Biomedical Signal Processing and Control 8 (2013) 667–674

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Short communication

# Spectral clustering independent component analysis for tissue classification from brain MRI

Sindhumol S.[a,*], Anil Kumar[b], Kannan Balakrishnan[a]

[a] Artificial Intelligence Lab, Department of Computer Applications, Cochin University of Science and Technology, Kochi, India
[b] Institute of Radiology and Imaging Sciences, Indira Gandhi Co-operative Hospital, Kochi, Kerala, India

ARTICLE INFO

*Article history:*
Received 19 April 2012
Received in revised form 13 May 2013
Accepted 14 June 2013
Available online 30 July 2013

*Keywords:*
Multispectral analysis
Magnetic resonance imaging
Spectral angle mapping
Independent component analysis
Support vector machine

ABSTRACT

A spectral angle based feature extraction method, Spectral Clustering Independent Component Analysis (SC-ICA), is proposed in this work to improve the brain tissue classification from Magnetic Resonance Images (MRI). SC-ICA provides equal priority to global and local features; thereby it tries to resolve the inefficiency of conventional approaches in abnormal tissue extraction. First, input multispectral MRI is divided into different clusters by a spectral distance based clustering. Then, Independent Component Analysis (ICA) is applied on the clustered data, in conjunction with Support Vector Machines (SVM) for brain tissue analysis. Normal and abnormal datasets, consisting of real and synthetic T1-weighted, T2-weighted and proton density/fluid-attenuated inversion recovery images, were used to evaluate the performance of the new method. Comparative analysis with ICA based SVM and other conventional classifiers established the stability and efficiency of SC-ICA based classification, especially in reproduction of small abnormalities. Clinical abnormal case analysis demonstrated it through the highest Tanimoto Index/accuracy values, 0.75/98.8%, observed against ICA based SVM results, 0.17/96.1%, for reproduced lesions. Experimental results recommend the proposed method as a promising approach in clinical and pathological studies of brain diseases.



## 1. Introduction

Magnetic Resonance Imaging (MRI) is a multi sequence medical imaging technique with which stacks of images are acquired with different tissue contrasts. Each sequence, namely, T1-weighted, T2-weighted, Proton Density (PD), Fluid-Attenuated Inversion Recovery (FLAIR), etc., highlights specific properties of tissues and pathologies, but none of them can provide completely decisive and reliable information. Intensity profile of 1000 continuous samples from T1-weighted, T2-weighted and FLAIR images is plotted in Fig. 1. Variations in tissue properties for same sample in different sequences can be observed from the given plots. Simultaneous observation and quantitative analysis of diagnostic information across MRI slices is a tedious job for radiologists. Multispectral MRI analysis [1,2] can simplify the job considerably, by combining unlimited number of co-registered sequences in a single suite. However, information extraction from this complicated structure is a challenging job, where the researchers and radiologists are focusing for last few decades [1–3].

Pattern recognition techniques are generally considered as the most effective methods for multi-spectral image analysis, where the classification methods are divided into unsupervised and supervised learning. Clarke et al. [3] presented a good review of relative merits of single versus multispectral image segmentation using supervised and unsupervised approaches. Unsupervised methods like Expectation Maximization (EM), k-means [2,3] and its fuzzy equivalent, the most widely used Fuzzy C-Means (FCM) [4] generally creates satisfactory results in MR image analysis [5,6]. But cluster is not a reliable method for accurate classification in pathological analysis [3]. The conventional supervised learning machines like Artificial Neural Networks (ANN) [3,7], Probabilistic Neural Networks (PNN) [8], Support Vector Machines (SVM) [7,9], and data conditioning approaches [3] have been effectively used in multispectral MRI analysis [10,11]. However, application of these conventional classification methods alone in multispectral MRI analysis often failed to provide expected clinical accuracy [3,5].

Recent researches on feature extraction using Independent Component Analysis (ICA) [12,13] and its extensions [14,15] have contributed a lot in high performance brain tissue classification and medical diagnosis. ICA and its non-linear extension,

* Corresponding author at: Artificial Intelligence Lab, Department of Computer Applications, Cochin University of Science and Technology, Kochi 682022, Kerala, India. Tel.: +91 9961558506.
*E-mail address:* sindhumol09@gmail.com (S. S.).

