through formula (11), and used to determine the plane in formula (9). Afterwards, a similar optimization problem is generated that is analogous to (12) by exchanging the symbols of X_1 and X_2 . z_2^* can be obtained in a similar way.

3.6. Twin Support Vector Machine

In 2007, Jayadeva *et al.* [41] presented a novel classifier as twin support vector machine (TSVM). The TSVM is similar to GEPSVM in the way that both obtain non-parallel hyperplanes. The difference lies in that GEPSVM and TSVM are formulated entirely different. Both quadratic programming (QP) problems in TSVM pair are formulated as a typical SVM. Reports have shown that TSVM is better

than both SVM and GEPSVM [42–44]. Mathematically, the TSVM is constructed by solving the two QP problems

$$\begin{aligned} \min_{w_1, b_1, q} \quad & \frac{1}{2} (X_1 \mathbf{w}_1 + o_1 b_1)^T (X_1 \mathbf{w}_1 + o_1 b_1) + c_1 o_2^T q \\ \text{s.t.} \quad & - (X_2 \mathbf{w}_1 + o_2 b_1) + q \geq o_2, q \geq 0 \end{aligned} \quad (18)$$

$$\begin{aligned} \min_{w_2, b_2, q} \quad & \frac{1}{2} (X_2 \mathbf{w}_2 + o_2 b_2)^T (X_2 \mathbf{w}_2 + o_2 b_2) + c_2 o_1^T q \\ \text{s.t.} \quad & - (X_1 \mathbf{w}_2 + o_1 b_2) + q \geq o_1, q \geq 0 \end{aligned} \quad (19)$$

here q is a nonnegative slack variance. c_i ($i = 1, 2$) are positive parameters, and o_i ($i = 1, 2$) is the same as in formula (10). By this mean, the TSVM constructed two hyperplanes [45]. The first term in equations of (18) and (19) is the sum of squared distances. The second one represents the sum of error variables. Therefore, minimizing Equations (18) and (19) will force the hyperplanes approximate to data in each class, and minimize the misclassification rate [46]. Finally, the constraint requires the hyperplane to be at a distance of more than one from points of the other class. Another advantage of TSVM is that its convergence rate is four times faster than conventional SVM [47].

3.7. Pseudocode of the Whole System

The implementation covers two phases: offline learning and online prediction, with the goal of training the classifier and prediction of new instances, respectively. Table 1 offers the pseudocode of proposed methods.

Table 1. Pseudocode of our system.

Phase I: Offline learning		
Step A	Wavelet Analysis	Perform s -level dual-tree complex wavelet transform (DTCWT) on every image in the ground-truth dataset
Step B	Feature Extraction	Obtain $12 \times s$ features ($6 \times s$ Variances and $6s$ Entropies, and s represents the decomposition level) from the subbands of DTCWT
Step C	Training	Submit the set of features together with the class labels to the classifier, in order to train its weights/biases.
Step D	Evaluation	Record the classification performance based on a $10 \times K$ -fold stratified cross validation.
Phase II: Online prediction		
Step A	Wavelet Analysis	Perform s -level DTCWT on the real query image (independent from training images)
Step B	Feature Extraction	Obtain VE feature set
Step C	Prediction	Feed the VE feature set into the trained classifier, and obtain the output.

4. Experiment Design

4.1. Statistical Setting

In order to carry out a strict statistical analysis, stratified cross validation (SCV) was used since it is a model validation technique for small-size data [48]. 6-fold SCV was employed for Dataset66, and 5-fold SCV was employed for Dataset160 and Dataset255. The SCV setting was listed in Table 2.

Table 2. Stratified cross validation (SCV) setting of all datasets.

Dataset	No. of Fold	Training		Validation		Total	
		H	P	H	P	H	P
Dataset66	6	15	40	3	8	18	48
Dataset160	5	16	112	4	28	20	140
Dataset255	5	28	176	7	44	35	220

(H = Healthy, P = Pathological).

The 10-fold is not used because of two reasons: One is that past literatures used the same setting as Table 2. Another is for stratification, viz., we expect to guarantee each fold covers the same class numbers. If we divide the dataset into 10-folds, then the stratification will be breached.

Figure 6 illustrates an example of K -fold SCV, by which the dataset is partitioned into K folds with the same class distributions. The $(K-1)$ folds are used for training, and the rest fold for test, *i.e.*, query images come from the rest fold. The evaluation is based on test images. This above process repeats K times so that each fold is used as test once. The final accuracy of K -fold SCV is obtained by averaging the K results. The K -fold SCV repeats 10 times to further remove randomness (See Section 5.4).

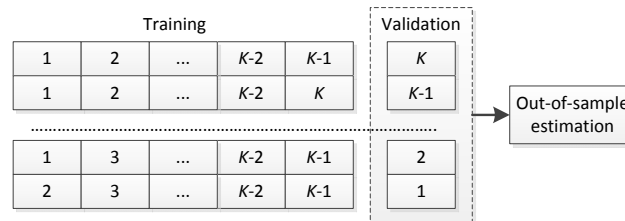


Figure 6. A K -fold SCV.

