

## **BRAIN MR IMAGE CLASSIFICATION USING MULTI-SCALE GEOMETRIC ANALYSIS OF RIPPLET**

**Sudeb Das<sup>\*</sup>, Manish Chowdhury, and Malay K. Kundu**

Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata-108, India

**Abstract**—We propose an automatic and accurate technique for classifying normal and abnormal magnetic resonance (MR) images of human brain. Ripplet transform Type-I (RT), an efficient multiscale geometric analysis (MGA) tool for digital images, is used to represent the salient features of the brain MR images. The dimensionality of the image representative feature vector is reduced by principal component analysis (PCA). A computationally less expensive support vector machine (SVM), called least square-SVM (LS-SVM) is used to classify the brain MR images. Extensive experiments were carried out to evaluate the performance of the proposed system. Two benchmark MR image datasets and a new larger dataset were used in the experiments, consisting 66, 160 and 255 images, respectively. The generalization capability of the proposed technique is enhanced by  $5 \times 5$  cross validation procedure. For all the datasets used in the experiments, the proposed system shows high classification accuracies (on an average  $> 99\%$ ). Experimental results and performance comparisons with state-of-the-art techniques, show that the proposed scheme is efficient in brain MR image classification.

### **1. INTRODUCTION**

Magnetic resonance imaging (MRI) is a low-risk, fast, non-invasive imaging technique with no ionizing radiation hazard. MRI provides high quality and high contrast images of anatomical structures as well as functional images of different organs [1]. Soft tissue structures (heart, lungs, liver, brain and other organs) are clearer and more detailed with MRI than other medical imaging modalities. The non-invasive nature of MRI together with its rich information provision,

---

*Received 1 January 2013, Accepted 4 February 2013, Scheduled 6 February 2013*

<sup>\*</sup> Corresponding author: Sudeb Das (to.sudeb@gmail.com).

makes MRI the widely used method for diagnosis and treatment planning [2–4]. Various researchers are not only trying to improve the MR image quality, but also seeking novel methods for easier and quicker inference from the images [5–7]. In recent years, MRI has emerged as one of the popular choice to study the human brain [8–10]. MRI can detect a variety of conditions of the brain such as cysts, tumors, bleeding, swelling, developmental and structural abnormalities, infections, inflammatory conditions, or problems with the blood vessels, etc.. It can determine whether a shunt is working or not, and detect damage to the brain caused by an injury or a stroke.

However, because of the huge amount of imaging data, the existing manual methods of analysis and interpretation of brain images are tedious, time consuming, costly and subject to fatigue of human observer. This necessitates the requirement of developing automated diagnosis tools to draw quicker and easier inferences from the MR images. These automated systems can be of great help for the medical personnel in diagnosis, prognosis, pre-surgical and post-surgical procedures, etc. [11,12]. One of the most distinguishable feature of a normal human brain is the symmetry, which is obvious in the axial and the coronal brain magnetic resonance (MR) images. Whereas, asymmetry in an axial MR brain image strongly indicates abnormality/disorder [11,13]. This symmetry-asymmetry can be modelled by various image and signal processing techniques, which can be used to classify the normal and abnormal brain MR images [6].

In recent years, various approaches of brain MR image classification have been proposed by different researchers. Chaplot et al. have achieved 94% and 98% accuracies through classifiers based on self-organizing map (SOM) and support vector machine (SVM), respectively. They have used discrete wavelet transform (DWT) for feature extraction [13]. In [14], Maitra and Chatterjee have shown that Slantlet transform can be combined with supervised classification (back-propagation neural network (BPNN)) technique to achieve 100% classification accuracy. Principal component analysis (PCA) is used to reduce the dimension of the feature vector obtained through DWT in [15] by El-Dahshan et al. They have achieved 97% and 98% success-rates through feed-forward BPNN and k-nearest neighbor (kNN) classifiers, respectively. Recently, Zhang et al. have proposed several advanced techniques for brain MR image classification with high classification accuracies [12,16–18]. In all the proposed techniques, they have used DWT for feature extraction, PCA for feature dimension reduction and different classifiers with various weight optimization schemes (forward neural network (FNN) + scaled chaotic artificial bee colony (SCABC) in [12], FNN + adaptive chaotic particle swarm

optimization (ACPSO) in [16], BPNN + scaled conjugate gradient (SCG) in [17], kernel SVM (KSVM) with different kernels: linear (LIN), homogeneous polynomial (HPOL), inhomogeneous polynomial (IPOL) and Gaussian radial basis (GRB) in [18]) for segregating the normal and abnormal MR images.

Most of the existing brain MR image classification systems suffer from several shortcomings:

- (i) These systems are based on DWT or variants of DWT, which has several problems: limited directionality, non-supportiveness to anisotropy, etc.. Therefore, DWT cannot capture the subtle and intrinsic details of the brain MR images, which are required for segregating normal and abnormal cases.
- (ii) Although, most of the state-of-the-art schemes utilize PCA for feature reduction to achieve computational efficiency. Because of the DWT + PCA combination, the dimension of the reduced feature vector is still comparatively high.
- (iii) Moreover, most of these schemes use neural network with complex weight optimization techniques, which leads to high computational complexity.
- (iv) Furthermore, most of the existing techniques lack the generalization capability, as these systems work well on small particular dataset, but, fail to work efficiently for different datasets of various sizes with varying classes of diseases.

