Least Squares SVM Approach for Abnormal Brain Detection in MRI using Multiresolution Analysis

Deepak Ranjan Nayak, Ratnakar Dash, Banshidhar Majhi
Department of Computer Science and Engineering,
National Institute of Technology, Rourkela, India- 769008
Email: depakranjannayak@gmail.com, ratnakar@nitrkl.ac.in, bmajhi@nitrkl.ac.in

***Abstract*—Developing automatic and accurate computer-aided diagnosis (CAD) systems for detecting brain disease in magnetic resonance imaging (MRI) are of great importance in recent years. These systems help the radiologists in accurate interpretation of brain MR images and also substantially reduce the time needed for it. In this paper, a new system for abnormal brain detection is presented. The proposed method employs a multiresolution approach (discrete wavelet transform) to extract features from the MR images. Kernel principal component analysis (KPCA) is harnessed to reduce the dimension of the features, with the goal of obtaining the discriminant features. Subsequently, a new version of support vector machine (SVM) with low computational cost, called least squares SVM (LS-SVM) is utilized to classify brain MR images as normal or abnormal. The proposed scheme is validated on a dataset of 90 images (18 normal and 72 abnormal). A 6-fold stratified cross-validation procedure is implemented and the results of the experiments indicate that the proposed scheme outperforms other competent schemes in terms of classification accuracy with relatively small number of features.**

***Keywords*—Magnetic resonance imaging (MRI); Discrete wavelet transform (DWT); Kernel principal component analysis (KPCA); Least squares support vector machine (LS-SVM)**

I. Introduction

Brain diseases are growing rapidly among children and adults throughout the world. According to the National Brain Tumor Foundation (NBTF) in the United States, it has been estimated that, in children, brain tumors are the reason for one-quarter of all cancer deaths [1]. In the year 2104, World Health Organization (WHO) reported that around 250,000 people globally were diagnosed with primary brain tumors every year. Therefore, early detection of brain disease is very important. Magnetic resonance imaging (MRI) has been used as the most suitable medical imaging technique for an accurate detection of various brain diseases in recent years [2]. It is a low-risk, non-invasive method that generates high-quality images of the anatomical structures of the human brain and gives rich information about the soft brain tissues anatomy [3], [4]. MRI provides better contrast for different brain tissues than all other imaging modalities [5]. These advantages have delineated MRI as the most well-known method of brain pathology diagnosis and treatment. However, the high volume of information leads difficulty in analyzing and interpreting MR images. Computer-aided diagnosis (CAD) systems are currently used which examines brain MR images with the help of image processing techniques. CAD systems help radiologists in accurate interpretation of brain MR images for detecting abnormal brain. Therefore, it is necessary to develop a CAD system to increase the diagnosis capability and to reduce the time required for it. One of the most important steps in this system is to find out a set of discriminative feature that can classify the normal brain MR image from the abnormal one. An assortment of techniques has been studied for this purpose.