4.2. Parameter Estimation for s

It remains an issue of finding the optimal value of decomposition level s . From the view of information provided, a smaller s offers less information and a larger s offers more information to the classifier. From the view of avoiding overfit, a smaller s may prevent overfitting in greater degree than a larger s . This study used grid-search [49] method to find the optimal value of s , *i.e.*, we vary the value of s from 1 to 5 with increment of 1, and check the corresponding average accuracies. The one associated with the largest accuracy is the optimal value of s .

4.3. Evaluation

The pathological brains are treated as positive, while the healthy brains as negative. To evaluate the performance, we first calculated overall confusion matrix of 10 runs, then calculate the TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). The pathological brains were set to true and healthy ones to false, following common convention. The classification accuracy (Acc) is defined as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (20)$$

5. Results and Discussions

The algorithms were in-house developed based on 64 bit Matlab 2015a (The Mathworks ©, Natick, MA, USA). Figure 7 shows the graphical user interface (GUI). The simulation experiments were implemented on the platform of P4 IBM with 3.2 GHz processor, and 8 GB random access memory (RAM), running under Windows 7 operating system.

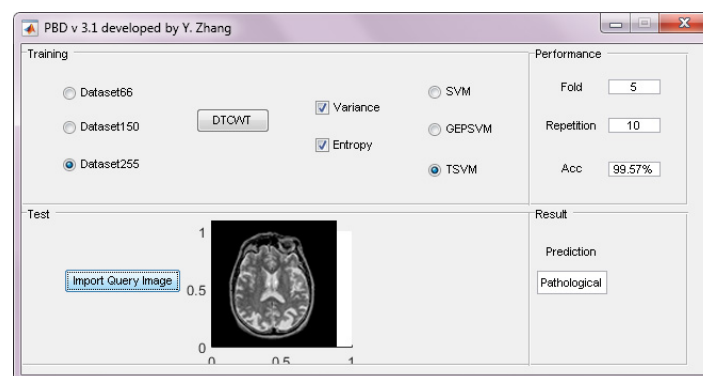


Figure 7. Graphical user interface (GUI) of our developed programs.

5.1. Classifier Comparison

In the second experiment, we compared three classifiers: SVM, GEPSVM, and TSVM. All three datasets are tested. A 10 runs of k-fold SCV was carried out. Accuracy was used for evaluation. The results are listed in Table 3. The SVM achieved 100.00%, 99.69%, and 98.43% accuracy for Dataset66, Dataset160, and Dataset255, respectively. The GEPSVM achieved accuracy of 100.00%, 99.75%, and 99.25% for three datasets. The TSVM yielded accuracy of 100.00%, 100.00%, and 99.57%, respectively.

Table 3. Accuracy Comparison based on 10 runs of k-fold SCV (Unit: %).

Our Methods	Dataset66	Dataset160	Dataset255
DTCWT + VE + SVM	100.00	99.69	98.43
DTCWT + VE + GEPSVM	100.00	99.75	99.25
DTCWT + VE + TSVM	100.00	100.00	99.57

Data in Table 3 indicate that GEPSVM is superior to standard SVM. For Dataset160, the *acc* of GEPSVM is higher than that of SVM by 0.06%. For Dataset255, the *acc* of GEPSVM is higher than that of SVM by 0.82%. Meanwhile, TSVM is superior to GEPSVM. For Dataset160, the *acc* of TSVM is 0.25 higher than that of GEPSVM. For Dataset255, the *acc* of TSVM is 0.32% higher than that of GEPSVM.

The parallel hyperplane setting restrains standard SVM to generate complicated and flexible hyperplanes. GEPSVM and TSVM discard this setting, so their performances are much better than SVM. TSVM is similar to GEPSVM in spirit, since both use non-parallel hyperplanes. The difference between them is TSVM uses simpler formulation than GEPSVM, and the former can be solved by merely two QP problems. Our results align with the finding in Kumar and Gopal [50], which says “*generalization performance of TSVM is better than GEPSVM and conventional SVM*”. In following experiments, TSVM is the default classifier

5.2. Optimal Decomposition Level Selection

The value of decomposition level s was set in the range of (1, 2, 3, 4, 5). We chose the TSVM. All datasets were tested. A 10 runs of k-fold SCV was implemented with varying s . The curve of average accuracy *versus* against decomposition level is shown in Figure 8.

