

AN MR BRAIN IMAGES CLASSIFIER VIA PRINCIPAL COMPONENT ANALYSIS AND KERNEL SUPPORT VECTOR MACHINE

Y. Zhang* and L. Wu

School of Information Science and Engineering, Southeast University,
Nanjing, China

Abstract—Automated and accurate classification of MR brain images is extremely important for medical analysis and interpretation. Over the last decade numerous methods have already been proposed. In this paper, we presented a novel method to classify a given MR brain image as normal or abnormal. The proposed method first employed wavelet transform to extract features from images, followed by applying principle component analysis (PCA) to reduce the dimensions of features. The reduced features were submitted to a kernel support vector machine (KSVM). The strategy of K -fold stratified cross validation was used to enhance generalization of KSVM. We chose seven common brain diseases (glioma, meningioma, Alzheimer's disease, Alzheimer's disease plus visual agnosia, Pick's disease, sarcoma, and Huntington's disease) as abnormal brains, and collected 160 MR brain images (20 normal and 140 abnormal) from Harvard Medical School website. We performed our proposed methods with four different kernels, and found that the GRB kernel achieves the highest classification accuracy as 99.38%. The LIN, HPOL, and IPOL kernel achieves 95%, 96.88%, and 98.12%, respectively. We also compared our method to those from literatures in the last decade, and the results showed our DWT+PCA+KSVM with GRB kernel still achieved the best accurate classification results. The averaged processing time for a 256×256 size image on a laptop of P4 IBM with 3 GHz processor and 2 GB RAM is 0.0448 s. From the experimental data, our method was effective and rapid. It could be applied to the field of MR brain image classification and can assist the doctors to diagnose where a patient is normal or abnormal to certain degrees.

Received 14 June 2012, Accepted 23 July 2012, Scheduled 19 August 2012

* Corresponding author: Yudong Zhang (zhangyudongnuaa@gmail.com).

1. INTRODUCTION

Magnetic resonance imaging (MRI) is an imaging technique that produces high quality images of the anatomical structures of the human body, especially in the brain, and provides rich information for clinical diagnosis and biomedical research [1–5]. The diagnostic values of MRI are greatly magnified by the automated and accurate classification of the MRI images [6–8].

Wavelet transform is an effective tool for feature extraction from MR brain images, because it allows analysis of images at various levels of resolution due to its multi-resolution analytic property. However, this technique requires large storage and is computationally expensive [9]. In order to reduce the feature vector dimensions and increase the discriminative power, the principal component analysis (PCA) was used [10]. PCA is appealing since it effectively reduces the dimensionality of the data and therefore reduces the computational cost of analyzing new data [11]. Then, the problem of how to classify on the input data arises.

In recent years, researchers have proposed a lot of approaches for this goal, which fall into two categories. One category is supervised classification, including support vector machine (SVM) [12] and k -nearest neighbors (k -NN) [13]. The other category is unsupervised classification [14], including self-organization feature map (SOFM) [12] and fuzzy c -means [15]. While all these methods achieved good results, and yet the supervised classifier performs better than unsupervised classifier in terms of classification accuracy (success classification rate). However, the classification accuracies of most existing methods were lower than 95%, so the goal of this paper is to find a more accurate method.

Among supervised classification methods, the SVMs are state-of-the-art classification methods based on machine learning theory [16–18]. Compared with other methods such as artificial neural network, decision tree, and Bayesian network, SVMs have significant advantages of high accuracy, elegant mathematical tractability, and direct geometric interpretation. Besides, it does not need a large number of training samples to avoid overfitting [19].

Original SVMs are linear classifiers. In this paper, we introduced the kernel SVMs (KSVMs), which extends original linear SVMs to nonlinear SVM classifiers by applying the kernel function to replace the dot product form in the original SVMs [20]. The KSVMs allow us to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space high dimensional; thus though the classifier is a hyperplane in the high-

dimensional feature space, it may be nonlinear in the original input space [21].

The structure of the rest of this paper is organized as follows. Next Section 2 gives the detailed procedures of preprocessing, including the discrete wavelet transform (DWT) and principle component analysis (PCA). Section 3 first introduces the motivation and principles of linear SVM, and then turns to the kernel SVM. Section 4 introduces the K -fold cross validation, protecting the classifier from overfitting. Experiments in Section 5 use totally 160 images as the dataset, showing the results of feature extraction and reduction. Afterwards, we compare our method with different kernels to the latest methods in the decade. Final Section 6 is devoted to conclusions and discussions.

2. PREPROCESSING

In total, our method consists of three stages:

- Step 1. Preprocessing (including feature extraction and feature reduction);
- Step 2. Training the kernel SVM;
- Step 3. Submit new MRI brains to the trained kernel SVM, and output the prediction.

As shown in Fig. 1, this flowchart is a canonical and standard classification method which has already been proven as the best classification method [22]. We will explain the detailed procedures of the preprocessing in the following subsections.