Over the last decade, several researches have been carried out for brain MR image classification. The most widely used approach for feature extraction is the multiresolution analysis that decomposes original MR image into several sub-images. These images preserve information about both low and high frequencies. Wavelet transform is one of the most important approaches for the texture analysis of the image. Various researchers have used wavelet transform to extract the features from the MR image. Chaplot *et al.* [3] have utilized the approximation coefficients of two-dimensional discrete wavelet transform (2D DWT) of level-2 decomposition as the features and employed self-organizing map (SOM) and support vector machine (SVM) classifiers. Maitra and Chatterjee [6] have introduced Slantlet transform (ST) which is an improved version of DWT, for feature extraction and applied back-propagation neural network (BPNN) classifier. El-Dahshan *et al.* [7] have used the approximation coefficients of level-3 decomposition of 2D DWT to represent each image. Principal component analysis (PCA) was employed to reduce the number of coefficients. They used feed forward back-propagation artificial neural network (FP-ANN) and k -nearest neighbor (k-NN) classifiers separately to detect the normal and pathological brain. In [4], [8]–[10], the researchers have used the coefficients of level-3 approximation sub-band of 2D DWT to extract features from images and then employed PCA for feature reduction. They have suggested different classifiers with some training parameter optimization approaches, namely, feed forward neural network (FNN) with scaled chaotic artificial bee algorithm (SCABC) [9], FNN with adaptive chaotic particle swarm optimization (ACPSO) [8], and BPNN with scale conjugate gradient (SCG) [4]. Zhang *et al.* [10] have used a kernel SVM (KSVM) classifier with three kernels, viz., linear (LIN), homogeneous polynomial (HPOL), inhomogeneous polynomial (IPOL) and Gaussian radial basis (GRB), to segregate the normal and pathological MR images. They have achieved high classification accuracy with GRB kernel. Das *et al.* [5] have presented an efficient mutiscale geometric analysis tool, Ripplet transform (RT) for feature extraction followed by PCA for dimensionality reduction. A less expensive SVM approach, called least square SVM (LS-SVM) was applied for classification and they have achieved suitable results over larger datasets. Saritha *et al.* [11] suggested the combined wavelet entropy based spider web plots (SWP) to extract features. The entropy values were calculated

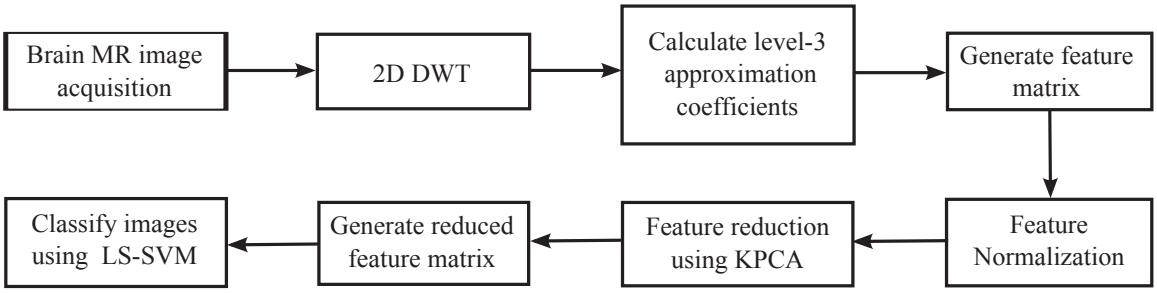


Fig. 1. Block diagram of the proposed scheme for detection of abnormal brain

for the approximation sub-bands of level-8 decomposition of Daubechies-4 wavelet. Finally, probabilistic neural network (PNN) was applied for classification. Zhang *et al.* [12] have used Shannon entropy (SE) and Tsallis entropy (TE) to get features from the discrete wavelet packet transform (DWPT) coefficients and suggested a generalized eigenvalue proximal SVM (GEPSVM) classifier. Zhou *et al.* [13] have achieved classification accuracy of 92.60% by using wavelet entropy values as the features for each image. They have applied a Naive Bayes classifier (NBC) to determine the normal and abnormal brain. Zhang *et al.* [14] have obtained 82.69% of accuracy using SVM classifier. To get the features, they have utilized wavelet-energy values of all the detail sub-bands of level-2 decomposition.

The literature review reveals different existing schemes for abnormal brain detection. Most of the schemes are not able to get a high classification accuracy. It has been observed that the dimension of the feature space is relatively high in many cases which may degrade the performance. When the extracted features have more complicated structures and can not be well represented in a linear subspace, then PCA will be not helpful for dimension reduction in such case. Hence, there is a need to use a new technique for nonlinear dimensionality reduction. Moreover, PCA requires a high computational cost for eigenvalue decomposition when the number of features is more than the number of images. To address the above issues, we have utilized the coefficients of approximation sub-band of 2D DWT for feature extraction. A kernel PCA (KPCA) approach is employed to handle the nonlinear feature values and to reduce the computational cost. To make the system more robust and computationally efficient, LS-SVM is used to classify the abnormal brain from the normal one. The proposed method is tested on a dataset of 90 images and the experimental results indicate that the scheme is superior to its competent schemes. The remainder of this paper is organized as follows. Section II deals with the working procedure of the proposed method. In Section III, the simulation results and comparisons are portrayed. Finally, Section IV gives the concluding remarks.