In general, most of the existing MR brain image classification systems consist of three different phases: feature extraction, feature reduction and classification. The proposed technique consists of similar process blocks.

The main motivation of this work is to develop an automatic MR brain image classification system with less computational complexity and high classification accuracy. The other motivation is to make the technique general so that it can work equally efficiently for different brain MR datasets, consisting varying number of diseases classes. The main advantages of the proposed scheme can be listed as follows:

- (i) A fully automatic and accurate system for MR brain image classification.
- (ii) The proposed scheme is based on RT, which is superior than DWT. RT has superior localization capability in both spatial and frequency domain. Moreover, RT can capture 2D singularities along different curves in any direction, because of the general scaling with arbitrary degree and support, and thus providing better image representation.

- (iii) Applying PCA on the coefficients of RT decomposed image, the feature vector dimension reduces to only 9 (much less than used by the existing schemes) which results in computational efficiency.
- (iv) The use of LS-SVM further reduces the computational cost by avoiding solving quadratic programming problem, and
- (v) The proposed system is more effective than the state-of-the-art techniques: it works efficiently on datasets with different sizes and consisting abnormal brain MR images from more number of disease classes.

All these above mentioned advantages make the proposed technique an effective and accurate system for MR brain image classification.

The rest of the paper is organized as follows. Theoretical backgrounds of RT, feature reduction and LS-SVM is described in Section 2, Section 3 and Section 4, respectively. Section 5, presents the proposed system. Experimental results and comparisons are given in Section 6 and we draw conclusion in Section 7.

## 2. RIPPLET TRANSFORM TYPE-I (RT)

DWT and its variants were used extensively by various researchers for feature extraction in MR brain image classification [12–18]. But the problem with DWT is that it is inherently non-supportive to directionality and anisotropy. To address these problems, Jun Xu et al. proposed a new MGA-tool called RT [19]. RT is a higher dimensional generalization of the Curvelet Transform (CVT), capable of representing images or  $2D$  signals at different scales and different directions. To achieve anisotropic directionality, CVT uses a parabolic scaling law [20]. From the perspective of micro-local analysis, the anisotropic property of CVT guarantees resolving  $2D$  singularities along  $C^2$  curves. Whereas, RT provides a new tight frame with sparse representation for images with discontinuities along  $C^d$  curves [19].

There are two questions regarding the scaling law used in CVT: 1) Is the parabolic scaling law optimal for all types of boundaries? and if not, 2) What scaling law will be optimal? To address these questions, Jun Xu et al. intended to generalize the scaling law, which resulted in RT. RT generalizes CVT by adding two parameters, i.e., support  $c$  and degree  $d$ . CVT is just a special case of RT with  $c = 1$  and  $d = 2$ . The anisotropy capability of representing singularities along arbitrarily shaped curves of RT, is due to these new parameters  $c$  and  $d$ .

As digital image processing needs discrete transform instead of continuous transform, here we describe the discretization of RT [19]. The discretization of continuous RT is based on the discretization of the

parameters of ripple functions. The scale parameter  $a$  is sampled at dyadic intervals. The position parameter  $b$  and the rotation parameter  $\theta$  are sampled at equal-spaced intervals.  $a_j$ ,  $\vec{b}_k$  and  $\theta_l$  substitute  $a$ ,  $\vec{b}$  and  $\theta$ , respectively, and satisfy that  $a_j = 2^{-j}$ ,  $\vec{b}_k = [c \cdot 2^{-j} \cdot k_1, 2^{-j/d} \cdot k_2]^T$  and  $\theta_l = \frac{2\pi}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot l$ , where  $\vec{k} = [k_1, k_2]^T$ , and  $j, k_1, k_2, l \in \mathbb{Z}$ .  $(\cdot)^T$  denotes the transpose of a vector.  $d \in \mathbb{R}$ , since any real number can be approximated by rational numbers, so we can represent  $d$  with  $d = n/m$ ,  $n, m \neq 0 \in \mathbb{Z}$ . Usually, we prefer  $n, m \in \mathbb{N}$  and  $n, m$  are both primes. In the frequency domain, the corresponding frequency response of ripple function is in the form

$$\hat{\rho}_j(r, \omega) = \frac{1}{\sqrt{c}} a^{\frac{m+n}{2n}} W(2^{-j} \cdot r) V\left(\frac{1}{c} \cdot 2^{-\lfloor j \frac{m-n}{n} \rfloor} \cdot \omega - l\right) \quad (1)$$

where,  $W$  and  $V$  are the *radial-window* and the *angular-window*, respectively. These two windows satisfy the following admissibility conditions:

$$\sum_{j=0}^{+\infty} |W(2^{-j} \cdot r)|^2 = 1 \quad (2)$$

$$\sum_{l=-\infty}^{+\infty} \left| V\left(\frac{1}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot \omega - l\right) \right|^2 = 1 \quad (3)$$

given  $c$ ,  $d$  and  $j$ . These two windows partition the polar frequency domain into ‘wedges’. The ‘wedge’ corresponding to the ripple function in the frequency domain is

$$H_{j,l}(r, \theta) = \left\{ 2^j \leq |r| \leq 2^{2j}, \left| \theta - \frac{\pi}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot l \right| \leq \frac{\pi}{2} 2^{-j} \right\} \quad (4)$$