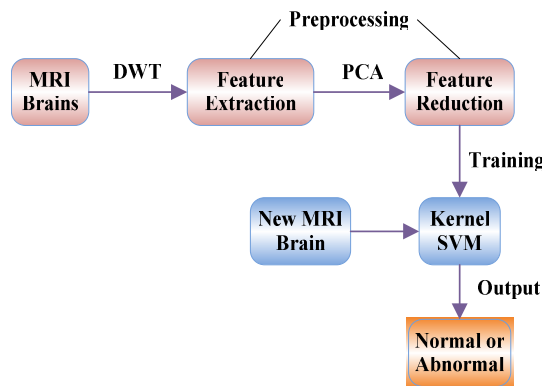


Figure 1. Methodology of our proposed algorithm.

2.1. Feature Extraction

The most conventional tool of signal analysis is Fourier transform (FT), which breaks down a time domain signal into constituent sinusoids of different frequencies, thus, transforming the signal from time domain to frequency domain. However, FT has a serious drawback as discarding the time information of the signal. For example, analyst can not tell when a particular event took place from a Fourier spectrum. Thus, the quality of the classification decreases as time information is lost.

Gabor adapted the FT to analyze only a small section of the signal at a time. The technique is called windowing or short time Fourier transform (STFT) [23]. It adds a window of particular shape to the signal. STFT can be regarded as a compromise between the time information and frequency information. It provides some information about both time and frequency domain. However, the precision of the information is limited by the size of the window.

Wavelet transform (WT) represents the next logical step: a windowing technique with variable size. Thus, it preserves both time and frequency information of the signal. The development of signal analysis is shown in Fig. 2.

Another advantage of WT is that it adopts “scale” instead of traditional “frequency”, namely, it does not produce a time-frequency view but a time-scale view of the signal. The time-scale view is a different way to view data, but it is a more natural and powerful way, because compared to “frequency”, “scale” is commonly used in daily life. Meanwhile, “in large/small scale” is easily understood than “in high/low frequency”.

2.2. Discrete Wavelet Transform

The discrete wavelet transform (DWT) is a powerful implementation of the WT using the dyadic scales and positions [24]. The fundamentals

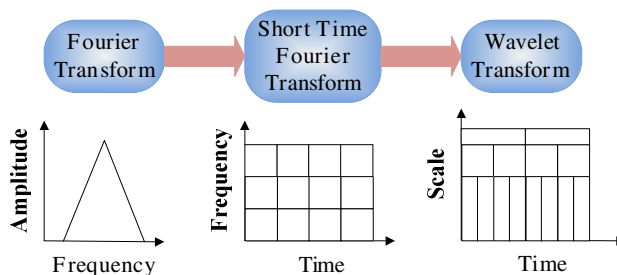


Figure 2. The development of signal analysis.

of DWT are introduced as follows. Suppose $x(t)$ is a square-integrable function, then the continuous WT of $x(t)$ relative to a given wavelet $\psi(t)$ is defined as

$$W_\psi(a, b) = \int_{-\infty}^{\infty} x(t) \psi_{a,b}(t) dt \quad (1)$$

where

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-a}{b}\right) \quad (2)$$

Here, the wavelet $\psi_{a,b}(t)$ is calculated from the mother wavelet $\psi(t)$ by translation and dilation: a is the dilation factor and b the translation parameter (both real positive numbers). There are several different kinds of wavelets which have gained popularity throughout the development of wavelet analysis. The most important wavelet is the Harr wavelet, which is the simplest one and often the preferred wavelet in a lot of applications [25–27].

Equation (1) can be discretized by restraining a and b to a discrete lattice ($a = 2^b$ & $a > 0$) to give the DWT, which can be expressed as follows.

$$\begin{aligned} ca_{j,k}(n) &= DS \left[\sum_n x(n) g_j^*(n - 2^j k) \right] \\ cd_{j,k}(n) &= DS \left[\sum_n x(n) h_j^*(n - 2^j k) \right] \end{aligned} \quad (3)$$

Here $ca_{j,k}$ and $cd_{j,k}$ refer to the coefficients of the approximation components and the detail components, respectively. $g(n)$ and $h(n)$ denote for the low-pass filter and high-pass filter, respectively. j and k represent the wavelet scale and translation factors, respectively. DS operator means the downsampling. Equation (3) is the fundamental of wavelet decomposes. It decomposes signal $x(n)$ into two signals, the approximation coefficients $ca(n)$ and the detail components $cd(n)$. This procedure is called one-level decompose.

The above decomposition process can be iterated with successive approximations being decomposed in turn, so that one signal is broken down into various levels of resolution. The whole process is called wavelet decomposition tree, shown in Fig. 3.

2.3. 2D DWT

In case of 2D images, the DWT is applied to each dimension separately. Fig. 4 illustrates the schematic diagram of 2D DWT. As a result, there are 4 sub-band (LL, LH, HH, and HL) images at each scale. The sub-band LL is used for next 2D DWT.

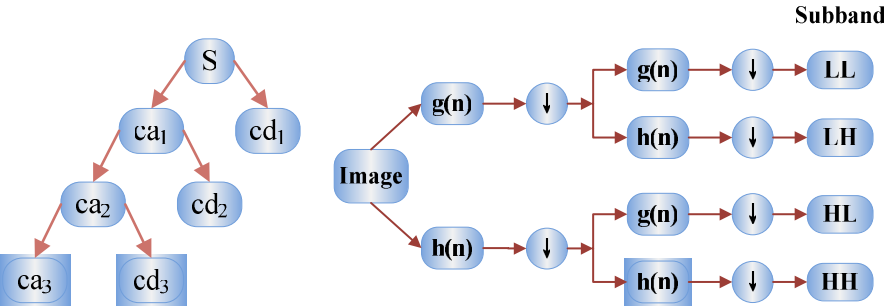


Figure 3. A 3-level wavelet decomposition tree. **Figure 4.** Schematic diagram of 2D DWT.