II. Proposed Method

The proposed method includes three important phases, namely, feature extraction, feature dimensionality reduction and classification. The overall block diagram of the proposed scheme is shown in Fig. 1. All the phases of the scheme are portrayed below in detail.

A. Feature Extraction using mutltiresolution technique

The proposed scheme uses a popular multiresolution technique, called DWT to extract features from the brain MR images. Wavelet transform is proven to be a powerful mathematical tool for feature extraction [15]. Compared to other transformation techniques, wavelet transform provides time-frequency localization of an image which is very important for classification.

A 2D DWT is implemented using low pass and high pass filters and down samplers. In case of images, the DWT is applied to each dimension individually, which results in four sub-band images (LL, LH, HL, HH) at each level. Among them, three sub-band images LH (low-high), HL (high-low) and HH (high-high) are the detail (high frequency) components in horizontal, vertical and diagonal directions, respectively. LL (low-low) sub-band image is the approximation (low pass) component which is used for next level 2D DWT calculation [4]. Fig. 2 illustrates the wavelet decomposition of a normal brain MR image up to three resolution levels. In this study, we have utilized the coefficients of the approximation sub-band of level-3 decomposition (LL_3) of Daubechies-4 wavelet to extract features. Daubechies-4 provides better resolution for smoothly varying signals in case if MR images of the brain. Therefore, we have selected Daubechies-4 wavelet, which gives better classification accuracy. The coefficients of LL_3 sub-band are arranged in row-major order to generate a feature vector. Then a feature matrix is created by combining the vectors corresponding to all brain MR images. The extracted features have been normalized before employing to KPCA. The feature z is normalized to z_n using the following formula.

$$z_n = \frac{z - \mu}{\sigma} \tag{1}$$

where, μ and σ are the mean and standard deviation of the features, respectively. The normalized feature vectors are then sent to the next phase.

B. Feature Reduction using KPCA

The size of the feature space becomes large if the approximation coefficients are directly used as the features, and all the features are not relevant for classification. Hence, to make the classification task feasible, the dimensionality of the feature vector needs to be significantly reduced, and informative features need to be extracted. PCA is often used for this purpose [16]. However, PCA only allows the linear dimensionality reduction and it doesn't perform well on the high-dimensional features having complicated structures. Therefore,

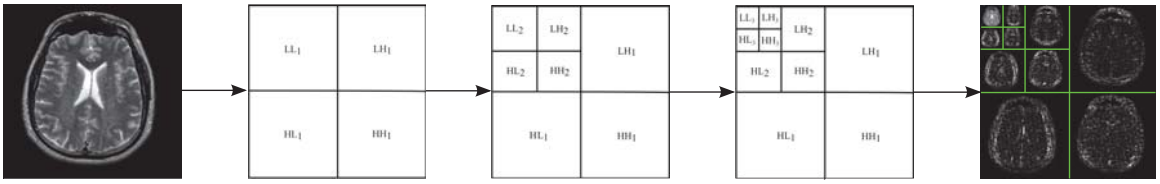


Fig. 2. A normal brain MR image and its wavelet decomposition at three resolution level

a non-linear form of PCA, called kernel PCA (KPCA) is employed in this paper for the dimensionality reduction of the features. Additionally, KPCA is computationally efficient than conventional PCA when the size of the feature space is greater than the number of samples [17], [18].

Consider a dataset $\{x_k \in X\}$ of N observations where $k = 1, 2, \ldots, N$ and each x_k is a d-dimensional feature vector. Suppose a nonlinear transformation $\phi(x)$ from original feature space to a high dimensional feature space is given, such that $\phi(x) : \Re^d \rightarrow \Re^D$, where $D >> d$. Each data point x_k is projected to a point $\phi(x_k)$. At first, we have assumed that the data is centered, *i.e.* $\frac{1}{N}\sum_{k=1}^{N}\phi(x_k) = 0$.