1746-8094/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bspc.2013.06.007

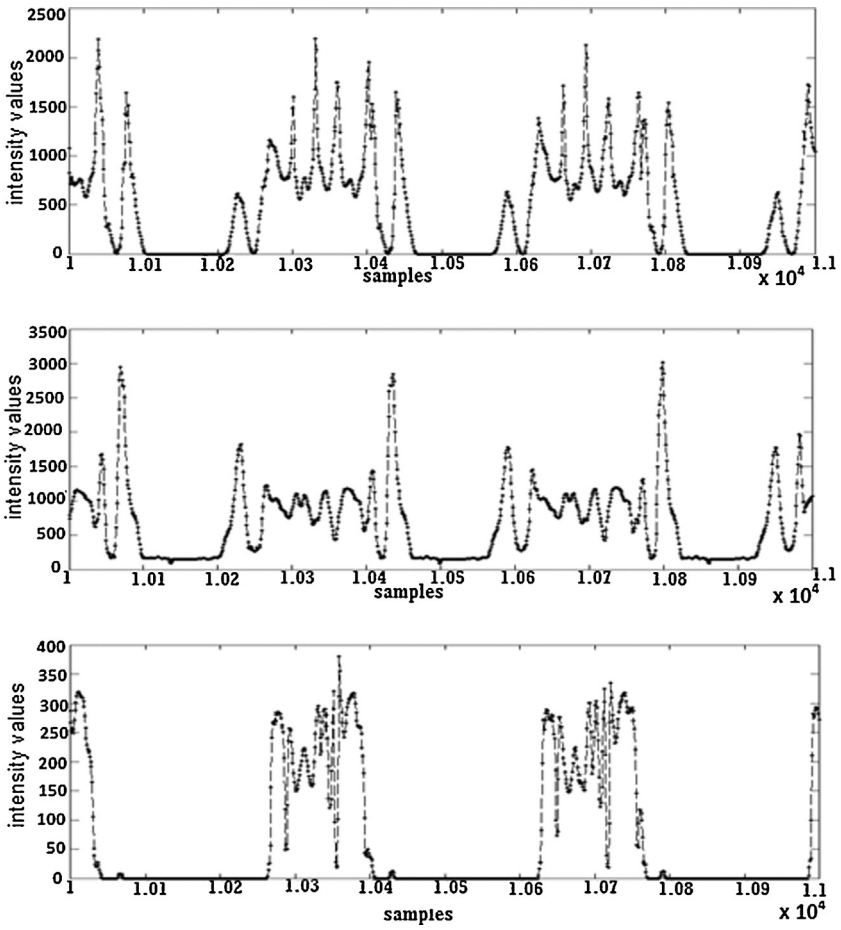


**Fig. 1.** Intensity distribution of samples from T1-weighted (top), T2-weighted (middle) and FLAIR (bottom) images.

kernel ICA [16], spectrally unmix the data into a set of statistically Independent Components (ICs) representing different brain tissues, Gray Matter (GM), White Matter (WM), Cerebro Spinal Fluid (CSF), etc. [13]. A thorough analysis of ICA based SVM (referred to as ICA + SVM) classification, and its performance in multispectral analysis with limited dimensionality is provided in [15,17]. As a global transform, ICA is good only at extracting the frequently occurred features and it fails to preserve the local characteristics [18]. Researchers have attempted to resolve these issues through extensions like local ICA [18,19], wavelet based ICA [20], spectral screening ICA [21], etc. However, the subset based approaches and approximations in these literatures find its difficulty in MRI analysis, where soft tissues and small abnormalities are to be extracted with high accuracy from a limited number of bands.

A new spectral clustering algorithm based on Spectral Angle Mapping (SAM) [22], Spectral Clustering Independent Component Analysis (SC-ICA), is proposed in this work to extend ICA for improved feature extraction from multispectral MRI. As a spectral angle based pre-processing technique, it can exploit intrinsic spectral characteristics of input signals and it is robust to input energy differences [22]. The main objective of this method is to retain less frequently occurred objects like small lesions, while dealing with massive amount of information in multispectral analysis. Performance evaluation of the proposed algorithm was done with SVM classification using synthetic and real, normal and abnormal, brain MRI images. Synthetic image analysis was performed on slices from T1-weighted, T2-weighted and PD sequences, whereas T1-weighted, T2-weighted and FLAIR sequences were selected for clinical image analysis. Detailed comparison of the proposed method with ICA based SVM [17], and other existing supervised classification methods like SVM, PNN, ANN and Bagging confirmed the superiority of the new approach in brain tissue and small abnormality analysis.

## 2. Materials and methods

### 2.1. Input datasets

Synthetic MR images included in this work contains normal and abnormal (multiple sclerosis) data obtained from BrainWeb [23], the Simulated Brain Database at the McConnell Brain Imaging Centre of the Montreal Neurological Institute (MNI), McGill University. Axial T1-Weighted Images (T1WI), T2-Weighted Images (T2WI), and PD Images (PDI) formed the input multispectral suite. Slices from each sequence have parameter settings, 1-mm slice thickness and noise level, 0% and 1%.

T1WI, T2WI and FLAIR images of 60 cases (40 normal and 20 abnormal), sampled by Siemens' whole body 3 T MR system (Siemens, AG Medical Solutions, Erlangen, Germany), were selected for clinical analysis with following settings: Axial spin echo T1WI with repetition time (TR) = 1600 ms, echo time (TE) = 8.9 ms, image dimension 320 × 320 pixels, and T2WI with TR/TE = 4000 ms/95 ms, size 416 × 512 pixels. FLAIR images with image dimension 464 × 512 pixels, TR/TE/inversion time = 6000 ms/94 ms/2026.5 ms for abnormal cases and with

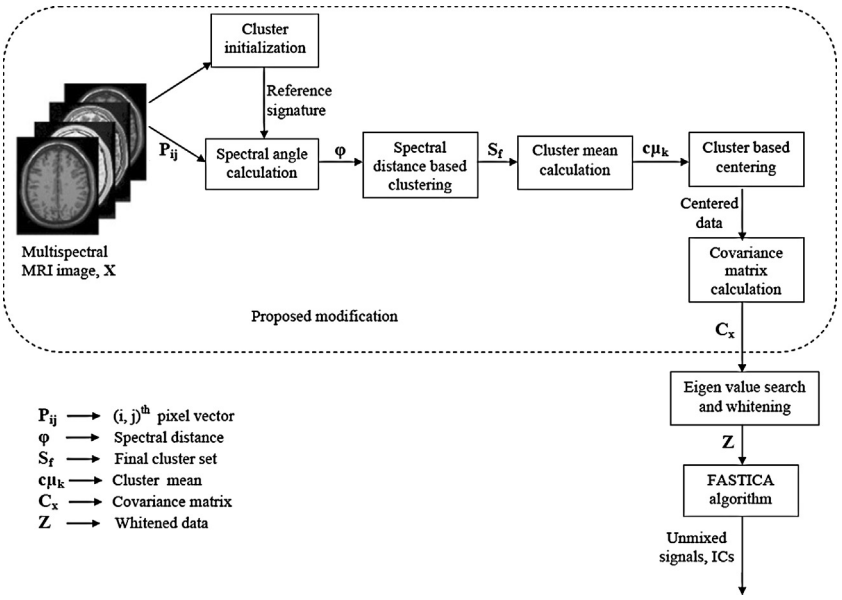


**Fig. 2.** Proposed algorithm.

TR/TE/inversion time = 9000 ms/94 ms/2500 ms for normal cases were also included in the analysis. Additional parameters were, slice gap 6.5 mm and slice thickness 5 mm.