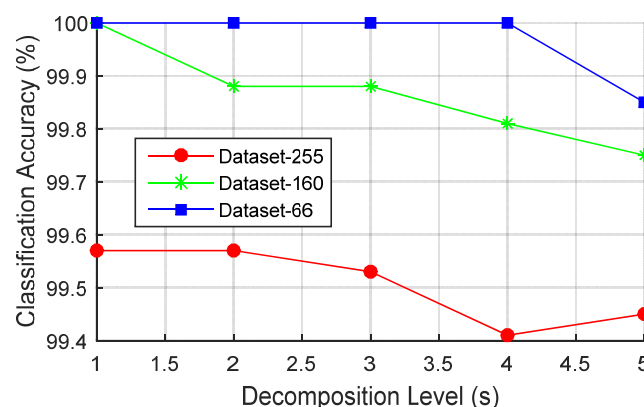


Figure 8. Classification Accuracy *versus* Decomposition Level (s).

Remember for $s = 1$, only 12 features are used. For $s = 2$, in total 24 features are used. The number of employed features is 12 times of the value of s . Figure 8 shows the relationship between *Acc versus s*. Here we find the *acc* has a tendency of decrease when the decomposition level s increases. The reason is more features will attenuate the classification performance [51]. Reducing the number of features

can simplify the model, cost shorter training time, and augment generalization performance through reduction of variance [52].

5.3. Comparison to State-of-the-Art Approaches

We have already compared the SVM with its variants in Section 5.1. In this section, we compared the best of the proposed methods (DTCWT + VE + TSVM) with 12 state-of-the-art methods: DWT + PCA + KNN [4], DWT + SVM + RBF [5], DWT + PCA + SCG-FNN [6], DWT + PCA + SVM + RBF [7], RT + PCA + LS-SVM [8], PCNN + DWT + PCA + BPNN [9], DWPT + TE + GEPSVM [10], SWT + PCA + HPA-FNN [11], WE + HMI + GEPSVM [13], SWT + PCA + GEPSVM [53], FRFE + WTT + SVM [54], and SWT + PCA + SVM + RBF [55]. The meaning of these abbreviations can be found in the Abbreviations Section. The accuracy results were extracted directly from above literatures. The comparison is based on results obtained in each individual study.

Table 4 shows the comparison results between the best proposed method “DTCWT + VE + TSVM” with the state-of-the-art approaches. The first column lists the method name, the second column the number of features employed, the third column the total run times (all algorithms run 10 times, except some old algorithms ran five times which were reported in literature [8]), and the last three columns the average *acc* over three datasets.

Table 4. Classification comparison.

Algorithms	Feature #	Run #	Acc		
			Dataset66	Dataset160	Dataset255
DWT + PCA + KNN [4]	7	5	98.00	97.54	96.79
DWT + SVM + RBF [5]	4761	5	98.00	97.33	96.18
DWT + PCA + SCG-FNN [6]	19	5	100.00	99.27	98.82
DWT + PCA + SVM + RBF [7]	19	5	100.00	99.38	98.82
RT + PCA + LS-SVM [8]	9	5	100.00	100.00	99.39
PCNN + DWT + PCA + BPNN [9]	7	10	100.00	98.88	98.24
DWPT + TE + GEPSVM [10]	16	10	100.00	100.00	99.33
SWT + PCA + HPA-FNN [11]	7	10	100.00	100.00	99.45
WE + HMI + GEPSVM [13]	14	10	100.00	99.56	98.63
SWT + PCA + GEPSVM [53]	7	10	100.00	99.62	99.02
FRFE + WTT + SVM [54]	12	10	100.00	99.69	98.98
SWT + PCA + SVM + RBF [55]	7	10	100.00	99.69	99.06
DTCWT + VE + TSVM (Proposed)	12	10	100.00	100.00	99.57

(# represents number)

After investigating the results in Table 4, it is clear that 11 out of 13 methods achieve perfect classification (100%) over Dataset66, which stems from its small size. For a larger dataset (Dataset160), only four methods yield perfect classification. They are RT + PCA + LS-SVM [8], DWPT + TE + GEPSVM [10], SWT + PCA + HPA-FNN [11], and the proposed “DTCWT + VE + TSVM”. A common point among the four methods is that they all used advanced feature extraction (RT, DWPT, SWT, and DTCWT) and classification techniques (LS-SVM, GEPSVM, HPA-FNN, and TSVM). This suggests us to learn and apply latest advanced artificial-intelligence and machine-learning approaches to the field of MR brain classification. For the largest dataset (Dataset255), no algorithm achieves perfect classification, because there are relatively various types of diseases. Among all methods, this proposed “DTCWT + VE + TSVM” achieves the highest accuracy of nearly 99.57%, which demonstrates its effectiveness and feasibility.

It is reasonable that all methods achieved high accuracy. Retrospect a similar problem of facial recognition system (FRS), the latest FRS achieved nearly perfect performance and been applied for banking customers [56], vehicle security [57], etc. The pathological detection is simpler than face

detection, because it does not need to identify each subject but identify the status (pathological or healthy). Hence, it is expected that our methods can achieve high classification accuracy.

The difference of accuracy in Table 4 is not significant, however, it is obtained by strict statistical analysis, viz., 10 runs of *K*-fold stratified cross validation. Hence, this slight improvement is reliable and convincing. Even the largest dataset only contains 255 images, so we will try to create a larger dataset that contains more images and more types of diseases.