The LL subband can be regarded as the approximation component of the image, while the LH, HL, and HH subbands can be regarded as the detailed components of the image. As the level of decomposition increased, compacter but coarser approximation component was obtained. Thus, wavelets provide a simple hierarchical framework for interpreting the image information. In our algorithm, level-3 decomposition via Harr wavelet was utilized to extract features.

The border distortion is a technique issue related to digital filter which is commonly used in the DWT. As we filter the image, the mask will extend beyond the image at the edges, so the solution is to pad the pixels outside the images. In our algorithm, symmetric padding method [28] was utilized to calculate the boundary value.

2.4. Feature Reduction

Excessive features increase computation times and storage memory. Furthermore, they sometimes make classification more complicated, which is called the curse of dimensionality. It is required to reduce the number of features.

PCA is an efficient tool to reduce the dimension of a data set consisting of a large number of interrelated variables while retaining most of the variations. It is achieved by transforming the data set to a new set of ordered variables according to their variances or importance. This technique has three effects: it orthogonalizes the components of the input vectors so that uncorrelated with each other, it orders the resulting orthogonal components so that those with the largest variation come first, and eliminates those components contributing the least to the variation in the data set.

It should be noted that the input vectors be normalized to

have zero mean and unity variance before performing PCA. The normalization is a standard procedure. Details about PCA could be seen in Ref. [10].

3. KERNEL SVM

The introduction of support vector machine (SVM) is a landmark in the field of machine learning. The advantages of SVMs include high accuracy, elegant mathematical tractability, and direct geometric interpretation [29]. Recently, multiple improved SVMs have grown rapidly, among which the kernel SVMs are the most popular and effective. Kernel SVMs have the following advantages [30]: (1) work very well in practice and have been remarkably successful in such diverse fields as natural language categorization, bioinformatics and computer vision; (2) have few tunable parameters; and (3) training often involves convex quadratic optimization [31]. Hence, solutions are global and usually unique, thus avoiding the convergence to local minima exhibited by other statistical learning systems, such as neural networks.

3.1. Motivation

Suppose some prescribed data points each belong to one of two classes, and the goal is to classify which class a new data point will be located in. Here a data point is viewed as a p -dimensional vector, and our task is to create a $(p-1)$ -dimensional hyperplane. There are many possible hyperplanes that might classify the data successfully. One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes, since we could expect better behavior in response to unseen data during training, i.e., better generalization performance. Therefore, we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized [32]. Fig. 5 shows the geometric interpolation of linear SVMs, here H1, H2, H3 are three hyperplanes which can classify the two classes successfully, however, H2 and H3 does not have the largest margin, so they will not perform well to new test data. The H1 has the maximum margin to the support vectors (S_{11} , S_{12} , S_{13} , S_{21} , S_{22} , and S_{23}), so it is chosen as the best classification hyperplane [33].

3.2. Principles of Linear SVMs

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

$$\{(x_n, y_n) | x_n \in R^p, y_n \in \{-1, +1\}\}, \quad n = 1, \dots, N \quad (4)$$

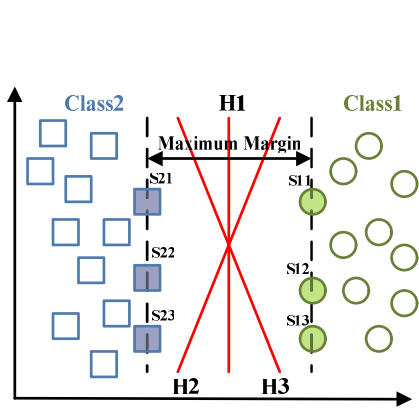


Figure 5. The geometric interpolation of linear SVMs (H denotes for the hyperplane, S denotes for the support vector).

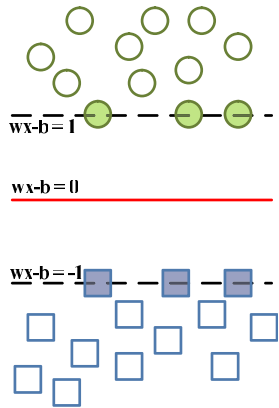


Figure 6. The concept of parallel hyperplanes (w denotes the weight, and b denotes the bias).

where y_n is either -1 or 1 corresponds to the class 1 or 2. Each x_n is a p -dimensional vector. The maximum-margin hyperplane which divides class 1 from class 2 is the support vector machine we want. Considering that any hyperplane can be written in the form of

$$\mathbf{w} \cdot \mathbf{x} - b = 0 \tag{5}$$

where $\cdot$ denotes the dot product and $\mathbf{W}$ the normal vector to the hyperplane. We want to choose the $\mathbf{W}$ and b to maximize the margin between the two parallel (as shown in Fig. 6) hyperplanes as large as possible while still separating the data. So we define the two parallel hyperplanes by the equations as

$$\mathbf{w} \cdot \mathbf{x} - b = \pm 1 \tag{6}$$

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

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \|\mathbf{w}\| \\ \text{s.t.} \quad & y_n (\mathbf{w} \cdot x_n - b) \geq 1, \quad n = 1, \dots, N \end{aligned} \tag{7}$$

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

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & y_n (\mathbf{w} \cdot x_n - b) \geq 1, \quad n = 1, \dots, N \end{aligned} \tag{8}$$

Table 1. Three common Kernels (HPOL, IPOL, and GRB) with their formula and parameters.