The discrete RT of an  $M \times N$  image  $f(n_1, n_2)$  will be in the form of

$$R_{j,\vec{k},l} = \sum_{n_1=0}^{M-1} \sum_{n_2=0}^{N-1} f(n_1, n_2) \overline{\rho_{j,\vec{k},l}(n_1, n_2)} \quad (5)$$

where,  $R_{j,\vec{k},l}$  are the ripple coefficients.

As a generalized version of CVT, RT is not only capable of resolving 2D singularities, but it also has some useful properties:

- (i) It forms a new tight frame in a function space. Having good capability of localization in both spatial and frequency domain, it provides a more efficient and effective representation for images or 2D signals.
- (ii) It has general scaling with arbitrary degree and support, which can capture 2D singularities along different curves in any directions.

Jun Xu et al. have showed that RT can provide a more effective representation for images with singularities along smooth curves [19]. It outperforms discrete cosine transform and DWT in nonlinear approximation, when the number of retained coefficients is small. RT can achieve roughly 2dB higher PSNR on average than JPEG, and provide better visual quality than JPEG2000 at low bit-rates, when applied to image compression. In case of image denoising application, RT performs better than CVT and DWT. RT produces high quality fused images, when applied in the medical image fusion domain [21]. All these experiments show that RT based image coding is suitable for representing texture or edges in images.

### 3. FEATURE REDUCTION

There are two different feature reduction phases in the proposed scheme. Here, we briefly describe these phases.

Let,  $N \times N$  be the size of an brain MR image  $I$ ,  $N = 2m$ ,  $m \in \mathbb{Z}^+$ . After applying RT on  $I$ , let the size of the decomposed low frequency subband (LFS) be  $M \times M$ ,  $M = 2p$ ,  $p \in \mathbb{Z}^+$ . Therefore, the first feature reduction happens for the transition:  $N \times N \rightarrow M \times M$ ,  $M < N$ .

The second feature reduction is achieved using PCA [22]. Let,  $T_{var}$  be the total variance of the original feature set, and  $S_{var}$  be the total variance of the reduced feature set having dimension  $d$ ,  $d \ll M^2$ . The proper value of  $d$  is selected:  $\frac{S_{var}}{T_{var}} \simeq 0.9$ . Let,  $PER_{red}$  be the achieved feature reduction percentage.

$$\therefore PER_{red} = \frac{N^2 - d}{N^2} \times 100 \% \quad (6)$$

### 4. CLASSIFICATION THROUGH LS-SVM

The main drawback of support vector machine (SVM) is its high computational complexity for large dimensional datasets. To reduce the computational burden, the least square version of SVM (LS-SVM) is adopted as classifier in this paper. Due to equality type constraints in the formulation, the solution follows from solving a set of linear equations, instead of quadratic programming for classical SVM's [23]. Considering a linearly separable binary classification problem:

$$(x_i, y_i)_{i=1}^n \quad \text{and} \quad y_i = \{+1, -1\} \quad (7)$$

where,  $x_i$  is an  $n$ -dimensional vector and  $y_i$  is the label of this vector, LS-SVM can be formulated as the optimization problem:

$$\min_{w,b,e} \mathcal{J}(w, b, e) = \frac{1}{2} w' w + \frac{1}{2} C \sum_{i=1}^n e_i^2 \quad (8)$$

subject to the equality constraint:

$$y_i [w' \varphi(x_i) + b] = 1 - e_i \quad (9)$$

where,  $C > 0$  is a regularization factor,  $b$  a bias term,  $w$  the weight vector,  $e_i$  the difference between the desired output and the actual output, and  $\varphi(x_i)$  a mapping function.

The lagrangian for the problem of Eq. (8) is defined as follows:

$$\mathcal{L}(w, e_i, b, \alpha_i) = \min_{w,b,e} \mathcal{J}(w, b, e) - \sum_{i=1}^n \alpha_i \{y_i [w' \varphi(x_i) + b] - 1 + e_i\} \quad (10)$$

where,  $\alpha_i$  are Lagrange multipliers. The Karush-Kuhn-Tucker (KKT) conditions for optimality  $\frac{\partial \mathcal{L}}{\partial w} = 0 \rightarrow w = \sum_{i=1}^n \alpha_i y_i \varphi(x_i)$ ;  $\frac{\partial \mathcal{L}}{\partial e_i} = 0 \rightarrow \alpha_i = C e_i$ ;  $\frac{\partial \mathcal{L}}{\partial b} = 0 \rightarrow \sum_{i=1}^n \alpha_i y_i = 0$ ;  $\frac{\partial \mathcal{L}}{\partial \alpha_i} = 0 \rightarrow y_i [w' \varphi(x_i) + b] - 1 + e_i = 0$ , is the solution to the following linear system

$$\begin{bmatrix} 0 & -Y \\ Y & \varphi \varphi' + C^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{1} \end{bmatrix} \quad (11)$$