The proposed CAD system cannot give physical meanings of particular brain regions. Nevertheless, after comparing classifier and human brains, we believe expert systems are similar to declarative memory, while support vector machines are similar to nondeclarative memory. Thus, it is impossible for SVMs (its variants) to give physical meanings. In the future, we may try to use expert systems that can mimic the reasoning process of doctors, but may not give as high accuracies as SVMs.

5.4. Results of Different Runs

The correct classification instance numbers, together with their accuracies, are listed in Table 5. In the table, each row lists the results of different runs, and each column lists the results of different folds. The last row averages the results, and the last column summarizes the results over different folds.

Table 5. Detailed Results of DTCWT + VE + TSVM over Dataset255.

Run	F1	F2	F3	F4	F5	Total
1	51(100.00%)	51(100.00%)	51(100.00%)	51(100.00%)	51(100.00%)	255(100.00%)
2	51(100.00%)	50(98.04%)	51(100.00%)	51(100.00%)	51(100.00%)	254(99.61%)
3	50(98.04%)	51(100.00%)	51(100.00%)	51(100.00%)	50(98.04%)	253(99.22%)
4	50(98.04%)	50(98.04%)	51(100.00%)	51(100.00%)	51(100.00%)	253(99.22%)
5	51(100.00%)	51(100.00%)	51(100.00%)	51(100.00%)	50(98.04%)	254(99.61%)
6	51(100.00%)	51(100.00%)	51(100.00%)	50(98.04%)	50(98.04%)	253(99.22%)
7	51(100.00%)	51(100.00%)	51(100.00%)	51(100.00%)	50(98.04%)	254(99.61%)
8	51(100.00%)	50(98.04%)	51(100.00%)	51(100.00%)	51(100.00%)	254(99.61%)
9	51(100.00%)	51(100.00%)	51(100.00%)	51(100.00%)	51(100.00%)	255(100.00%)
10	51(100.00%)	50(98.04%)	51(100.00%)	51(100.00%)	51(100.00%)	254(99.61%)
Average						253.9 (99.57%)

(F = Fold, R = Run).

5.5. Computation Time

The computation time of each step of our method was calculated and recorded. The training time was recorded over Dataset255. The results of offline-learning and online-prediction are listed in Tables 6 and 7, respectively.

Table 6. Offline-Learning Computation Time.

Process	Time (second)
DTCWT	8.41
VE	1.81
TSVM Training	0.29

Table 7. Online-Prediction Computation Time.

Process	Time (second)
DTCWT	0.037
VE	0.009
TSVM Test	0.003

The computation time results in Table 6 provides that DTCWT costs 8.41 s, VE costs 1.81 s, and TSVM training costs 0.29 s, in the offline-learning procedure. This is because there are 255 images in the dataset, and the training process need to handle all images. The total time is 10.51 s.

The online-prediction time only deals with one query image, so the computation time reduces sharply. Table 7 provides that DTCWT costs 0.037 s, VE costs 0.009 s, and TSVM costs 0.003 s for one query image. Its total time is 0.049 s. Therefore, the proposed system is feasible in practice.

5.6. Comparison to Human Reported Results

In the final experiment, we invited three senior neuroradiologists who have over ten years of experiences. We scanned 20 subjects (5 healthy and 15 pathological), and we pick five slices from each subject. For the healthy subjects, the five slices were selected randomly. For the pathological subjects, the five slices should contain the lesions by confirmation of all the three senior neuroradiologists.

Afterwards, a double blind test was performed. Four junior neuroradiologists with less than 1 year of experiences were required to predict the status of the brain (either pathological or healthy). Each image was assigned with three minutes. Their diagnosis accuracies were listed in Table 8.

Table 8. Comparison to human reported results.

Neuroradiologist	Accuracy
O1	74%
O2	78%
O3	77%
O4	79%
Our method	96%

(O = Observer).

From Table 8, we see our computer-aided diagnosis method achieves an accuracy of 96%. Compared to Table 4, the performance of our proposed method is affected in real-world scenario.

The reasons are complicated. First, the source dataset is downloaded from Harvard medical school, which is for teaching. Hence, the source dataset itself highlighted the difference between pathological and healthy brains intentionally, and the slice positions were selected with care. Second, images from real hospitals are of poorer quality and of poorer localization of lesions. All these contribute to the worsen performance of our method. After all, an accuracy of 96% in rea-world scenario is good and promising.

Another finding is Table 8 is that the four junior neuroradiologists obtained accuracies of 74%, 78%, 77%, and 79%, respectively. All are below 80%. This validates the power of computer vision and machine learning, since computers have proven to deliver better performance in face recognition, video security, *etc.* Nevertheless, there are other more visible symptoms suggesting that something may be wrong in the brain for neuroradiologists. Therefore, this simple test does not reflect the realistic diagnosis accuracy in real hospitals.