The covariance matrix of the projected features is defined by,

$$C_D = \frac{1}{N}\sum_{j=1}^{N}\phi(x_j)\phi(x_j)^T \tag{2}$$

Now, we have to find out the eigenvalues λ and eigenvectors V of C_D using

$$\lambda V = C_D V \tag{3}$$

Multiplying $\phi(x_k)$ in both sides of (3), we get,

$$\lambda(\phi(x_k).V) = (\phi(x_k).C_D V)\,; \quad k = 1, 2, \ldots, N \tag{4}$$

The eigenvectors can be expressed as the linear combinations of projected features and is given by,

$$V = \sum_{i=1}^{N}\beta_i\phi(x_i) \tag{5}$$

where β_i, $i = 1, 2, \ldots, N$ are the coefficients. Substituting (2) and (5) in (4), we have,

$$\lambda\sum_{i=1}^{N}\beta_i\left(\phi(x_k)\phi(x_i)\right) = \frac{1}{N}\sum_{i=1}^{N}\beta_i\left(\phi(x_k)\sum_{j=1}^{N}\phi(x_j)^T\right)\left(\phi(x_j)\phi(x_i)\right), \quad \forall k = 1, 2, \ldots N \tag{6}$$

and defining an $N \times N$ matrix K by,

$$K_{ij} = (\phi(x_i).\phi(x_j)) \tag{7}$$

we can write (6) as,

$$N\lambda K\beta = K^2\beta \tag{8}$$

where β is a N-dimensional column vector with entries $\beta_1, \beta_2, \ldots, \beta_N$ and can be solved by the eigenvalue problem

$$N\lambda\beta = K\beta \tag{9}$$

Here, K is a symmetric and positive semidefinite matrix. We then normalize the eigenvectors β to ensure that the corresponding eigenvectors V are orthonormal. The resulting k^{th} principal component of a test sample x is calculated using

$$\begin{aligned} y_k(x) = \left(V^k.\phi(x)\right) &= \left(\sum_{i=1}^{N}\beta_i^k\phi(x_i)\right).\phi(x) \\ &= \sum_{i=1}^{N}\beta_i^k k(x_i, x) \end{aligned} \tag{10}$$

In general, $\{\phi(x_k)\}$ may not be zero mean, and the Gram matrix $\widetilde{K}$ can be used for this purpose to substitute the kernel matrix K [19]. The Gram matrix is defines as,

$$\widetilde{K} = K - 1_N K - K 1_N + 1_N K 1_N \tag{11}$$

where 1_N is the $N \times N$ matrix with all the elements equal to $1/N$. The main advantage of kernel methods is that we do not need to compute $\phi(x_k)$ explicitly. Rather, we can directly construct the kernel matrix from the input dataset X [18]. Three commonly used kernels are polynomial kernel, Gaussian kernel and sigmoid kernel.

In this paper, we have used polynomial kernel to construct K. Using the above procedure, we have obtained an reduced feature matrix from the normalized feature matrix. The reduced matrix and a target vector is sent to the classifier to determine the abnormal brain.

C. Classification using LS-SVM

Standard SVM leads to high computational complexity in the case when it deals with a large dimensional dataset. To overcome the computational overhead, a least squares version of SVM (LS-SVM) is exploited as the classifier in this paper. Because of the equality constraints in the LS-SVM formulation, a set of linear equations has to be solved, instead of a quadratic programming problem for standard SVM [20].

Given a training set of N data points $\{x_k, d_k\}_{k=1}^{N}$ with input data $x_k \in \Re^m$ and class labels $d_k \in \Re$, where $d_k = \{-1, +1\}$, LS-SVM can be formulated as the optimization problem:

$$\min_{w,b,e} J(w, b, e) = \frac{1}{2}w^T w + \zeta\frac{1}{2}\sum_{k=1}^{N}e_k^2 \tag{12}$$

subject to the equality constraint

$$d_k\left[w^T\varphi(x_k) + b\right] = 1 - e_k, \quad k = 1, 2, \ldots, N \tag{13}$$

where w is the weight vector, $\varphi(.)$ the mapping function, $\zeta > 0$ the regularization factor, b a bias term and e_k the error variables.