Name	Formula	Parameter
Homogeneous Polynomial (HPOL)	$k(x_i, x_j) = (x_i \cdot x_j)^d$	d
Inhomogeneous Polynomial (IPOL)	$k(x_i, x_j) = (x_i \cdot x_j + 1)^d$	d
Gaussian Radial Basis (GRB)	$k(x_i, x_j) = \exp(-\gamma \ x_i - x_j\ ^2)$	γ

The reason leans upon the fact that $\|\mathbf{w}\|$ is involved in a square root calculation. After it is superseded with formula (8), the solution will not change, but the problem is altered into a quadratic programming optimization that is easy to solve by using Lagrange multipliers [34] and standard quadratic programming techniques and programs [35, 36].

3.3. Kernel SVMs

Traditional SMVs constructed a hyperplane to classify data, so they cannot deal with classification problem of which the different types of data located at different sides of a hypersurface, the kernel strategy is applied to SVMs [37]. The resulting algorithm is formally similar, except that every dot product is replaced by a nonlinear kernel function. The kernel is related to the transform $\varphi(x_i)$ by the equation $k(x_i, x_j) = \varphi(x_i)\varphi(x_j)$. The value w is also in the transformed space, with $w = \sum_i \alpha_i \gamma_i \varphi(x_i)$. Dot products with w for classification can be computed by $w \cdot \varphi(x) = \sum_i \alpha_i \gamma_i k(x_i, x)$. In another point of view, the KSVMs allow to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space higher dimensional; thus though the classifier is a hyperplane in the higher-dimensional feature space, it may be nonlinear in the original input space. Three common kernels [38] are listed in Table 1. For each kernel, there should be at least one adjusting parameter so as to make the kernel flexible and tailor itself to practical data.

4. K-FOLD STRATIFIED CROSS VALIDATION

Since the classifier is trained by a given dataset, so it may achieve high classification accuracy only for this training dataset not yet other independent datasets. To avoid this overfitting, we need to integrate cross validation into our method. Cross validation will not increase the final classification accuracy, but it will make the classifier reliable and can be generalized to other independent datasets.

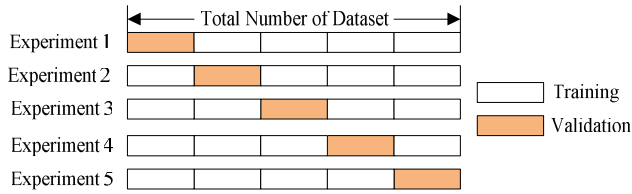


Figure 7. A 5-fold cross validation.

Cross validation methods consist of three types: Random subsampling, K -fold cross validation, and leave-one-out validation. The K -fold cross validation is applied due to its properties as simple, easy, and using all data for training and validation. The mechanism is to create a K -fold partition of the whole dataset, repeat K times to use $K - 1$ folds for training and a left fold for validation, and finally average the error rates of K experiments. The schematic diagram of 5-fold cross validation is shown in Fig. 7.

The K folds can be purely randomly partitioned, however, some folds may have a quite different distributions from other folds. Therefore, stratified K -fold cross validation was employed, where every fold has nearly the same class distributions [39]. Another challenge is to determine the number of folds. If K is set too large, the bias of the true error rate estimator will be small, but the variance of the estimator will be large and the computation will be time-consuming. Alternatively, if K is set too small, the computation time will decrease, the variance of the estimator will be small, but the bias of the estimator will be large [40]. In this study, we empirically determined K as 5 through the trial-and-error method, which means that we suppose parameter K varying from 3 to 10 with increasing step as 1, and then we train the SVM by each value. Finally we select the optimal K value corresponding to the highest classification accuracy.

5. EXPERIMENTS AND DISCUSSIONS

The experiments were carried out on the platform of P4 IBM with 3 GHz processor and 2 GB RAM, running under Windows XP operating system. The algorithm was in-house developed via the wavelet toolbox, the biostatistical toolbox of Matlab 2011b (The Mathworks ©). We downloaded the open SVM toolbox, extended it to Kernel SVM, and applied it to the MR brain images classification. The programs can be run or tested on any computer platforms where Matlab is available.

5.1. Database

The datasets consists of T2-weighted MR brain images in axial plane and 256×256 in-plane resolution, which were downloaded from the website of Harvard Medical School (URL: <http://med.harvard.edu/AANLIB/>), OASIS dataset (URL: <http://www.oasis-brains.org/>), and ADNI dataset (URL: <http://adni.loni.ucla.edu/>). We choose T2 model since T2 images are of higher-contrast and clearer vision compared to T1 and PET modalities.