6. Conclusions and Future Research

We proposed a novel CAD system for pathological brain detection (PBD) using DTCWT, VE, and TSVM. The results show that the proposed method yields better results than 12 existing methods in terms of classification accuracy.

Our contributions include three points: (1) We investigate the potential use of dual-tree complex wavelet transform (DTCWT) in MR image classification, and prove DTCWT is effective; (2) We utilize twin support vector machine (TSVM) and prove it is better than canonical SVM and GEPSVM; (3) The proposed system “DTCWT + VE + TSVM” is superior to nineteen state-of-the-art systems.

The limitation of our method is the dataset size is too small. We will try to re-check our methods by creating a large dataset. Another limitation is the dataset cannot reflect real-word scenario, thus we

need to obtain more data from hospitals directly in the future. The third limitation is that our data involves only middle and late stage of diseases; hence, our method performs not so good for MR images with diseases in early stage.

In the future, we will consider to validate our method use real clinical data and use advanced classification methods, such as RBFNN, deep learning, least-square techniques. Besides, we will try to apply our method to remote-sensing related fields or hearing loss detection [58]. The advanced parameter estimation, case-based reasoning [59], and optimization [60] techniques will be carried out in a thorough way. Fuzzy method [61] may be applied to remove outliers in the dataset. Coarse-graining [62] can help extract more efficient entropy that is robust to the noises. Video-on-demand [63,64] services may be applied to help reduce computation resources. Particularly, we shall acquire more datasets and compare our method with human interpretation.

Acknowledgments: This paper was supported by Natural Science Foundation of Jiangsu Province (BK20150983), Open Fund of Key laboratory of symbolic computation and knowledge engineering of ministry of education, Jilin University (93K172016K17), Open Fund of Key Laboratory of Statistical information technology and data mining, State Statistics Bureau, (SDL201608), Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), Nanjing Normal University Research Foundation for Talented Scholars (2013119XGQ0061, 2014119XGQ0080), Open Fund of Guangxi Key Laboratory of Manufacturing System & Advanced Manufacturing Technology (15-140-30-008K), Open Project Program of the State Key Lab of CAD & CG, Zhejiang University (A1616), Fundamental Research Funds for the Central Universities (LGYB201604).

Author Contributions: Shuihua Wang & Yudong Zhang conceived the study. Yudong Zhang designed the model. Ming Yang acquired the data. Siyuan Lu & Jiquan Yang analyzed the data. Zhengchao Dong interpreted the data. Shuihua Wang developed the program. Yudong Zhang wrote the draft. All authors gave critical revisions and approved the submission.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

(A)(BP)(F)(PC)NN	(Artificial) (Back-propagation) (Feed-forward) (Pulse-coupled) neural network
(B)PSO(-MT)	(Binary) Particle Swarm Optimization (-Mutation and TVAC)
(D)(S)W(P)T	(Discrete) (Stationary) wavelet (packet) transform
(k)(F)(LS)(GEP)SVM	(kernel) (Fuzzy) (Least-Squares) (Generalized eigenvalue proximal) Support vector machine
(W)(P)(T)E	(Wavelet) (Packet) (Tsallis) entropy
CAD	Computer-aided diagnosis
CS	Cost-sensitivity
FRFE	Fractional Fourier entropy
HMI	Hu moment invariant
KNN	K-nearest neighbors
MR(I)	Magnetic resonance (imaging)
PCA	Principal Component Analysis
RBF	Radial Basis Function
TVAC	Time-varying Acceleration Coefficients
WTT	Welch's <i>t</i> -test

References

1. Thorsen, F.; Fite, B.; Mahakian, L.M.; Seo, J.W.; Qin, S.P.; Harrison, V.; Johnson, S.; Ingham, E.; Caskey, C.; Sundstrom, T.; *et al.* Multimodal imaging enables early detection and characterization of changes in tumor permeability of brain metastases. *J. Controll. Release* **2013**, *172*, 812–822. [[CrossRef](#)] [[PubMed](#)]
2. Gorji, H.T.; Haddadnia, J. A novel method for early diagnosis of Alzheimer's disease based on pseudo Zernike moment from structural MRI. *Neuroscience* **2015**, *305*, 361–371. [[CrossRef](#)] [[PubMed](#)]
3. Goh, S.; Dong, Z.; Zhang, Y.; DiMauro, S.; Peterson, B.S. Mitochondrial dysfunction as a neurobiological subtype of autism spectrum disorder: Evidence from brain imaging. *JAMA Psychiatry* **2014**, *71*, 665–671. [[CrossRef](#)] [[PubMed](#)]
4. El-Dahshan, E.S.A.; Hosny, T.; Salem, A.B.M. Hybrid intelligent techniques for MRI brain images classification. *Digit. Signal Process.* **2010**, *20*, 433–441. [[CrossRef](#)]