### 2.2. Data analysis

The proposed method is divided into two parts: (i) spectral distance based clustering of multispectral MRI data and (ii) modifying the centering and whitening [24] part of the ICA algorithm with statistical measures from clustered multispectral data. Proposed algorithm is outlined in Fig. 2. The grouped portion explains the details of proposed extensions to simple ICA process. Initially, input multispectral MRI cube is formed by placing registered T1WI, T2WI and PD/FLAIR image in 1st, 2nd and 3rd dimensions, respectively. Let '**x**' be the $L$ (here $L = 3$) dimensional pixel vector, which is linearly mixed by a set of $m$ statistically independent tissues or tumor information, $s_1, s_2, \ldots, s_m$, by means of a $L \times m$ mixing matrix '**A**'. Proposed SC-ICA algorithm extracts the unknown sources, $s_1, s_2, \ldots, s_m$, from the image cube, as described below.

$\mathbf{P}_{ij}$ in Fig. 2 represents $(i, j)$th pixel vector, also known as the spectral signature at pixel location $(i, j)$. In this work, $\mathbf{P}_{ij} = [x_1\, x_2\, x_3]^T$, where $x_1$, $x_2$ and $x_3$ are $(i, j)$th intensity values for T1WI, T2WI and PD/FLAIR images, respectively. First, cluster set, **S**, is initialized with the first pixel vector. The next step is the similarity checking between spectral signatures by SAM [22], in which spectral angle, $\varphi$, is calculated as

$$\phi = \cos^{-1}\left(\frac{\mathbf{x}\cdot\mathbf{y}}{||\mathbf{x}||||\mathbf{y}||}\right) \quad (1)$$

where **x** and **y** are the input signature and reference signature, respectively.

For each input $\mathbf{P}_{ij}$, we considered the cluster mean as the reference signature to calculate $\varphi$. The pixel vectors are classified into appropriate groups by thresholding with some spectral angle threshold value, $\alpha$. Centering of the data for each cluster is separately done, and covariance matrix $\mathbf{C_X}$ is calculated. After that, common ICA procedures [24], as shown in Fig. 2, are followed to generate ICs representing different brain tissues. The entire procedure is summarized into two algorithms: spectral clustering algorithm and Spectral Clustering Independent Component Analysis (SC-ICA), as given below.

#### 2.2.1. Spectral clustering algorithm

*Input*: Multispectral MRI cube, **X**, with pixel vectors $\mathbf{P}_{ij}$ and spectral angle threshold, $\alpha$.
*Step1*: Initialize unique cluster set **S** with a set containing first pixel vector in the image cube.
*Step2*: ***loop1*** – for each $\mathbf{P}_{ij}$ do,
  begin
    ***loop2*** – for each unique cluster $c_k$ from **S**
      begin
        Calculate reference spectral signature,
          $\mathbf{c\mu}_k$ = average pixel vector for $c_k$
        Find $\varphi_{ij}$ = Angle between $\mathbf{P}_{ij}$ and $\mathbf{c\mu}_k$.
        if $\varphi_{ij} < \alpha$, add $\mathbf{P}_{ij}$ to $c_k$. Continue with ***loop1***
        else continue.
    end
  If $\mathbf{P}_{ij}$ is not added to any $c_k$ in **S**,
    Create a new cluster $c_{k+1}$ in set **S.**
    Add $\mathbf{P}_{ij}$ to $c_{k+1}$.
  end
*Step3*: Output unique cluster set $\mathbf{S_f}$ with elements $c_1, c_2, c_3, \ldots$

#### 2.2.2. SC-ICA algorithm

*Step1*: For each cluster $c_j$ in $\mathbf{S_f}$ do,
  i. Find mean $\mathbf{c\mu}_j$
  ii. Cluster based centering:
    for each cluster elements (pixel vectors) $\mathbf{P}_k$ in $c_j$,
      Calculate, $\mathbf{P}_k - \mathbf{c\mu}_j$
*Step2*: Calculate covariance matrix $\mathbf{C_X}$ with the statistics computed from previous step.
*Step3*: Compute **D**, diagonal matrix of $\mathbf{C_X}$'s Eigen values, and **E,** the orthogonal matrix of $\mathbf{C_X}$'s eigenvectors.
  Calculate whitening matrix [24] **P** as follows,
    $\mathbf{P} = \mathbf{D}^{-1/2}\mathbf{E}^T$
*Step4*: Whitening process: Use **P** from *Step 3* to calculate whitened data, $\mathbf{Z} = \mathbf{PX}$.
*Step5*: Apply FastICA [24] algorithm on **Z** to obtain object specific unmixed independent components.

Best feature vectors for training and testing process are selected from SC-ICA results, under the supervision of an experienced radiologist. SVM classification using RBF kernel [7] is applied on selected features to analyze the reproduced tissues, both visually and quantitatively.