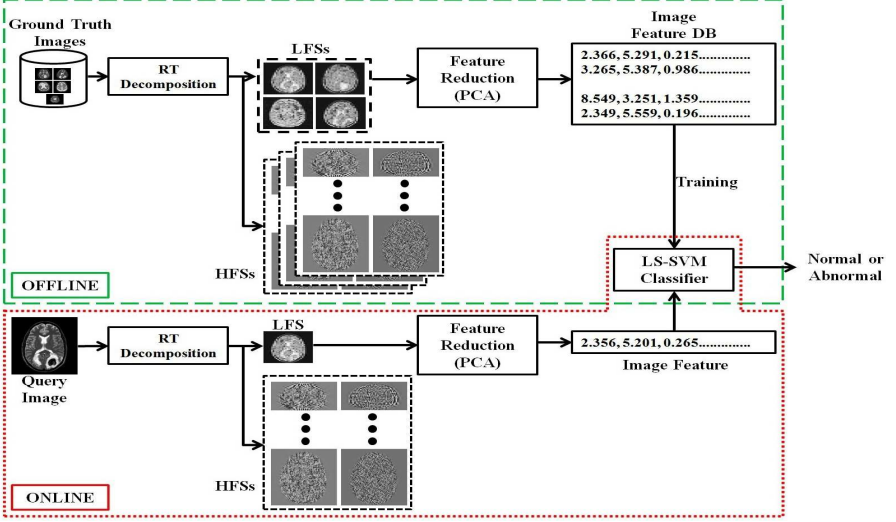
where,  $\varphi = [\varphi(x_1)' y_1, \dots, \varphi(x_n)' y_n]$ ,  $Y = [y_1, \dots, y_n]$ ,  $\bar{1} = [1, \dots, 1]$ , and  $\alpha = [\alpha_1, \dots, \alpha_n]$ .

For a given kernel function  $K(\cdot)$  and a new test sample point  $x$ , the LS-SVM classifier is given by

$$f(x) = \text{sgn} \left[ \sum_{i=1}^n \alpha_i y_i K(x, x_i) + b \right] \quad (12)$$

## 5. PROPOSED SYSTEM

The proposed system consists of two phases as shown in the block diagram of Fig. 1: an offline phase and an online phase. Both the phases, consist of the following steps: feature extraction based on RT, feature reduction through PCA and classification by LS-SVM classifier.



**Figure 1.** Block diagram of the proposed system.

### 5.1. Offline Phase

Let,  $n$  be the number of training images of size  $N \times N$ ,  $N = 2m$ ,  $m \in \mathbb{Z}^+$ . The salient steps of the offline (training) phase are listed next:

**Step 1:** The training images are decomposed by RT to get the low-frequency subbands (LFSs) and high-frequency subbands (HFSs). Only the LFSs (one LFS/image) are used as features. Let,  $M \times M$  ( $M < N$ ) be the size of each LFS,  $M = 2p$ ,  $p \in \mathbb{Z}^+$ . A feature matrix  $X$  of size  $n \times M^2$  is constructed by the coefficients of the LFSs. Each row of  $X$  consists of  $M^2$  coefficients belonging to a particular LFS, representing feature vector of dimension  $M^2$  of that image.

**Step 2:** The dimension of the feature vectors representing the training images is reduced by applying PCA on  $X$  from  $M^2$  to  $d$  (say), where  $d \ll M^2$ , following the criteria mentioned in Section 3.

**Step 3:** The set of reduced feature vectors, along with the class information are used to train a LS-SVM classifier. Cross validation is used for improving the generalization capability of the system.

### 5.2. Online Phase

The online phase of the proposed system consists of the following steps:



**Step 1:** The user (doctors, radiologist, etc.) inputs the brain MR image of size  $N \times N$  to be classified. RT is applied on the input image to get the LFS of size  $M \times M$ .

**Step 2:** The dimension of the feature vector representing the input query image is reduced from  $M^2$  to  $d$  by applying PCA.