The Lagrangian can be defined as

$$\mathcal{L}(w,b,e,\alpha) = J(w,b,e) - \sum_{k=1}^{N} \alpha_k \left\{ d_k \left[w^T \varphi(x_k) + b \right] - 1 + e_k \right\} \quad (14)$$

where α_k are Lagrange multipliers. The conditions for optimality are: $\frac{\partial \mathcal{L}}{\partial w} = 0 \rightarrow w = \sum_{k=1}^{N} \alpha_k d_k \varphi(x_k)$; $\frac{\partial \mathcal{L}}{\partial b} = 0 \rightarrow \sum_{k=1}^{N} \alpha_k d_k = 0$; $\frac{\partial \mathcal{L}}{\partial e_k} = 0 \rightarrow \alpha_k = \zeta e_k$; and $\frac{\partial \mathcal{L}}{\partial \alpha_k} = 0 \rightarrow d_k \left[w^T \varphi(x_k) + b \right] - 1 + e_k = 0$, which can be written as the solution to the following set of linear equations

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -D^T \\ 0 & 0 & \zeta I & -I \\ Z & D & I & 0 \end{bmatrix} \begin{bmatrix} w \\ b \\ e \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vec{1} \end{bmatrix} \quad (15)$$

where $Z = \left[\varphi(x_1)^T d_1; \ldots; \varphi(x_N)^T d_N \right]$, $D = [d_1; \ldots; d_N]$, $\vec{1} = [1; \ldots; 1], e = [e_1; \ldots; e_N], \alpha = [\alpha_1; \ldots; \alpha_N]$. We can also obtain the solution as

$$\begin{bmatrix} 0 & -D^T \\ D & \Omega + \zeta^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \quad (16)$$

where $\Omega = ZZ^T$ and according to Mercer's condition [20],

$$\Omega_{kl} = d_k d_l \varphi(x_k)^T \varphi(x_l) = d_k d_l \mathcal{K}(x_k, x_l) \quad (17)$$

where, $\mathcal{K}(.,.)$ is the kernel function. The LS-SVM classifier is obtained by

$$f(x) = sign \left[\sum_{k=1}^{N} \alpha_k d_k \mathcal{K}(x, x_k) + b \right] \quad (18)$$

The kernels that we have used for training the LS-SVM classifier are listed in Table I. The parameter θ indicates the degree of the polynomial and σ is a free parameter which controls the shape of the kernel.

TABLE I. DIFFERENT KERNEL FUNCTIONS USED IN LS-SVM

Kernel	Definition
Linear	$\mathcal{K}(x, x_k) = x_k^T x$
Polynomial	$\mathcal{K}(x, x_k) = (x_k^T x + 1)^\theta$
Radial Basis Function (RBF)	$\mathcal{K}(x, x_k) = exp\left\{ -\|\|x - x_k\|\|^2 / 2\sigma^2 \right\}$

III. EXPERIMENTAL RESULTS AND COMPARISONS

The experiments were carried out on a PC with 3.40 GHz Core-i7 processor and 4 GB of RAM, running under Windows 8 operating system. The proposed algorithm is simulated using MATLAB toolbox. The pseudocode of the proposed CAD system is presented in *Algorithm 1*.

1) Experimental setup: The dataset includes 90 images, 18 from each of the five category normal, brain tumor, stroke, degenerative disease and infectious disease. So there are total 18 normal and $18 \times 4 = 72$ abnormal images. The dataset consists of T2-weighted MR brain images in the axial plane and 256×256 in-plane resolutions which are downloaded from the Harvard Medical School website [21]. A sample from each of the category is shown in Fig. 3.