### 2.3. Performance evaluation

Tanimoto Index (TI), the most commonly used criterion in medical imaging [12,17] is used to measure the similarity of the

**Table 1**
Evaluation details of three sets of experiments.

| Experiment | Category | Training data count | | Testing data count | | Total patients | Cross validation |
|---|---|---|---|---|---|---|---|
| | | Cases | Slices | Cases | Slices | | |
| Set1 | Synthetic normal | 2 | 10 | 2 | 4 | 2 | Holdout (training) |
| | Synthetic abnormal | 2 | 10 | 2 | 4 | 2 | Holdout (training) |
| | Clinical normal | 32 | 125 | 8 | 35 | 40 | Holdout (training) |
| | Clinical abnormal | 16 | 83 | 4 | 17 | 20 | Holdout (training) |
| Set2 | Clinical abnormal | 10 | 42 | 10 | 42 | 20 | Internal Leave-One-Out<br>External holdout<br>External 10-fold |
| Set3 | Clinical abnormal | 20 | 160 | | | 20 | External 10-fold |

reproduced tissues with the ground truth values, and it is given by,

$$\mathrm{TI} = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

where $A$ represents the set of binary information corresponding to reproduced brain tissues like CSF, GM, WM, Tumor, etc., and $B$ is the ground truth dataset. In this case, the ground truth is defined as a binary image representing a brain tissue, collected from traditional brain MRI tissue segmentation by a human expert.

Classifier performance evaluation in this work is conducted with widely used statistical measures, sensitivity, specificity, accuracy and error rate [25] given by

$$\text{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$$\text{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{4}$$

$$\text{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{5}$$

True Positive (TP) is defined as the number of correctly identified positive pixels; True Negative (TN) is defined as correctly identified negative pixels. For example, in a diagnostic test evaluation focusing on the presence of abnormal tissues, tumor samples are considered in the positive category and normal tissues will be in the negative category. False Positive (FP) represents the count of normal tissues incorrectly identified as tumor, and False Negative (FN) gives the count of abnormal samples incorrectly identified as normal tissues. Higher values of sensitivity, the proportion of correctly classified positives, indicate the good performance of the method in predicting positives. Specificity measures how well the system can predict the negatives. Accuracy measures the overall correctness of the classifier in predicting both positives and negatives, and overall error rate is calculated as

$$\text{Error rate} = 1 - \text{Accuracy} \tag{6}$$

## 3. Results and discussion

The proposed algorithm was evaluated by three sets of experiments using Mathworks Matlab 7.0 (R2009a). First set performed feature extraction from input data using SC-ICA for varying thresholds, classification by SVM, and comparison of reproduced tissues with those from ICA + SVM. Non-linear SVM (RBF kernel) functions in Pattern Recognition Toolbox, with default parameter settings were applied for training and classification. One-against-all SVM strategy was adopted to solve the classification problem. Evaluation of the proposed classifier, with error rate estimation using different validation techniques [25] was included in the second set. Third set was the statistical and time based comparative analysis of the proposed system with existing approaches in the literature. Table 1 explains the details of test plan and evaluation procedure for three sets. GM, WM, CSF and abnormality were the main brain tissue classes considered in this study. Feature extraction techniques were first applied on each training/test data to generate ICs.

### 3.1. Feature extraction

Both synthetic and clinical datasets were given as input to SC-ICA as well as to ICA for feature extraction. Then, we analyzed the improvement in extracted features through a performance comparison of generated ICs from SC-ICA and ICA. The same steps were repeated for SC-ICA on varying spectral angle threshold values. ICs extracted by SC-ICA and ICA for 0% noise level synthetic sample images are available from Fig. 3A (normal) and B (abnormal). Unmixed components from SC-ICA for threshold value 0.06 (referred to as SC-ICA_.06) were found to be yielding best results. It is evident from Fig. 3A and B that SC-ICA results are more specific compared to background dominating results from ICA. A clear picture of major tissues CSF, WM and GM was available from normal case ICs (Fig. 3A last row). In abnormal case analysis using SC-ICA, we could extract a unique feature set for each class, CSF, WM, GM and lesions (total 4 classes) from three ICs as shown in Fig. 3B last row. Contribution of SC-ICA in identifying presence of abnormalities in other tissues, especially in the case of WM, was very evident without any indexing (Fig. 3B last row).

### 3.2. Visual and quantitative analysis

Feature vectors, representing each class, for SVM training and classification were selected by an experienced radiologist, with the help of a $3 \times 3$ pixel window. The positive impact of SC-ICA in brain tissue reproduction and classification was demonstrated by a detailed visual and quantitative analysis. 50% holdout cross validation on selected feature vectors from training set was repeated in 100 trials, to train the classifier for experiment set1. Then, we applied the selected model on the preprocessed test dataset to reproduce the brain tissues. ICA based classification was also repeated in the same environment to confirm the improvement by SC-ICA.

#### 3.2.1. Synthetic image analysis

Reproduced tissues from normal and abnormal data, Fig. 3C (normal) and D (abnormal), indicated that SC-ICA_.06 + SVM yields maximum information on CSF, GM and WM. As described in feature extraction, last two columns of set B demonstrates the contribution of SC-ICA in local feature analysis, specifying better details of abnormality and its presence in WM. Tanimoto Index (TI) values were measured for reproduced CSF, GM, WM, and White Matter Lesions (WML), using ground truth images in Brainweb database. Observed results for normal and abnormal cases, with ICA + SVM and SC-ICA + SVM on varying $\alpha$ values, 0.1, 0.06 and 0.03 are summarized