The abnormal brain MR images of the dataset consist of the following diseases: glioma, meningioma, Alzheimer’s disease, Alzheimer’s disease plus visual agnosia, Pick’s disease, sarcoma, and Huntington’s disease. The samples of each disease are illustrated in Fig. 8.

We randomly selected 20 images for each type of brain. Since there are one type of normal brain and seven types of abnormal brain in the dataset, 160 images are selected consisting of 20 normal and 140 ($= 7$ types of diseases $\times$ 20 images/diseases) abnormal brain images. The setting of the training images and validation images is shown in Table 2 since 5-fold cross validation was used.

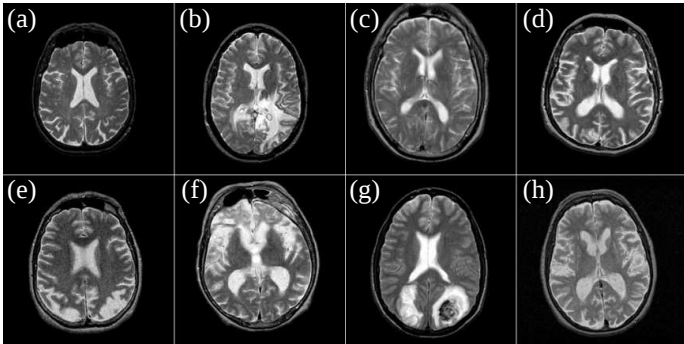


Figure 8. Sample of brain MRIs: (a) normal brain; (b) glioma; (c) meningioma; (d) Alzheimer’s disease; (e) Alzheimer’s disease with visual agnosia; (f) Pick’s disease; (g) sarcoma; (h) Huntington’s disease.

Table 2. Setting of training and validation images (5-fold stratified cross validation).

Total No. of images	Training (128)		Validation (32)	
	Normal	Abnormal	Normal	Abnormal
160	16	112	4	28

5.2. Feature Extraction

The three levels of wavelet decomposition greatly reduce the input image size as shown in Fig. 9. The top left corner of the wavelet coefficients image denotes the approximation coefficients of level-3, whose size is only $32 \times 32 = 1024$.

5.3. Feature Reduction

As stated above, the number of extracted features was reduced from 65536 to 1024. However, it is still too large for calculation. Thus, PCA is used to further reduce the dimensions of features to a higher degree. The curve of cumulative sum of variance versus the number of principle components is shown in Fig. 10.

The variances versus the number of principle components from 1 to 20 are listed in Table 3. It shows that only 19 principle components (bold font in table), which are only 1.86% of the original features, could preserve 95.4% of total variance.

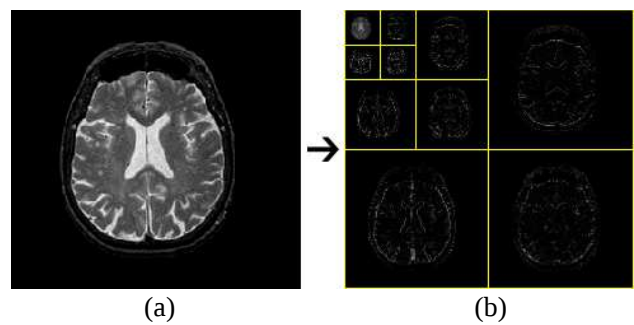


Figure 9. The procedures of 3-level 2D DWT: (a) normal brain MRI; (b) level-3 wavelet coefficients.

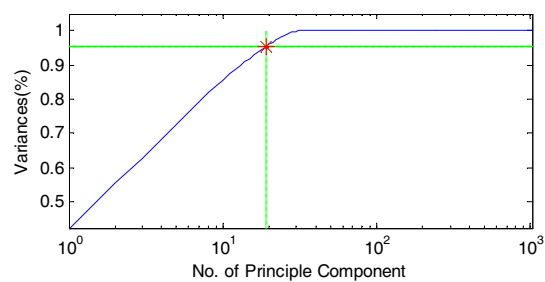


Figure 10. Variances against No. of principle components (x axis is log scale).

Table 3. Detailed data of PCA.

No. of Prin. Comp.	1	2	3	4	5	6	7	8	9	10
Variance (%)	42.3	55.6	62.4	68.1	72.3	76.2	79.3	82.1	84.0	85.6
No. of Prin. Comp.	11	12	13	14	15	16	17	18	19	20
Variance (%)	87.3	88.6	89.8	91.0	92.0	93.0	93.9	94.6	95.4	96.1

Table 4. Confusion matrix of our DWT+PCA+KSVM method (Kernel chose LIN, HPOL, IPOL, and.

LIN	Normal (O)	Abnormal (O)
Normal (T)	17	3
Abnormal (T)	5	135
HPOL	Normal (O)	Abnormal (O)
Normal (T)	19	1
Abnormal (T)	4	136
IPOL	Normal (O)	Abnormal (O)
Normal (T)	18	2
Abnormal (T)	1	139
GRB	Normal (O)	Abnormal (O)
Normal (T)	20	0
Abnormal (T)	1	139

(O denotes for output, T denotes for Target)

5.4. Classification Accuracy

We tested four SVMs with different kernels (LIN, HPOL, IPOL, and GRB). In the case of using linear kernel, the KSVM degrades to original linear SVM.