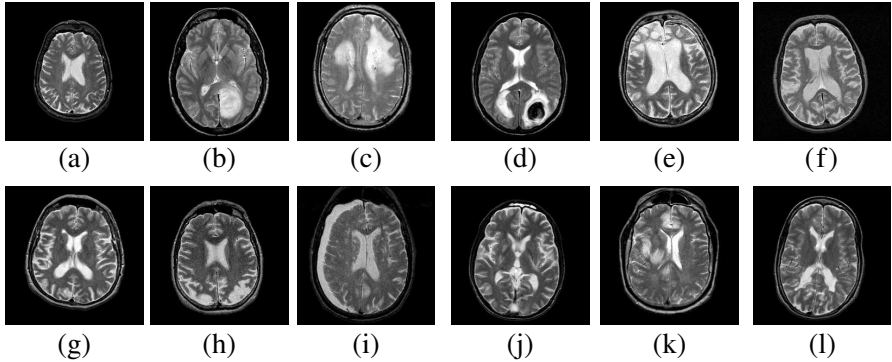
**Step 3:** This reduced feature vector of dimension  $d$  is used as input to the previously trained LS-SVM classifier. The classifier classified the input query image as normal or abnormal.

## 6. EXPERIMENTAL RESULTS AND COMPARISONS

Extensive experiments were carried out to evaluate the performance of the proposed system in brain MR image classification.

### 6.1. Experimental Setup

We implemented the proposed technique in MATLAB, and experiments were carried out on a PC with 2.66 GHz CPU and 4 GB RAM. The decomposition parameter of RT was levels = [1, 2, 4, 4], and we used '9/7' and 'pkva' as the pyramid filter and orientation filter, respectively. Three (3) different MR image datasets were used in the experiments. All the datasets consist of T2-weighted MR brain images in axial plane and  $256 \times 256$  in-plane resolution, which were downloaded from the website of Harvard Medical School (URL: <http://med.harvard.edu/AANLIB/>). The first two datasets are benchmark datasets, widely used in brain MR image classification problem, and consist of abnormal images from 7 types of diseases along with the normal images. The abnormal brain MR images of the benchmark datasets consist images of the following diseases: glioma, meningioma, Alzheimer's disease, Alzheimer's disease plus visual agnosia, Pick's disease, sarcoma and Huntington's disease. The first benchmark dataset (Dataset-66) consists of 66 (18 normal and 48 abnormal) brain MR images. There are in total 160 (20 normal and 140 abnormal) brain MR images in the second benchmark dataset (Dataset-160). The third new larger dataset (Dataset-255) consists of 255 (35 normal and 220 abnormal) brain MR images. Abnormal brain MR images of the third dataset are from 11 types of diseases, among which 7 types of diseases are same as the two benchmark datasets, mentioned before. The third dataset also consists abnormal images of 4 new types of diseases: chronic subdural hematoma, cerebral toxoplasmosis, herpes encephalitis and multiple sclerosis. Each of the 11 types of diseases' consists of 20 abnormal brain MR images. Fig. 2, shows samples of the brain MR images used in the experiments.



**Figure 2.** Sample brain MR images: (a) Normal, (b) glioma, (c) meningioma, (d) sarcoma, (e) Pick's disease, (f) Huntington's disease, (g) Alzheimer's disease, (h) Alzheimer's disease with visual agnosia, (i) chronic subdural hematoma, (j) cerebral toxoplasmosis, (k) herpes encephalitis, (l) multiple sclerosis.

To make the LS-SVM classifier more reliable and generalize to independent datasets,  $5 \times 5$ -fold and  $5 \times 6$ -fold stratified cross validations (CV) were employed.  $5 \times 6$ -fold stratified CV was used for Dataset-66, and for the other two datasets  $5 \times 5$ -fold stratified CV was carried out. For training of the LS-SVM, we used the Radial Basis Function (RBF):  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ ,  $\gamma > 0$ , as the kernel. There are two tunable parameters while using the RBF kernel in LS-SVM classifier:  $C$  and  $\gamma$ . The kernel parameter  $\gamma$  controls the shape of the kernel and regularization parameter  $C$  controls the tradeoff between margin maximization and error minimization. It is not known beforehand which values of  $C$  and  $\gamma$  are the best for the classification problem at hand. Hence, various pairs of  $(C, \gamma)$  were tried with over the course of the CV procedure, and the one with the lowest CV error rate was picked, where  $C \in [1, 10]$  and  $\gamma \in [1, 3]$ . After finding the best values for the parameters  $C$  and  $\gamma$ , these values were used to train the LS-SVM model, and the test set was used to measure the error rate of the classification system. Tables 1–3, show the settings of the training and the validation images for the datasets used in the experiments.

To compare the performance of our proposed method with the state-of-the-art techniques, we implemented several of the existing schemes. Quantitative evaluation of the proposed system and its performance comparison with other state-of-the-art techniques were analyzed using the following statistical measures:

**Table 1.** Setting of one pass of 6-fold stratified CV for Dataset-66.

| Total No. of Images |               | Training (55) |          | Validation (11) |          |
|---------------------|---------------|---------------|----------|-----------------|----------|
| Normal (18)         | Abnormal (48) | Normal        | Abnormal | Normal          | Abnormal |
| 66                  |               | 15            | 40       | 3               | 8        |

**Table 2.** Setting of one pass of 5-fold stratified CV for Dataset-160.

| Total No. of Images |                | Training (128) |          | Validation (32) |          |
|---------------------|----------------|----------------|----------|-----------------|----------|
| Normal (20)         | Abnormal (140) | Normal         | Abnormal | Normal          | Abnormal |
| 160                 |                | 16             | 112      | 4               | 28       |