5. Patnaik, L.M.; Chaplot, S.; Jagannathan, N.R. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomed. Signal Process. Control* **2006**, *1*, 86–92.
6. Dong, Z.; Wu, L.; Wang, S.; Zhang, Y. A hybrid method for MRI brain image classification. *Expert Syst. Appl.* **2011**, *38*, 10049–10053.
7. Wu, L. An MR brain images classifier via principal component analysis and kernel support vector machine. *Prog. Electromagn. Res.* **2012**, *130*, 369–388.
8. Das, S.; Chowdhury, M.; Kundu, M.K. Brain MR image classification using multiscale geometric analysis of Ripplet. *Progress Electromagn. Res.-Pier* **2013**, *137*, 1–17. [[CrossRef](#)]
9. El-Dahshan, E.S.A.; Mohsen, H.M.; Revett, K.; Salem, A.B.M. Computer-Aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert Syst. Appl.* **2014**, *41*, 5526–5545. [[CrossRef](#)]
10. Dong, Z.; Ji, G.; Yang, J. 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* **2015**, *17*, 1795–1813.
11. Wang, S.; Dong, Z.; Du, S.; Ji, G.; Yan, J.; Yang, J.; Wang, Q.; Feng, C.; Phillips, P. Feed-Forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. *Int. J. Imaging Syst. Technol.* **2015**, *25*, 153–164. [[CrossRef](#)]
12. Nazir, M.; Wahid, F.; Khan, S.A. A simple and intelligent approach for brain MRI classification. *J. Intell. Fuzzy Syst.* **2015**, *28*, 1127–1135.
13. Sun, P.; Wang, S.; Phillips, P.; Zhang, Y. Pathological brain detection based on wavelet entropy and Hu moment invariants. *Bio-Med. Mater. Eng.* **2015**, *26*, 1283–1290.
14. Mount, N.J.; Abrahart, R.J.; Dawson, C.W.; Ab Ghani, N. The need for operational reasoning in data-driven rating curve prediction of suspended sediment. *Hydrol. Process.* **2012**, *26*, 3982–4000. [[CrossRef](#)]
15. Abrahart, R.J.; Dawson, C.W.; See, L.M.; Mount, N.J.; Shamseldin, A.Y. Discussion of “Evapotranspiration modelling using support vector machines”. *Hydrol. Sci. J.-J. Sci. Hydrol.* **2010**, *55*, 1442–1450. [[CrossRef](#)]
16. Si, Y.; Zhang, Z.S.; Cheng, W.; Yuan, F.C. State detection of explosive welding structure by dual-tree complex wavelet transform based permutation entropy. *Steel Compos. Struct.* **2015**, *19*, 569–583. [[CrossRef](#)]
17. Hamidi, H.; Amirani, M.C.; Arashloo, S.R. Local selected features of dual-tree complex wavelet transform for single sample face recognition. *IET Image Process.* **2015**, *9*, 716–723. [[CrossRef](#)]
18. Murugesan, S.; Tay, D.B.H.; Cooke, I.; Faou, P. Application of dual tree complex wavelet transform in tandem mass spectrometry. *Comput. Biol. Med.* **2015**, *63*, 36–41. [[CrossRef](#)] [[PubMed](#)]
19. Smaldino, P.E. Measures of individual uncertainty for ecological models: Variance and entropy. *Ecol. Model.* **2013**, *254*, 50–53. [[CrossRef](#)]
20. Yang, G.; Zhang, Y.; Yang, J.; Ji, G.; Dong, Z.; Wang, S.; Feng, C.; Wang, Q. Automated classification of brain images using wavelet-energy and biogeography-based optimization. *Multimed. Tools Appl.* **2015**, 1–17. [[CrossRef](#)]
21. Guang-Shuai, Z.; Qiong, W.; Chunmei, F.; Elizabeth, L.; Genlin, J.; Shuihua, W.; Yudong, Z.; Jie, Y. Automated Classification of Brain MR Images using Wavelet-Energy and Support Vector Machines. In Proceedings of the 2015 International Conference on Mechatronics, Electronic, Industrial and Control Engineering, Shenyang, China, 24–26 April 2015; Liu, C., Chang, G., Luo, Z., Eds.; Atlantis Press: Shenyang, China, 2015; pp. 683–686.
22. Carrasco, M.; Lopez, J.; Maldonado, S. A second-order cone programming formulation for nonparallel hyperplane support vector machine. *Expert Syst. Appl.* **2016**, *54*, 95–104. [[CrossRef](#)]
23. Wei, Y.C.; Watada, J.; Pedrycz, W. Design of a qualitative classification model through fuzzy support vector machine with type-2 fuzzy expected regression classifier preset. *IEEE Trans. Electr. Electron. Eng.* **2016**, *11*, 348–356. [[CrossRef](#)]
24. Wu, L.; Zhang, Y. Classification of fruits using computer vision and a multiclass support vector machine. *Sensors* **2012**, *12*, 12489–12505.
25. Johnson, K.A.; Becker, J.A. The Whole Brain Atlas. Available online: <http://www.med.harvard.edu/AANLIB/home.html> (accessed on 1 March 2016).
26. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489. [[CrossRef](#)] [[PubMed](#)]