We computed hundreds of simulations in order to estimate the optimal parameters of the kernel functions, such as the order d in HPOL and IPOL kernel, and the scaling factor γ in GRB kernel. The confusion matrices of our methods are listed in Table 4. The element of i th row and j th column represents the classification accuracy belonging to class i are assigned to class j after the supervised classification.

The results showed that the proposed DWT+PCA+KSVM method obtains quite excellent results on both training and validation images. For LIN kernel, the whole classification accuracy was (17 +

Table 5. Classification accuracy comparison of 10 different algorithms for the same MRI dataset and same number of images.

Approach from literatures	Classification Accuracy (%)
DWT+SOM [12]	94
DWT+SVM with linear kernel [12]	96
DWT+SVM with RBF based kernel [12]	98
DWT+PCA+ANN [41]	97
DWT+PCA+kNN [41]	98
DWT+PCA+ACPSO+FNN [25]	98.75
Approach from this paper	Classification Accuracy (%)
DWT+PCA+KSVM (LIN)	95%
DWT+PCA+KSVM (HPOL)	96.88%
DWT+PCA+KSVM (IPOL)	98.12%
DWT+PCA+KSVM (GRB)	99.38%

135)/160 = 95%; for HPOL kernel, was $(19 + 136)/160 = 96.88\%$; for IPOL kernel, was $(18 + 139)/160 = 98.12\%$; and for the GRB kernel, was $(20 + 139)/160 = 99.38\%$. Obviously, the GRB kernel SVM outperformed the other three kernel SVMs.

Moreover, we compared our method with six popular methods (DWT+SOM [12], DWT+SVM with linear kernel [12], DWT+SVM with RBF based kernel [12], DWT+PCA+ANN [41], DWT+PCA+kNN [41], and DWT+PCA+ACPSO+FNN [25]) described in the recent literature using the same MRI datasets and same number of images. The comparison results were shown in Table 5. It indicates that our proposed method DWT+PCA+KSVM with GRB kernel performed best among the 10 methods, achieving the best classification accuracy as 99.38%. The next is DWT+PCA+ACPSO+FNN method [25] with 98.75% classification accuracy. The third is our proposed DWT+PCA+KSVM with IPOL kernel with 98.12% classification accuracy.

5.5. Time Analysis

Computation time is another important factor to evaluate the classifier. The time for SVM training was not considered, since the parameters of the SVM keep unchanged after training. We sent all the 160 images into the classifier, recorded corresponding computation time, computed the average value, depicted consumed time of different stages shown in Fig. 11.

For each 256×256 image, the averaged computation time on

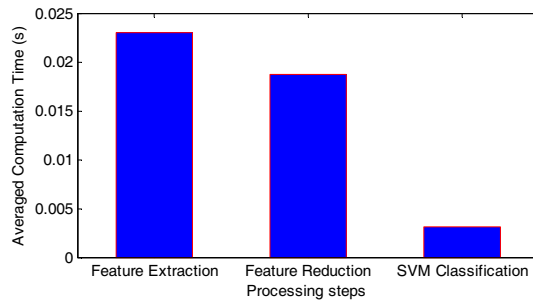


Figure 11. Computation times at different stages.

feature extraction, feature reduction, and SVM classification is 0.023 s, 0.0187 s, and 0.0031 s, respectively. The feature extraction stage is the most time-consuming as 0.023 s. The feature reduction costs 0.0187 s. The SVM classification costs the least time only 0.0031 s.

The total computation time for each 256×256 size image is about 0.0448 s, which is rapid enough for a real time diagnosis.

6. CONCLUSIONS AND DISCUSSIONS

In this study, we have developed a novel DWT+PCA+KSVM method to distinguish between normal and abnormal MRIs of the brain. We picked up four different kernels as LIN, HPOL, IPOL and GRB. The experiments demonstrate that the GRB kernel SVM obtained 99.38% classification accuracy on the 160 MR images, higher than HPOL, IPOL and GRB kernels, and other popular methods in recent literatures.

Future work should focus on the following four aspects: First, the proposed SVM based method could be employed for MR images with other contrast mechanisms such as T1-weighted, Proton Density weighted, and diffusion weighted images. Second, the computation time could be accelerated by using advanced wavelet transforms such as the lift-up wavelet. Third, Multi-classification, which focuses on specific disorders studied using brain MRI, can also be explored. Forth, novel kernels will be tested to increase the classification accuracy.

The DWT can efficiently extract the information from original MR images with little loss. The advantage of DWT over Fourier Transforms is the spatial resolution, viz., DWT captures both frequency and location information. In this study we choose the Harr wavelet, although there are other outstanding wavelets such as Daubechies series. We will compare the performance of different families of wavelet in future work. Another research direction lies in the stationary wavelet

transform and the wavelet packet transform.

The importance of PCA is demonstrated in the discussion section. If we omitted the PCA procedures, we meet a huge search space (as shown in Fig. 10 and Table 3, PCA reduced the 1024 dimensional search space to 19 dimensional search space) which will cause heavy computation burden and worsened classification accuracy. There are some other excellent feature transformation methods such as ICA, manifold learning. In the future, we will focus on investigating the performance of these algorithms.