**Table 3.** Setting of one pass of 5-fold stratified CV for Dataset-255.

| Total No. of Images |                | Training (204) |          | Validation (51) |          |
|---------------------|----------------|----------------|----------|-----------------|----------|
| Normal (35)         | Abnormal (220) | Normal         | Abnormal | Normal          | Abnormal |
| 255                 |                | 28             | 177      | 7               | 43       |

**Sensitivity (true positive fraction):** is the probability that a diagnostic test is positive, given that the person has the disease.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (13)$$

**Specificity (true negative fraction):** is the probability that a diagnostic test is negative, given that the person does not has the disease.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (14)$$

**Accuracy:** is the probability that a diagnostic test is correctly performed.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

where,

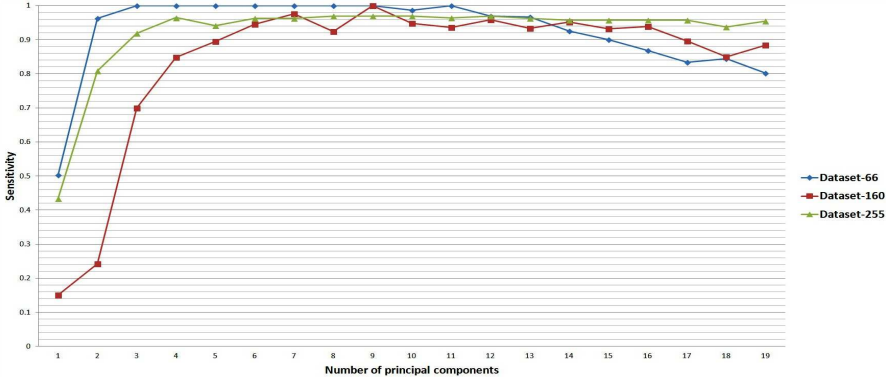
- TP (True Positive) — correctly classified positive cases,
- TN (True Negative) — correctly classified negative cases,
- FP (False Positive) — incorrectly classified negative cases, and
- FN (False Negative) — incorrectly classified positive cases.

## 6.2. Results and Discussions

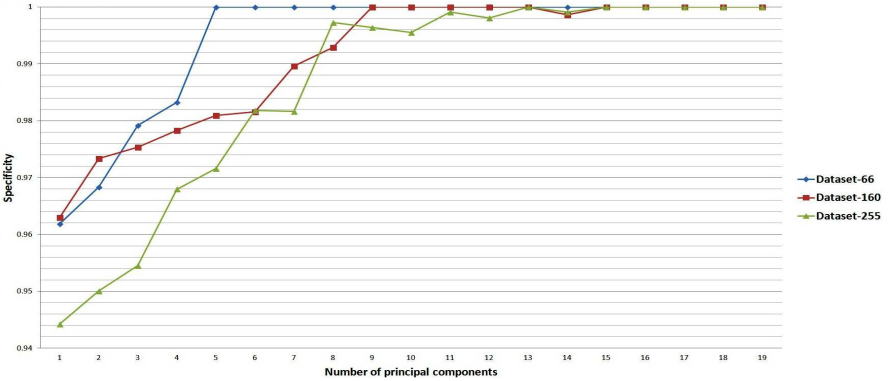
Several different experiments were carried out to evaluate the performance of the proposed system in light of feature reduction

efficiency, classification accuracy, comparisons with other state-of-the-art schemes and computation complexity analysis.

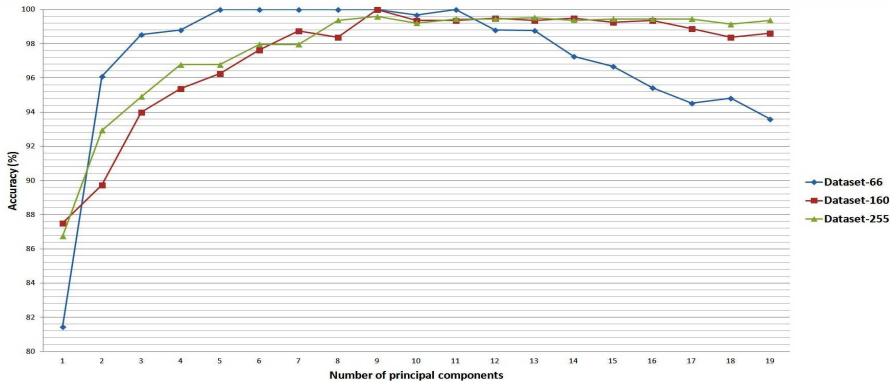
The proposed system is based on the MGA of the LFS coefficients obtained by RT decomposition. With the above mentioned RT decomposition configuration, the size of the LFS is  $32 \times 32$ . PCA is used to reduce the feature vector size to only 9, where these 9 features are the first 9 principal components, preserving  $\simeq 90\%$  of total variance of the RT decomposed features. This reduced feature set is only 0.88% and 0.014% of the initial feature set considering LFS and original image, respectively. Therefore, due to the RT + PCA combination we have achieved 99% feature reduction. The systems proposed in [12, 16–18] have used 19 principal components as the image representing feature vector. We have not only achieved