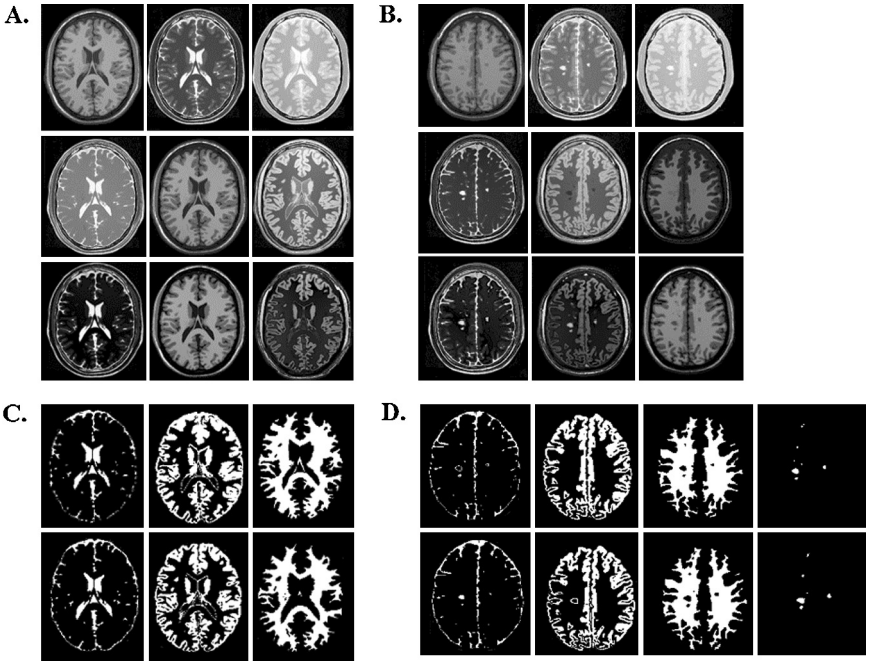


**Fig. 3.** Synthetic image analysis of normal (left sets: A and C) and abnormal (right sets: B and D) slices. (A and B): Upper row, input T1WI, T2WI and PDI. Middle row, three ICs from ICA. Last row, three ICs from SC-ICA_.06. (C and D): Reproduced tissues, CSF, GM, WM (from left to right) and WML (only for set D). 1st row, results from ICA + SVM; 2nd row, results from SC-ICA_.06 + SVM.

in Table 2. Potential of the proposed method in WML detection was evaluated through sensitivity, specificity and accuracy values (measured in percentage). Compared to ICA + SVM results, overall performance of SC-ICA + SVM was very good except for $\alpha$ value 0.1. SC-ICA_.06 + SVM was found to be yielding best results for synthetic brain tissue analysis. We observed a very small TI value, 0.0357, for SC-ICA_.03 + SVM. But improved sensitivity value, 97.3%, demonstrated that lesion details were not lost from the result. Low TI value resulted from false classification due to over clustering by SC-ICA_.03.

### 3.2.2. Clinical image analysis

Control points based image registration using MATLAB functions were applied on normal and abnormal clinical data to provide co-registered T1WI, T2WI and FLAIR slices (Fig. 4 top row). Dimensions of the registered images were fixed to $209 \times 276$ pixels and $227 \times 260$ pixels for normal and abnormal cases. Feature extraction using ICA, and SC-ICA for different threshold values was followed as described in Section 3.1. CSF, GM and WM were reproduced from normal case (Fig. 4A), and CSF, GM, WM and abnormal tissues were classified from abnormal case (Fig. 4B). For normal case tissue analysis, GM classification was highly affected by intensity levels of FLAIR image, and best results were provided by SC-ICA_.12 + SVM. Abnormality in clinical case shown in Fig. 4B was observed as a lesion surrounded by edema. ICA + SVM results best described the lesion part, but failed to reproduce surrounding edema. SC-ICA + SVM results could specify the abnormality and its effect on other tissues in a better way. SC-ICA_.03 + SVM (Fig. 4B last row) extracted maximum abnormality, which seems to be very useful in clinical analysis.

TI values of reproduced CSF, GM, WM, and lesions were calculated as explained in Section 2.3 (Table 2). Manually segmented images from experienced clinical experts were considered as ground truth images. SC-ICA_.1 + SVM provided good results for normal tissue analysis, but optimal values were observed from SC-ICA_.12 + SVM. Normal case results for SC-ICA_.03 was not satisfactory in visual and quantitative analysis. But it was highly attractive for abnormality analysis, which was emphasized through sensitivity, specificity and accuracy values of abnormal tissues as given in Table 2. High accuracy values observed for SC-ICA_.03 + SVM and SC-ICA_.06 + SVM, 98.8% and 97.6%, are very promising in clinical abnormal tissue analysis. Analysis of reproduced results demonstrated the feasibility of the proposed algorithm for improved MRI analysis. To confirm the efficiency of

**Table 2**
Quantitative analysis of reproduced tissues from synthetic and clinical images.

| | | Tanimoto Index (TI) | | | | | | | | | Statistical measures (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Normal cases | | | | Abnormal cases | | | | | Lesions | | |
| | | α | CSF | WM | GM | α | CSF | WM | GM | Lesion | Sensitivity | Specificity | Accuracy |
| Synthetic | SC-ICA + SVM | 0.1 | 0.71 | 0.92 | 0.59 | 0.1 | 0.80 | 0.92 | 0.80 | 0.79 | 84.2 | 99.9 | 99.9 |
| | | 0.06 | 0.78 | 0.94 | 0.71 | 0.06 | 0.81 | 0.94 | 0.84 | 0.85 | 91.1 | 99.9 | 99.9 |
| | | 0.03 | 0.77 | 0.91 | 0.62 | 0.03 | 0.77 | 0.97 | 0.79 | 0.04 | 97.3 | 90.2 | 90.2 |
| | ICA + SVM | | 0.69 | 0.86 | 0.61 | | 0.75 | 0.91 | 0.84 | 0.85 | 86.2 | 99.9 | 99.9 |
| Clinical | SC-ICA + SVM | 0.1 | 0.72 | 0.83 | 0.43 | 0.1 | 0.42 | 0.87 | 0.43 | 0.38 | 38.2 | 99.4 | 96.5 |
| | | 0.12 | 0.72 | 0.87 | 0.46 | 0.06 | 0.74 | 0.89 | 0.36 | 0.50 | 50.3 | 99.9 | 97.6 |
| | | 0.03 | 0.60 | 0.51 | 0.57 | 0.03 | 0.62 | 0.89 | 0.29 | 0.75 | 76.4 | 99.9 | 98.8 |
| | ICA + SVM | | 0.62 | 0.74 | 0.42 | | 0.20 | 0.83 | 0.27 | 0.17 | 17.4 | 99.9 | 96.1 |