Algorithm 1 Pseudocode of the proposed system

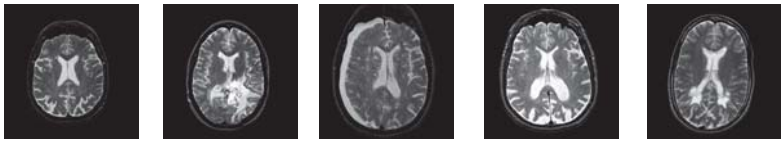
Require: Brain MR images of size 256×256
N: Total number of images
M: Total number of features
R: Number of reduced features
Ensure: Normal or abnormal brain
Step 1: Feature extraction using 2D DWT
for $i \leftarrow 1$ to N **do**
Read the brain MR images
Compute the LL_3 coefficients of 2D DWT
Arrange the coefficients in row-major order and store in a matrix $Q(N \times M)$
end for
Normalize the matrix Q using (1) and obtain a new matrix $Q_n(N \times M)$
Step 2: Feature reduction using KPCA
Choose a kernel function and find out the kernel matrix $K(N \times N)$ using Q_n
Calculate the normalized kernel matrix of the data *i.e.* $\widetilde{K}$
Solve the eigenvalue problem, $\lambda\beta = \widetilde{K}\beta$
Select R principal components corresponding to the R largest eigenvalues
Generate a reduced feature matrix $X(N \times R)$
Step 3: Classification using LS-SVM
Create a training dataset $\{x_k, d_k\}_{k=1}^{N}$ with input data $x_k \in X$ and class label $d_k \in D = \{-1, +1\}$
Apply n-fold cross validation procedure to find training and testing samples
Choose a kernel function $\mathcal{K}$ and train the LS-SVM classifier
Classify test images as normal or abnormal

To make the classifier more reliable and more generalize to independent datasets, 6-fold stratified cross-validation (CV) procedure is employed. The setting of the training and the validation images of the dataset is shown in Table II.

TABLE II. SETTING OF 6-FOLD STRATIFIED CV PROCEDURE

Total number of images (90)		Training images (75)		Validation images (15)	
Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
18	72	15	60	3	12

In this work, we consider the abnormal and normal class as the positive and negative class, respectively. Table III lists the measures which are used to calculate the performance of the proposed scheme and its competent schemes. TPR (Sensitivity) is the probability that a diagnostic test is positive, given that the person has the disease, whereas TNR (Specificity) is the probability that a diagnostic test is negative, given that the person does not have the disease. ACC is the probability that a diagnostic test is correctly performed. The classifier's performance can also be evaluated using an important index value, called the area under the curve (AUC) which is calculated by the receiver operating characteristic (ROC) curve. For an ideal classifier, the value of AUC is 1.



(a) (b) (c) (d) (e)

Fig. 3. Sample of brain MR images with (a) Normal brain, (b) brain tumor, (c) stroke, (d) degenerative disease, (e) infectious disease

TABLE III. DIFFERENT CLASSIFICATION PERFORMANCE MEASURES

Measures	Definition
True Positive Rate (TPR)	TP/(TP+FN)
True Negative Rate (TNR)	TN/(TN+FP)
Accuracy (ACC)	(TP+TN)/(TP+TN+FP+FN)

TP (True Positive): correctly classified positive cases, TN (True Negative): correctly classified negative cases, FN(False Negative): incorrectly classified positive cases, FP(False Positive): incorrectly classified negative cases

2) Results and Discussion: The proposed method utilizes the coefficients of LL_3 sub-band as the primary features of each MR image. However, the size of the feature space is 1444 for Daubechies-4 wavelet, which is quite large for computation. Thus, KPCA approach is used to reduce the dimensions of features to only 7. These reduced features are the first 7 principal components(PCs) which are only 0.48% of the primary features. The polynomial kernel has been selected to calculate the kernel matrix in KPCA method. The performance of the method is tested with different number of principal components to find out the required number of features. It has been observed that the proposed system works efficiently with 7 PCs on the given dataset. The LS-SVM classifier has been trained with the three kernels (linear, polynomial, and RBF). To estimate the optimal value of the parameters, viz., θ, σ and ζ, various pairs of the (ζ, θ) and (ζ, σ) are tested and finally the pair with the low error rate is chosen to train the classifier. The confusion matrix for linear, polynomial, and RBF kernel is illustrated in Table IV, V, and VI, respectively.