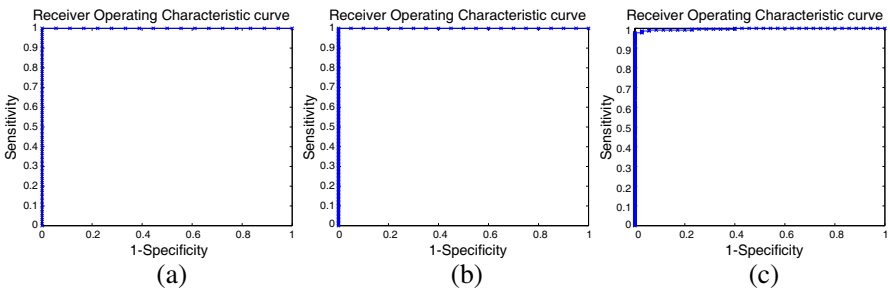
**Figure 3.** Performance evaluation in terms of sensitivity.



**Figure 4.** Performance evaluation in terms of specificity.



**Figure 5.** Performance evaluation in terms of classification accuracy.



**Figure 6.** ROC curves of performance evaluation: (a) Dataset-66, (b) Dataset-160, and (c) Dataset-255.

47.39% feature reduction from the state-of-the-art brain MR image classification techniques, but also higher performance in terms of accuracy. To find out the proper number of principal components, which give the best result, the performance of the proposed system was experimented with different numbers of principal components (1)–(19). The graphs of the Figs. 3, 4, and 5, show the performance of the proposed system in terms of Sensitivity, specificity and classification accuracy for the three datasets used in the experiments with different numbers of principal components, respectively. It is clear from the results given in Figs. 3, 4, and 5, that our proposed system works efficiently for all the three datasets using only 9 principal components for image representation. We have achieved the best results for all the statistical measures used to evaluate the performance of the proposed system, considering only 9 principal components: Sensitivity (1.00, 1.00, 0.97), specificity (1.00, 1.00, 0.99) and classification accuracy

**Table 4.** Performance comparison using the three datasets.

| Scheme                        | Feature<br>dimension | Accuracy (%)  |               |              |
|-------------------------------|----------------------|---------------|---------------|--------------|
|                               |                      | Dataset-66    | Dataset-160   | Dataset-255  |
| DWT + SOM [13]                | 4761                 | 94.00         | 93.17         | 91.65        |
| DWT + SVM + LIN [13]          | 4761                 | 96.15         | 95.38         | 94.05        |
| DWT + SVM + POLY [13]         | 4761                 | 98.00         | 97.15         | 96.37        |
| DWT + SVM + RBF [13]          | 4761                 | 98.00         | 97.33         | 96.18        |
| DWT + PCA + FNN [15]          | 7                    | 97.00         | 96.98         | 95.29        |
| DWT + PCA + kNN [15]          | 7                    | 98.00         | 97.54         | 96.79        |
| DWT + PCA + FNN + ACP SO [16] | 19                   | 100.00        | 98.75         | 97.38        |
| DWT + PCA + BPNN + SCG [17]   | 19                   | 100.00        | 98.29         | 97.14        |
| DWT + PCA + FNN + SCABC [12]  | 19                   | 100.00        | 98.93         | 97.81        |
| DWT + PCA + KSVM + LIN [18]   | 19                   | 96.01         | 95.00         | 94.29        |
| DWT + PCA + KSVM + HPOLY [18] | 19                   | 98.34         | 96.88         | 95.61        |
| DWT + PCA + KSVM + IPOLY [18] | 19                   | 100.00        | 98.12         | 97.73        |
| DWT + PCA + KSVM + GRB [18]   | 19                   | 100.00        | 99.38         | 98.82        |
| Proposed Scheme               | 9                    | <b>100.00</b> | <b>100.00</b> | <b>99.39</b> |

(100%, 100%, 99.39%), for the datasets-66, 160 and 255, respectively. The classification accuracy of the proposed system is also evaluated through receiver operating characteristic (ROC) curves, shown in Fig. 6. The proposed technique correctly classified the MRI images of Dataset-66 and Dataset-160 with an average area under curve (AUC) of 100%, with 0% standard deviation. For the Dataset-255 we achieved AUC of 99.45% ( $\pm 0.0046\%$ ).

We compared the performance of the proposed system with 13 state-of-the-art brain MR image classification schemes. As mentioned earlier, we implemented all these methods for proper performance comparison, and the performances of the implemented methods were evaluated using all the three datasets. The comparison results are shown in Table 4. The Table 4, also shows the feature vector dimension for each of the schemes. The schemes described in [13] give the worst performance results in terms of accuracy. Moreover, these methods have the highest feature dimension (4761 features/image), which results in high computational complexity. The dimension (7) of the feature vector used in [15], is less than our proposed method (9). But it is obvious from the results of Table 4, that the method described in [15] is less efficient and general than the proposed scheme, in terms of classification accuracy. The techniques described in [12, 16–18] show improved results in brain MR image classification, with lower feature

vector dimension (19). But, these schemes use various complex weight optimization techniques, which themselves require high computational complexity. Whereas, our proposed system only requires feature vector of dimension 9, with the highest retrieval accuracies.