27. Ng, E.Y.K.; Borovetz, H.S.; Soudah, E.; Sun, Z.H. Numerical Methods and Applications in Biomechanical Modeling. *Comput. Math. Methods Med.* **2013**, *2013*, 727830:1–727830:2. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, Y.; Peng, B.; Liang, Y.-X.; Yang, J.; So, K.; Yuan, T.-F. Image processing methods to elucidate spatial characteristics of retinal microglia after optic nerve transection. *Sci. Rep.* **2016**, *6*, 21816. [[CrossRef](#)] [[PubMed](#)]
29. Shin, D.K.; Moon, Y.S. Super-Resolution image reconstruction using wavelet based patch and discrete wavelet transform. *J. Signal. Process. Syst. Signal Image Video Technol.* **2015**, *81*, 71–81. [[CrossRef](#)]
30. Yu, D.; Shui, H.; Gen, L.; Zheng, C. Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging. *IEEJ Trans. Electr. Electron. Eng.* **2015**, *10*, 116–117.
31. Beura, S.; Majhi, B.; Dash, R. Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing* **2015**, *154*, 1–14. [[CrossRef](#)]
32. Ayatollahi, F.; Raie, A.A.; Hajati, F. Expression-Invariant face recognition using depth and intensity dual-tree complex wavelet transform features. *J. Electron. Imaging* **2015**, *24*, 13. [[CrossRef](#)]
33. Hill, P.R.; Anantrasirichai, N.; Achim, A.; Al-Mualla, M.E.; Bull, D.R. Undecimated Dual-Tree Complex Wavelet Transforms. *Signal Process-Image Commun.* **2015**, *35*, 61–70. [[CrossRef](#)]
34. Kadir, M.; Djebbouri, M.; Carré, P. Magnitude-Phase of the dual-tree quaternionic wavelet transform for multispectral satellite image denoising. *EURASIP J. Image Video Process.* **2014**, *2014*, 1–16. [[CrossRef](#)]
35. Singh, H.; Kaur, L.; Singh, K. Fractional M-band dual tree complex wavelet transform for digital watermarking. *Sadhana-Acad. Proc. Eng. Sci.* **2014**, *39*, 345–361. [[CrossRef](#)]
36. Celik, T.; Tjahjadi, T. Multiscale texture classification using dual-tree complex wavelet transform. *Pattern Recognit. Lett.* **2009**, *30*, 331–339. [[CrossRef](#)]
37. Zhang, Y.; Wang, S.; Dong, Z.; Phillips, P.; Ji, G.; Yang, J. Pathological brain detection in magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimization. *Progress Electromagn. Res.* **2015**, *152*, 41–58. [[CrossRef](#)]
38. Mangasarian, O.L.; Wild, E.W. Multisurface proximal support vector machine classification via generalized eigenvalues. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 69–74. [[CrossRef](#)] [[PubMed](#)]
39. Khemchandani, R.; Karpatne, A.; Chandra, S. Generalized eigenvalue proximal support vector regressor. *Expert Syst. Appl.* **2011**, *38*, 13136–13142. [[CrossRef](#)]
40. Shao, Y.H.; Deng, N.Y.; Chen, W.J.; Wang, Z. Improved Generalized Eigenvalue Proximal Support Vector Machine. *IEEE Signal Process. Lett.* **2013**, *20*, 213–216. [[CrossRef](#)]
41. Jayadeva; Khemchandani, R.; Chandra, S. Twin support vector machines for pattern classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 905–910. [[CrossRef](#)] [[PubMed](#)]
42. Nasiri, J.A.; Charkari, N.M.; Mozafari, K. Energy-Based model of least squares twin Support Vector Machines for human action recognition. *Signal Process.* **2014**, *104*, 248–257. [[CrossRef](#)]
43. Xu, Z.J.; Qi, Z.Q.; Zhang, J.Q. Learning with positive and unlabeled examples using biased twin support vector machine. *Neural Comput. Appl.* **2014**, *25*, 1303–1311. [[CrossRef](#)]
44. Shao, Y.H.; Chen, W.J.; Zhang, J.J.; Wang, Z.; Deng, N.Y. An efficient weighted Lagrangian twin support vector machine for imbalanced data classification. *Pattern Recognit.* **2014**, *47*, 3158–3167. [[CrossRef](#)]
45. Zhang, Y.-D.; Wang, S.-H.; Yang, X.-J.; Dong, Z.-C.; Liu, G.; Phillips, P.; Yuan, T.-F. Pathological brain detection in MRI scanning by wavelet packet Tsallis entropy and fuzzy support vector machine. *SpringerPlus* **2015**, *4*, 716. [[CrossRef](#)] [[PubMed](#)]
46. Zhang, Y.-D.; Chen, S.; Wang, S.-H.; Yang, J.-F.; Phillips, P. Magnetic resonance brain image classification based on weighted-type fractional Fourier transform and nonparallel support vector machine. *Int. J. Imaging Syst. Technol.* **2015**, *25*, 317–327. [[CrossRef](#)]
47. Zhang, Y.; Wang, S. Detection of Alzheimer's disease by displacement field and machine learning. *PeerJ* **2015**, *3*. [[CrossRef](#)] [[PubMed](#)]
48. Purushotham, S.; Tripathy, B.K. Evaluation of classifier models using stratified tenfold cross validation techniques. In *Global Trends in Information Systems and Software Applications*; Krishna, P.V., Babu, M.R., Ariwa, E., Eds.; Springer-Verlag Berlin: Berlin, Germany, 2012; Volume 270, pp. 680–690.
49. Ng, E.Y.K.; Jamil, M. Parametric sensitivity analysis of radiofrequency ablation with efficient experimental design. *Int. J. Thermal Sci.* **2014**, *80*, 41–47. [[CrossRef](#)]
50. Kumar, M.A.; Gopal, M. Least squares twin support vector machines for pattern classification. *Expert Syst. Appl.* **2009**, *36*, 7535–7543. [[CrossRef](#)]