The proposed DWT+PCA+KSVM with GRB kernel method shows superiority to the LIN, HPOL, and IPOL kernels SVMs. The reason is the GRB kernel takes the form of exponential function, which can enlarge the distance between samples to the extent that HPOL can't reach. Therefore, we will apply the GRB kernel to other industrial fields.

There are two different schools of classification. One is *while-box* classification, such as the decision-trees or rule-based models. The readers can extract reasonable rules from this kind of classifiers. For example, a typical decision tree can be interpreted as "If age is less than 15, turn to left node, and then if gender is male, then turn to right node, and ...". Therefore, the white-box classifiers make sense to patients.

The other school is *black-box* classification, which means that the classifier is intuitionistic, so the reader cannot extract reasonable rules even the kind of classifiers works better and gets higher classification accuracy than the white-box classifiers. From another point of view, this kind of classifiers is really designed by "artificial intelligence" or "computer intelligence". The computer constructed the classifier using its own intelligence not the human sense.

Our method belongs to the latter one. Our goal is to construct a universal classifier not regarding to the age, gender, brain structure, focus of disease, and the like [42], but merely centering on the classification accuracy and highly robustness. This kind of classifier may need further improvements since the patients may need convincing and irrefutable proof to accept the diagnosis of their diseases.

There are literatures describing wavelet transforms, PCA, and kernel SVMs. The most important contribution of this paper is to propose a method which combines them as a powerful tool for identifying normal MR brain from abnormal MR brain. Meanwhile, we tested four kernels, and find GRB kernel as the most successful one. This technique of brain MRI classification based on PCA and KSVM is a potentially valuable tool to be used in computer assisted clinical diagnosis.

REFERENCES

1. Zhang, Y., L. Wu, and S. Wang, "Magnetic resonance brain image classification by an improved artificial bee colony algorithm," *Progress In Electromagnetics Research*, Vol. 116, 65–79, 2011.
2. Mohsin, S. A., N. M. Sheikh, and U. Saeed, "MRI induced heating of deep brain stimulation leads: Effect of the air-tissue interface," *Progress In Electromagnetics Research*, Vol. 83, 81–91, 2008.
3. Golestanirad, L., A. P. Izquierdo, S. J. Graham, J. R. Mosig, and C. Pollo, "Effect of realistic modeling of deep brain stimulation on the prediction of volume of activated tissue," *Progress In Electromagnetics Research*, Vol. 126, 1–16, 2012.
4. Mohsin, S. A., "Concentration of the specific absorption rate around deep brain stimulation electrodes during MRI," *Progress In Electromagnetics Research*, Vol. 121, 469–484, 2011.
5. Oikonomou, A., I. S. Karanasiou, and N. K. Uzunoglu, "Phased-array near field radiometry for brain intracranial applications," *Progress In Electromagnetics Research*, Vol. 109, 345–360, 2010.
6. Scapaticci, R., L. Di Donato, I. Catapano, and L. Crocco, "A feasibility study on microwave imaging for brain stroke monitoring," *Progress In Electromagnetics Research B*, Vol. 40, 305–324, 2012.
7. Asimakis, N. P., I. S. Karanasiou, P. K. Gkonis, and N. K. Uzunoglu, "Theoretical analysis of a passive acoustic brain monitoring system," *Progress In Electromagnetics Research B*, Vol. 23, 165–180, 2010.
8. Chaturvedi, C. M., V. P. Singh, P. Singh, P. Basu, M. Singaravel, R. K. Shukla, A. Dhawan, A. K. Pati, R. K. Gangwar, and S. P. Singh, "2.45 GHz (CW) microwave irradiation alters circadian organization, spatial memory, DNA structure in the brain cells and blood cell counts of male mice, *mus musculus*," *Progress In Electromagnetics Research B*, Vol. 29, 23–42, 2011.
9. Emin Tagluk, M., M. Akin, and N. Sezgin, "Classification of sleep apnea by using wavelet transform and artificial neural networks," *Expert Systems with Applications*, Vol. 37, No. 2, 1600–1607, 2010.
10. Zhang, Y., L. Wu, and G. Wei, "A new classifier for polarimetric SAR images," *Progress in Electromagnetics Research*, Vol. 94, 83–104, 2009.
11. Camacho, J., J. Picó, and A. Ferrer, "Corrigendum to 'The best approaches in the on-line monitoring of batch processes based on PCA: Does the modelling structure matter?' [Anal. Chim. Acta Volume 642 (2009) 59-68]," *Analytica Chimica Acta*, Vol. 658,