During time requirement analysis we used all the images of all the three datasets for computing the overall time requirement of the proposed scheme. The computation times (except the LS-SVM training) of all the constituting stages (feature extraction, feature reduction and classification) of the proposed system was recorded, and we found out the average values as the time requirement. The feature extraction, feature reduction and LS-SVM classification average time requirements for an MRI image of size  $256 \times 256$  are 0.026 seconds, 0.014 seconds and 0.002 seconds, respectively. The overall average computation time for each MRI image of size  $256 \times 256$  is about 0.042 seconds. Even though, the time requirement for feature extraction through RT is slightly greater than DWT, but due to the use of LS-SVM with only 9 principal components results in lower overall time requirement. Moreover, with 9 dimensional feature vector for each training image, the storage cost of stored image feature database is also reduced. From the above mentioned results and discussions, it is clear that the proposed system not only performs the best, among all the mentioned state-of-the-art brain MR image classification techniques, but also works efficiently with different sizes' of datasets and various disease classes.

## 7. CONCLUSION

The manual process of diagnosis brain MR images has several problems. This necessitates the requirement for developing diagnostic tools, which can automatically and accurately classify brain MR images as normal or abnormal. Even though, there exists several advanced schemes to achieve this goal. Most of the state-of-the-art techniques have various shortcomings: based on DWT which can not capture the subtle and intrinsic details of the brain MR images, relatively high feature vector dimensionality, use of neural network with computationally expensive complex weight optimization techniques, and lack of generalization capability, etc. In this article, we propose to combine the benefits of MGA of RT, and a computationally less expensive SVM (LS-SVM) to build a fully automatic and accurate brain MR image classification system. With this combination, we not only achieve higher feature reduction, but also acquire superior performance than the state-of-the-art schemes. Extensive experiments and comparison show the effectiveness of the proposed system. In

future, we will investigate the effectiveness of other transforms along with other supervised and un-supervised classification schemes for brain MR image classification.

## REFERENCES

1. Westbrook, C., *Handbook of MRI Technique*, John Wiley & Sons, 2008.
2. 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.
3. Prasad, P. V., *Magnetic Resonance Imaging: Methods and Biologic Applications (Methods in Molecular Medicine)*, Humana Press, 2005.
4. 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.
5. 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.
6. Maji, P., M. K. Kundu, and B. Chanda, "Second order fuzzy measure and weighted co-occurrence matrix for segmentation of brain MR images," *Fundamenta Informaticae*, Vol. 88, Nos. 1–2, 161–176, 2008.
7. 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.
8. 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.
9. Rombouts, S. A., F. Barkhof, and P. Scheltens, *Clinical Applications of Functional Brain MRI*, Oxford University Press, 2007.
10. 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.
11. Maji, P., B. Chanda, M. K. Kundu, and S. Dasgupta, "Deformation correction in brain MRI using mutual information



- and genetic algorithm,” *Proc. Int. Conf. Computing: Theory and Applications*, 372–376, 2007.
12. 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.
  13. Chaplot, S., L. M. Patnaik, and N. R. Jagannathan, “Classification 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.
  14. Maitra, M. and A. Chatterjee, “A Slantlet transform based intelligent system for magnetic resonance brain image classification,” *Biomedical Signal Processing and Control*, Vol. 1, No. 4, 299–306, 2006.
  15. 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.
  16. 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.
  17. Zhang, Y., Z. Dong, L. Wu, and S. Wang, “A hybrid method for MRI brain image classification,” *Expert Systems with Applications*, Vol. 38, No. 8, 10049–10053, 2011.
  18. Zhang, Y. and L. Wu, “An MR brain images classifier via principal component analysis and kernel support vector machine,” *Progress In Electromagnetics Research*, Vol. 130, 369–388, 2012.
  19. Xu, J., L. Yang, and D. Wu, “Ripplet: A new transform for image processing,” *Journal of Visual Communication and Image Representation*, Vol. 21, No. 7, 627–639, 2010.
  20. Candes, E. J. and D. L. Donoho, “Continuous curvelet transform: I. Resolution of the wavefront set,” *Applied and Computational Harmonic Analysis*, Vol. 19, No. 2, 162–197, 2005.
  21. Das, S., M. Chowdhury, and M. K. Kundu, “Medical image fusion based on ripplet transform type-I,” *Progress In Electromagnetics Research B*, Vol. 30, 355–370, 2011.
  22. Jolliffe, I. T., *Principal Component Analysis*, Springer, 2002.
  23. Suykens, J. A. K. and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, Vol. 9, No. 3, 293–300, 1999.