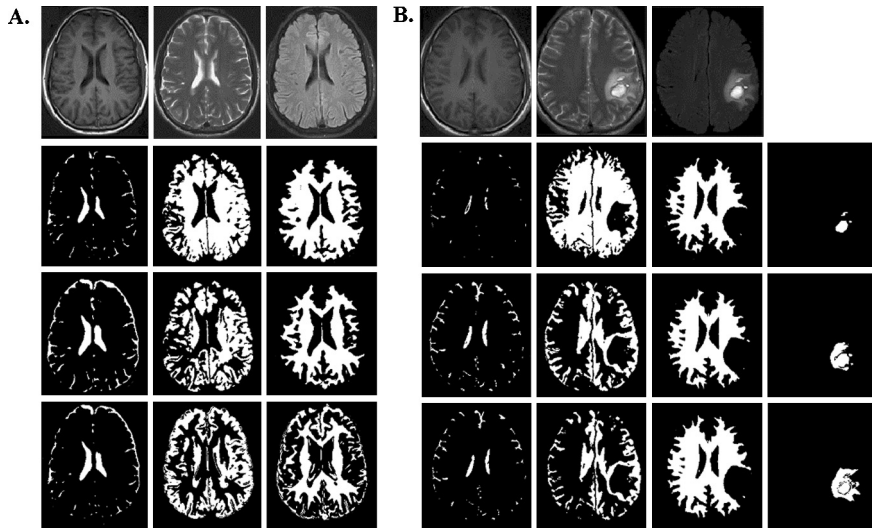


**Fig. 4.** Classified results for normal (set A) and abnormal (set B) clinical MRI datasets, in the order of CSF, GM, WM and lesion (only for set B) from left to right. 1st row: input images, T1WI, T2WI, FLAIR image. 2nd row: results from ICA + SVM. 3rd row: results from SC-ICA_12 + SVM for set A, results from SC-ICA_06 + SVM for set B. Last row: results from SC-ICA_03 + SVM.

the method in classification, we conducted a detailed bias estimation and comparative analysis also. These experiments were carried out with clinical data, and observed optimal threshold value, 0.06 was fixed for SC-ICA.

### 3.3. Error rate estimation with respect to number of selected feature vectors

Second set of experiments, error rate estimation, followed an equal size distribution of input data into training and test sets as given in Table 1. Cross Validation (CV) [26] techniques like Internal CV Leave One Out (ICV LOO), External 50% Holdout (EHoldout) and External 10-fold CV (ECV 10) were considered for bias estimation. Error estimates were first observed for 3000 feature vectors in 100 random splits. Same procedures were repeated for different number of feature vectors 2500, 2000, etc., and average error rates were estimated for a detailed bias estimation as plotted in Fig. 5. Results obtained for unbiased prediction error (TE) were also added to study the error variation. A detailed explanation of selection bias and error rate estimation is available in [25,26]. As shown in Fig. 5, ICV LOO results were found to be more biased compared to EHoldout, whereas ECV 10 and TE keep almost similar behavior.

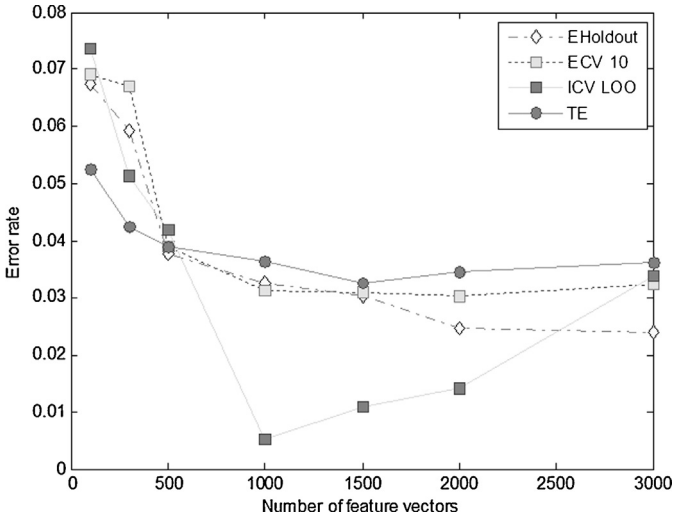


**Fig. 5.** Classification error rates for abnormal tissues in clinical datasets, measured by decreasing the number of feature vectors.

### 3.4. Comparison of different supervised approaches

Detailed comparative analysis of six classification algorithms including proposed method was performed in the third set of experiments. Total 4000 (2000 for training and 2000 for test) feature vectors from preprocessed data were selected in a single sample database. Details of the evaluation plan are available from Table 1. Matlab implementations for 4-layer feed-forward back propagation network with three hidden layer sizes [1085], probabilistic neural network with spread value 0.9, default SVM, and ensemble with 100 bagged decision trees were also employed in the study. Optimal parameters yielding best results were selected by trial and error. Average sensitivity, specificity, accuracy and standard deviation (std.) of the measures were observed with tumor and WM classes. SC-ICA + SVM, ICA + SVM and Bagging could give relatively good results as summarized in Table 3. In order to justify the performance of the proposed method in real environment, we measured average time required for classification and training (for single iteration) on a Windows7 PC with Pentium Dual CPU of 2.0 GHz/2 GB RAM for lesion analysis. Comparatively less classification time, 0.05 s and 0.13 s, was observed for SC-ICA + SVM and ICA + SVM, respectively (Table 3). However, additional feature extraction cost was found to be challenging for SC-ICA and ICA. We observed, 8.5 + 1.27 + 0.05 s, as the minimum time for SC-ICA + SVM in a typical clinical abnormality analysis.