51. Zhuang, J.; Widschwendter, M.; Teschendorff, A.E. A comparison of feature selection and classification methods in DNA methylation studies using the Illumina Infinium platform. *BMC Bioinform.* **2012**, *13*, 14. [[CrossRef](#)] [[PubMed](#)]
52. Shamsinejadbabki, P.; Saraee, M. A new unsupervised feature selection method for text clustering based on genetic algorithms. *J. Intell. Inf. Syst.* **2012**, *38*, 669–684. [[CrossRef](#)]
53. Dong, Z.; Liu, A.; Wang, S.; Ji, G.; Zhang, Z.; Yang, J. Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine. *J. Med. Imaging Health Inform.* **2015**, *5*, 1395–1403.
54. Yang, X.; Sun, P.; Dong, Z.; Liu, A.; Yuan, T.-F. Pathological Brain Detection by a Novel Image Feature—Fractional Fourier Entropy. *Entropy* **2015**, *17*, 7877.
55. Zhou, X.-X.; Yang, J.-F.; Sheng, H.; Wei, L.; Yan, J.; Sun, P.; Wang, S.-H. Combination of stationary wavelet transform and kernel support vector machines for pathological brain detection. *Simulation* **2016**. [[CrossRef](#)]
56. Sun, N.; Morris, J.G.; Xu, J.; Zhu, X.; Xie, M. iCARE: A framework for big data-based banking customer analytics. *IBM J. Res. Dev.* **2014**, *58*, 9. [[CrossRef](#)]
57. Rajeshwari, J.; Karibasappa, K.; Gopalakrishna, M.T. Three phase security system for vehicles using face recognition on distributed systems. In *Information Systems Design and Intelligent Applications*; Satapathy, S.C., Mandal, J.K., Udgata, S.K., Bhateja, V., Eds.; Springer-Verlag Berlin: Berlin, Germany, 2016; Volume 435, pp. 563–571.
58. Wang, S.; Yang, M.; Zhang, Y.; Li, J.; Zou, L.; Lu, S.; Liu, B.; Yang, J.; Zhang, Y. Detection of Left-Sided and Right-Sided Hearing Loss via Fractional Fourier Transform. *Entropy* **2016**, *18*, 194. [[CrossRef](#)]
59. Shubati, A.; Dawson, C.W.; Dawson, R. Artefact generation in second life with case-based reasoning. *Softw. Qual. J.* **2011**, *19*, 431–446. [[CrossRef](#)]
60. Zhang, Y.; Yang, X.; Cattani, C.; Rao, R.; Wang, S.; Phillips, P. Tea Category Identification Using a Novel Fractional Fourier Entropy and Jaya Algorithm. *Entropy* **2016**, *18*, 77. [[CrossRef](#)]
61. Ji, L.Z.; Li, P.; Li, K.; Wang, X.P.; Liu, C.C. Analysis of short-term heart rate and diastolic period variability using a refined fuzzy entropy method. *Biomed. Eng. Online* **2015**, *14*, 13. [[CrossRef](#)] [[PubMed](#)]
62. Li, P.; Liu, C.Y.; Li, K.; Zheng, D.C.; Liu, C.C.; Hou, Y.L. Assessing the complexity of short-term heartbeat interval series by distribution entropy. *Med. Biol. Eng. Comput.* **2015**, *53*, 77–87. [[CrossRef](#)] [[PubMed](#)]
63. Lau, P.Y.; Park, S. A new framework for managing video-on-demand servers: Quad-Tier hybrid architecture. *IEICE Electron. Express* **2011**, *8*, 1399–1405. [[CrossRef](#)]
64. Lau, P.Y.; Park, S.; Lee, J. Cohort-Surrogate-Associate: A server-subscriber load sharing model for video-on-demand services. *Malayas. J. Comput. Sci.* **2011**, *24*, 1–16.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).