TABLE IV. CONFUSION MATRIX FOR 'LS-SVM + LINEAR' CLASSIFIER

		Output (predicted) class	
		Abnormal (positive)	Normal (negative)
Target class	Abnormal (positive)	67	5
	Normal (negative)	0	18

TABLE V. CONFUSION MATRIX FOR 'LS-SVM + POLYNOMIAL' CLASSIFIER

		Output (predicted) class	
		Abnormal (positive)	Normal (negative)
Target class	Abnormal (positive)	71	1
	Normal (negative)	0	18

The performance measures of the three classifiers are shown in Table VII. From the table, it is evident that the classification accuracy and AUC value for RBF kernel is higher than other two kernels. However, all the three kernels achieve a specificity of 100%. The sensitivity of the three kernels

TABLE VI. CONFUSION MATRIX FOR 'LS-SVM + RBF' CLASSIFIER

		Output (predicted) class	
		Abnormal (positive)	Normal (negative)
Target class	Abnormal (positive)	72	0
	Normal (negative)	0	18

is 93.06%, 98.61% and 100%, respectively. We have also compared the ROC curves obtained by LS-SVM classifier with the three kernels and are shown in Fig. 4. Table VIII presents the classification performance comparison of our proposed method with the existing schemes. It is observed that the suggested scheme is superior to its competent schemes while it requires relatively less number of features.

TABLE VII. PERFORMANCE METRICS FOR THREE CLASSIFIERS

Classifier	Sensitivity (%)	Specificity (%)	ACC(%)	AUC
LS-SVM+Linear	93.06	100	94.44	0.965
LS-SVM+Polynomial	98.61	100	98.89	0.986
LS-SVM+RBF	100	100	100	1

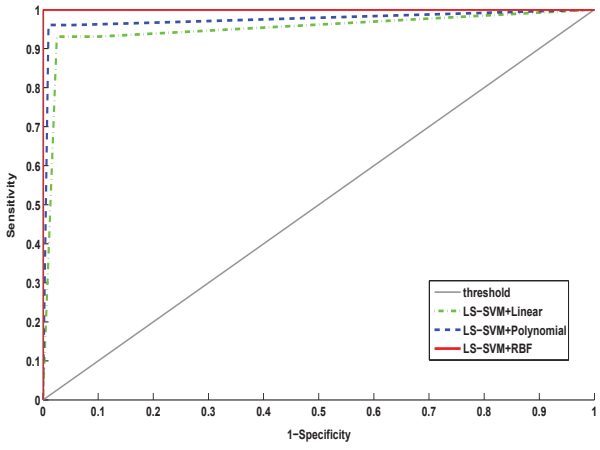


Fig. 4. ROC curves for LS-SVM classifier with three kernels

IV. CONCLUSION

This paper presents an efficient scheme to detect the brain MR images as abnormal. The scheme uses DWT to extract features from the images. The features have been normalized to enhance the efficiency. A computationally less expensive approach, KPCA is employed to select the most significant features from the high-dimensional normalized features. Finally, LS-SVM classifier has been used to build an automatic and

TABLE VIII. PERFORMANCE COMPARISON BETWEEN THE PROPOSED METHOD AND EXISTING SCHEMES

References	Total number of images	Feature extraction	Feature reduction	Classifier	Number of features	ACC (%)
Chaplot *et al.*, 2006 [3]	52	DWT	—	SVM with RBF	4761	98
El-Dahshan *et al.*, 2010 [7]	70	DWT	PCA	FP-ANN	7	97
El-Dahshan *et al.*, 2010 [7]	70	DWT	PCA	k-NN	7	98.6
Zhang *et al.*, 2011 [9]	66	DWT	PCA	FNN with SCABC	19	100
Zhang *et al.*, 2011 [4]	66	DWT	PCA	FNN with SCG	19	100
Zhou *et al.*, 2015 [13]	64	Wavelet entropy	—	Naive Bayes	7	92.60
Zhang *et al.*, 2015 [14]	66	Wavelet energy	—	SVM	6	82.69
Proposed method	**90**	**DWT**	**KPCA**	**LS-SVM+Linear**	**7**	**94.44**
				LS-SVM+Polynomial	**7**	**98.89**
				LS-SVM+RBF	**7**	**100**