In this work, a new feature extraction method, which avoids the loss of local characteristics in multispectral analysis, has been presented to improve the brain tissue classification. The proposed algorithm was validated with synthetic and real MRI data. Compared to ICA based SVM for MRI analysis by Chai et al. [17], SC-ICA + SVM provides better performance in small object detection. In normal case analysis, SC-ICA + SVM outperforms ICA + SVM with best reproduced brain tissues (Figs. 3C and 4A) and improved TI values (Table 2) for CSF/GM/WM, 0.78/0.71/0.94 (synthetic case) and 0.72/0.57/0.87 (clinical case). Improved performance is confirmed in lesion analysis by observed best-case sensitivity/TI values, 76.4%/0.75, against 17.4%/0.17for ICA + SVM. The positive impact of SC-ICA + SVM in locating presence of abnormalities on other tissues is demonstrated through high TI value for WM, 0.89 (Table 2),

**Table 3**
Performance comparison of supervised classification methods for clinical abnormal data.

| | Method | Sensitivity ± std. (%) | Specificity ± std. (%) | Accuracy ± std. (%) | Feature extraction time (s) | Training time (s) | Classification time (s) |
|---|---|---|---|---|---|---|---|
| TUM | SC-ICA + SVM | 96.7 ± 0.2 | 97.7 ± 0.2 | 97.0 ± 0.2 | 8.5 | 1.27 | 0.05 |
| | ICA + SVM | 96.1 ± 0.4 | 86.7 ± 0.3 | 92.3 ± 0.3 | 0.6 | 3.17 | 0.13 |
| | SVM | 96.6 ± 0.4 | 48.1 ± 1.8 | 80.5 ± 0.8 | | 2.82 | 0.17 |
| | ANN | 91.9 ± 0.4 | 70.1 ± 1.1 | 84.5 ± 0.4 | | 13.03 | 0.05 |
| | PNN | 97.8 ± 0.2 | 63.3 ± 2.7 | 86.4 ± 1.1 | | 0.25 | 0.32 |
| | Bagging | 95.8 ± 0.4 | 82.2 ± 2.0 | 91.3 ± 0.8 | | | |
| WM | SC-ICA + SVM | 95.2 ± 1.2 | 96.1 ± 0.2 | 95.1 ± 0.7 | | | |
| | ICA + SVM | 85.6 ± 0.4 | 94.8 ± 1.4 | 91.9 ± 0.6 | | | |
| | SVM | 85.1 ± 0.3 | 71.4 ± 1.6 | 80.8 ± 0.4 | | | |
| | ANN | 89.3 ± 0.9 | 80.6 ± 2.6 | 86.6 ± 1.2 | | | |
| | PNN | 62.9 ± 0.6 | 94.1 ± 0.2 | 72.6 ± 0.4 | | | |
| | Bagging | 94.1 ± 0.6 | 85.7 ± 0.5 | 91.5 ± 0.3 | | | |

and best sensitivity value, 95.2% (Table 3) in clinical image analysis. By examining the performance results from existing multispectral classifiers (Table 3), we observed that the least standard deviations, and best sensitivity/specificity/accuracy values (%), 96.7/97.7/97.0, ensure the stability and superiority of SC-ICA + SVM in classification.

However, experimental results indicate that classification performance highly varies on selected threshold values. Low threshold values can improve classification of local features. But it may lead to over-clustering, which adversely affects normal tissue analysis. Distorted WM tissue in Fig. 4A for SC-ICA_.03 + SVM explains this with the lowest TI value, 0.51 (Table 2). Input data characteristics have a big role in optimal threshold selection. SC-ICA_.06 performs well for all tissues in synthetic data analysis, whereas clinical normal case analysis supports threshold 0.12 and abnormal tissue analysis finds its best results with threshold 0.03. Feature extraction cost due to clustering in SC-ICA is another issue, but it can be justified by the quality improvement in classified results.

Qualitative and quantitative analysis results support SC-ICA + SVM as a promising approach in brain tissue classification and pathological analysis. It retains local and global characteristics with equal priority in classified results. Efficiency of the method in simultaneous analysis of small abnormalities like WML, and their effect on other tissues can help doctors in disease progress evaluation and treatment. Real environment artifacts such as Gaussian noise, intensity inhomogeneity, and so forth are not considered in this work. In future, the proposed method can be refined with an adaptive threshold selection scheme. Expansion of multispectral data with more informative MRI sequences in diagnostic studies can extend the applications to analysis of several brain diseases.

## 4. Conclusion

A spectral clustering extension to ICA (SC-ICA) has been proposed for multispectral brain MRI analysis. The algorithm is innovative in the way it approaches the problem as a data mining problem, to extract small details from dominating background and other tissues using spectral clustering, ICA and SVM. The proposed method is evaluated both quantitatively and qualitatively, using clinical and synthetic MRI data, for normal and abnormal cases. A detailed comparison is performed with results from ICA + SVM and other conventional classification methods. SC-ICA based SVM provides a high performance brain tissue classification, especially in the case of small lesions and tumors in clinical analysis. However, threshold value selection plays a significant role in accuracy and reproducibility of classified results. Future work focuses on an adaptive system that can solve the threshold and time complexity issues. Clinical accuracy of results can be improved significantly by addition of high informative MRI sequences to input data.