- No. 1, 106–106, 2010.
12. Chaplot, S., L. M. Patnaik, and N. R. Jagannathan, “Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network,” *Biomedical Signal Processing and Control*, Vol. 1, No. 1, 86–92, 2006.
 13. Cocosco, C. A., A. P. Zijdenbos, and A. C. Evans, “A fully automatic and robust brain MRI tissue classification method,” *Medical Image Analysis*, Vol. 7, No. 4, 513–527, 2003.
 14. Zhang, Y. and L. Wu, “Weights optimization of neural network via improved BCO approach,” *Progress In Electromagnetics Research*, Vol. 83, 185–198, 2008.
 15. Yeh, J.-Y. and J. C. Fu, “A hierarchical genetic algorithm for segmentation of multi-spectral human-brain MRI,” *Expert Systems with Applications*, Vol. 34, No. 2, 1285–1295, 2008.
 16. Patil, N. S., et al., “Regression models using pattern search assisted least square support vector machines,” *Chemical Engineering Research and Design*, Vol. 83, No. 8, 1030–1037, 2005.
 17. Wang, F.-F. and Y.-R. Zhang, “The support vector machine for dielectric target detection through a wall,” *Progress In Electromagnetics Research Letters*, Vol. 23, 119–128, 2011.
 18. Xu, Y., Y. Guo, L. Xia, and Y. Wu, “An support vector regression based nonlinear modeling method for Sic mesfet,” *Progress In Electromagnetics Research Letters*, Vol. 2, 103–114, 2008.
 19. Li, D., W. Yang, and S. Wang, “Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine,” *Computers and Electronics in Agriculture*, Vol. 4, No. 2, 274–279, 2010.
 20. Gomes, T. A. F., et al., “Combining meta-learning and search techniques to select parameters for support vector machines,” *Neurocomputing*, Vol. 75, No. 1, 3–13, 2012.
 21. Hable, R., “Asymptotic normality of support vector machine variants and other regularized kernel methods,” *Journal of Multivariate Analysis*, Vol. 106, 92–117, 2012.
 22. Ghosh, A., B. Uma Shankar, and S. K. Meher, “A novel approach to neuro-fuzzy classification,” *Neural Networks*, Vol. 22, No. 1, 100–109, 2009.
 23. Gabor, D., “Theory of communication. Part 1: The analysis of information,” *Journal of the Institution of Electrical Engineers Part III: Radio and Communication Engineering*, Vol. 93, No. 26, 429–441, 1946.
 24. Zhang, Y. and L. Wu, “Crop classification by forward neural

- network with adaptive chaotic particle swarm optimization,” *Sensors*, Vol. 11, No. 5, 4721–4743, 2011.
25. Zhang, Y., S. Wang, and L. Wu, “A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO,” *Progress In Electromagnetics Research*, Vol. 109, 325–343, 2010.
 26. Ala, G., E. Francomano, and F. Viola, “A wavelet operator on the interval in solving Maxwell’s equations,” *Progress In Electromagnetics Research Letters*, Vol. 27, 133–140, 2011.
 27. Iqbal, A. and V. Jeoti, “A novel wavelet-Galerkin method for modeling radio wave propagation in tropospheric ducts,” *Progress In Electromagnetics Research B*, Vol. 36, 35–52, 2012.
 28. Messina, A., “Refinements of damage detection methods based on wavelet analysis of dynamical shapes,” *International Journal of Solids and Structures*, Vol. 45, Nos. 14–15, 4068–4097, 2008.
 29. Martiskainen, P., et al., “Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines,” *Applied Animal Behaviour Science*, Vol. 119, Nos. 1–2, 32–38, 2009.
 30. Bermejo, S., B. Monegal, and J. Cabestany, “Fish age categorization from otolith images using multi-class support vector machines,” *Fisheries Research*, Vol. 84, No. 2, 247–253, 2007.
 31. Muniz, A. M. S., et al., “Comparison among probabilistic neural network, support vector machine and logistic regression for evaluating the effect of subthalamic stimulation in Parkinson disease on ground reaction force during gait,” *Journal of Biomechanics*, Vol. 43, No. 4, 720–726, 2010.
 32. Bishop, C. M., *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., 2006.
 33. Vapnik, V., *The Nature of Statistical Learning Theory*, Springer-Verlag New York, Inc., 1995.
 34. Jeyakumar, V., J. H. Wang, and G. Li, “Lagrange multiplier characterizations of robust best approximations under constraint data uncertainty,” *Journal of Mathematical Analysis and Applications*, Vol. 393, No. 1, 285–297, 2012.
 35. Cucker, F. and S. Smale, “On the mathematical foundations of learning,” *Bulletin of the American Mathematical Society*, Vol. 39, 1–49, 2002.
 36. Poggio, T. and S. Smale, “The mathematics of learning: Dealing with data,” *Notices of the American Mathematical Society (AMS)*,

- Vol. 50, No. 5, 537–544, 2003.
37. Acevedo-Rodríguez, J., et al., “Computational load reduction in decision functions using support vector machines,” *Signal Processing*, Vol. 89, No. 10, 2066–2071, 2009.
 38. Deris, A. M., A. M. Zain, and R. Sallehuddin, “Overview of support vector machine in modeling machining performances,” *Procedia Engineering*, Vol. 24, 308–312, 2011.
 39. May, R. J., H. R. Maier, and G. C. Dandy, “Data splitting for artificial neural networks using SOM-based stratified sampling,” *Neural Networks*, Vol. 23, No. 2, 283–294, 2010.
 40. Armand, S., et al., “Linking clinical measurements and kinematic gait patterns of toe-walking using fuzzy decision trees,” *Gait & Posture*, Vol. 25, No. 3, 475–484, 2007.
 41. El-Dahshan, E.-S. A., T. Hosny, and A.-B. M. Salem, “Hybrid intelligent techniques for MRI brain images classification,” *Digital Signal Processing*, Vol. 20, No. 2, 433–441, 2010.
 42. Evans, A. C., et al., “Brain templates and atlases,” *NeuroImage*, Vol. 62, No. 2, 911–922, 2012.