accurate CAD system for brain MR image classification. The classification accuracies of LS-SVM with respect to the linear, polynomial, and RBF kernel are 94.44%, 98.89% and 100%, respectively. The results show the efficacy of the suggested scheme with considerably less number of features as compared to other schemes. Though the feature reduction technique and the classifier used in this paper are less expensive than the schemes proposed in the literature, however, the feature extraction step is more time consuming. The proposed work can be experimented with larger datasets, and the performance of the feature extraction stage can be enhanced using other advanced transformation techniques.

REFERENCES

[1] E. A. El-Dahshan, H. M. Mohsen, K. Revett, and A. B. M. Salem, "Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5526–5545, 2014.

[2] C. Westbrook, *Handbook of MRI technique*. John Wiley & Sons, 2014.

[3] S. Chaplot, 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, pp. 86–92, 2006.

[4] Y. Zhang, Z. Dong, L. Wu, and S. Wang, "A hybrid method for MRI brain image classification," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10 049–10 053, 2011.

[5] S. Das, M. Chowdhury, and K. Kundu, "Brain MR image classification using multiscale geometric analysis of ripplet," *Progress In Electromagnetics Research*, vol. 137, pp. 1–17, 2013.

[6] M. Maitra and A. Chatterjee, "A Slantlet transform based intelligent system for magnetic resonance brain image classification," *Biomedical Signal Processing and Control*, vol. 1, no. 4, pp. 299–306, 2006.

[7] E.-S. A. El-Dahshan, T. Honsy, and A.-B. M. Salem, "Hybrid intelligent techniques for MRI brain images classification," *Digital Signal Processing*, vol. 20, no. 2, pp. 433–441, 2010.

[8] Y. Zhang, 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, pp. 325–343, 2010.

[9] Y. Zhang, L. Wu, and S. Wang, "Magnetic resonance brain image classification by an improved artificial bee colony algorithm," *Progress In Electromagnetics Research*, vol. 116, pp. 65–79, 2011.

[10] Y. Zhang and L. Wu, "An MR brain images classifier via principal component analysis and kernel support vector machine," *Progress In Electromagnetics Research*, vol. 130, pp. 369–388, 2012.

[11] M. Saritha, K. P. Joseph, and A. T. Mathew, "Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network," *Pattern Recognition Letters*, vol. 34, no. 16, pp. 2151–2156, 2013.

[12] Y. Zhang, Z. Dong, S. Wang, G. Ji, and J. Yang, "Preclinical diagnosis of MR brain images via discrete wavelet packet transform with tsallis entropy and generalized eigenvalue proximal support vector machine (GEPSVM)," *Entropy*, vol. 17, no. 4, pp. 1795–1813, 2015.

[13] X. Zhou, S. Wang, W. Xu, G. Ji, P. Phillips, P. Sun, and Y. Zhang, "Detection of pathological brain in MRI scanning based on wavelet-entropy and naive bayes classifier," in *Bioinformatics and Biomedical Engineering*, 2015, pp. 201–209.

[14] G. Zhang, Q. Wang, C. Feng, E. Lee, G. Ji, S. Wang, Y. Zhang, and J. Yan, "Automated classification of brain MR images using wavelet-energy and support vector machines," in *2015 International Conference on Mechatronics, Electronic, Industrial and Control Engineering (MEIC-15)*, 2015.

[15] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1999.

[16] S. Theodoridis and K. Koutroumbas, *Pattern recgnition*. Academic Press, 2008.

[17] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Artificial Neural NetworksICANN'97*, 1997, pp. 583–588.

[18] ——, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[19] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[20] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[21] "Harvard medical school data," http://www.med.harvard.edu/AANLIB/.