## Acknowledgment

We would like to thank Institute of Radiology and Imaging Sciences (IRIS) Pvt. Ltd., Kochi for supporting us with required medical guidance in this work.

## References

[1] T. Taxt, A. Lundervold, Multispectral analysis of the brain using magnetic resonance imaging, IEEE Transactions on Medical Imaging 13 (3) (1994) 470–481.
[2] Y. Kvinnsland, N. Brekke, T.M. Taxt, R. Gruner, Multispectral analysis of multimodal images, Acta Oncologica 48 (2) (2009) 277–284.
[3] L.P. Clarke, R.P. Velthuizen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, R.W. Thatcher, M.L. Silbiger, MRI segmentation: methods and applications, Magnetic Resonance Imaging 13 (3) (1995) 343–368.
[4] R. He, S. Datta, B.R. Sajja, P.A. Narayana, Generalized fuzzy clustering for segmentation of multi-spectral magnetic resonance images, Computerized Medical Imaging and Graphics 32 (5) (2008) 353–366.
[5] X. Llado, A. Oliver, M. Cabezas, J. Freixenet, J.C. Vilanova, A. Quiles, L. Valls, L. Ramio-Torrenta, A. Rovira, Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches, Information Sciences 186 (1) (2012) 164–185.
[6] R. de Boer, H.A. Vrooman, M.A. Ikram, M.W. Vernooij, M.M. Breteler, A. van der Lugt, W.J. Niessen, Accuracy and reproducibility study of automatic MRI brain tissue segmentation methods, Neuroimage 51 (3) (2010) 1047–1056.
[7] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[8] P. Georgiadis, D. Cavouras, I. Kalatzis, A. Daskalakis, G. Kagadis, K. Sifaki, M. Malamas, G. Nikiforidis, E. Solomou, Non-linear least squares features transformation for improving the performance of probabilistic neural networks in classifying human brain tumors on MRI, Lecture Notes in Computer Science 4707 (2007) 239–247.
[9] V.N. Vapnik, Statistical Learning Theory, John Wiley and Sons, Inc., New York, 1998.
[10] C. Valdes Hernandez Mdel, P.J. Gallacher, M.E. Bastin, N.A. Royle, S.M. Maniega, I.J. Deary, J.M. Wardlaw, Automatic segmentation of brain white matter and white matter lesions in normal aging: comparison of five multispectral techniques, Magnetic Resonance Imaging 30 (2) (2012) 222–229.
[11] N. Abdullah, U.K. Ngah, S.A. Aziz, Image classification of brain MRI using support vector machine, in: Proc. IEEE. Conf. on Imaging Systems and Techniques (IST), Malaysia, 2011, pp. 242–247.
[12] Y.C. Ouyang, H.M. Chen, J.W. Chai, C.C. Chen, C.C.C. Chen, S.K. Poon, C.W. Yang, S.K. Lee, Independent component analysis for magnetic resonance image analysis, EURASIP Journal on Advances in Signal Processing 2008 (780656) (2008).
[13] T. Nakai, S. Muraki, E. Bararinao, Y. Miki, Y. Takehara, K. Matsuo, C. Kato, H. Sakabara, H. Isoda, Application of independent component analysis to magnetic resonance imaging for enhancing the contrast of gray matter and white matter, Neuroimage 21 (1) (2004) 251–260.
[14] C.J. James, C.W. Hesse, Independent component analysis for biomedical signals, Physiological Measurement 26 (1) (2005) 15–39.
[15] Y.C. Ouyang, H.M. Chen, J.W. Chai, C.C. Chen, S.K. Poon, C.W. Yang, S.K. Lee, C.I. Chang, Band expansion-based over-complete independent component analysis for multispectral processing of magnetic resonance images, IEEE Transactions on Biomedical Engineering 55 (6) (2008) 1666–1677.
[16] T. Tateyama, Z. Nakao, Y. Chen, Classification of brain matters in MRI by kernel independent component analysis, in: IIH-MSP IEEE Computer Society, 2008, pp. 713–716.
[17] J.W. Chai, C.C. Chen, C.M. Chiang, Y.J. Ho, H.M. Chen, Y.C. Ouyang, C.W. Yang, S.K. Lee, C.I. Chang, Quantitative analysis in clinical applications of brain MRI using independent component analysis coupled with support vector machine, Journal of Magnetic Resonance Imaging 32 (1) (2010) 24–34.

[18] J. Karhunen, S. Malaroiu, M. Ilmoniemi, Local independent component analysis using clustering, International Journal of Neural Systems 10 (6) (2000) 439–451.
[19] C. Bauer, F.J. Theis, W. Baumler, E.W. Lang, Local features in biomedical image clusters extracted with independent component analysis, in: Proc. Int. Joint Conf. Neural Networks (IJCNN), 2003, pp. 81–84.
[20] H. Han, A high performance profile-biomarker diagnosis for mass spectral profiles, BMC Systems Biology 5 (2) (2011) S5.
[21] S.A. Robila, P.K. Varshney, A fast source separation algorithm for hyperspectral image processing, in: Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS), vol. 6, 2002, pp. 3516–3518.
[22] F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, A.F.H. Goetz, The spectral image processing system (SIPS) – interactive visualization and analysis of imaging spectrometer data, Remote Sensing of Environment 44 (2–3) (1993) 145–163.
[23] C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, A.C. Evans, Brainweb: online interface to a 3D MRI simulated brain database, Neuroimage 5 (4) (1997) S425 http://www.bic.mni.mcgill.ca/brainweb
[24] A. Hyvarinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.
[25] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Elsevier, 2009.
[26] C. Ambroise, G.J. McLachlan, Selection bias in gene extraction on the basis of microarray gene-expression data, Proc. Natl. Acad. Sci. U.S.A. 99 (10) (2002) 6562–6566.