

Research Article

Hybrid Technique for Human Spine MRI Classification

Gurjeet Singh^{A*}, Harpreet Kaur^A and Daljit Singh^B

^AComputer Science Engineering, Punjab Technical University, SBBSIET, Jalandhar, India

^CElectronics and Communication Engineering, Lovely Professional University, LPU, Jalandhar, India

Accepted 30 May 2014, Available online 01 June 2014, Vol.4, No.3 (June 2014)

Abstract

MRI has become a proficient appliance for clinical diagnosis and research. For the identification of different diseases; it has become a useful medical modality. The purpose of optimized hybrid technique is to classify the spinal metastases and further these classifications can be used to help surgical planning and further research. Spine is the third most sites for Metastatic disease. The purpose of this study is to measure and characterize the different features of spine MRI which results in the diagnosis of exact spine disorder. Techniques used for research are GLCM and PCA for feature extraction and SVM for classification. Features extracted by GLCM give cent percent accuracy along with SVM-RBF classifier with minimum execution time of 3.7617 seconds. Software used for this research is MATLAB2011a. Choosing the right option for spine diagnosis is often difficult; as it includes life expectancy and balance of risk of surgery.

Keywords: Metastases, PCA, GLCM, SVM, Spine, SCI and Feature.

1. Introduction

Spinal cord related problems are common among many industrialized countries which had enormous impact on patients and their families [Bunheang Tay *et.al*, 2014]. According to health care data; 20 million MRI exams are conducted annually and 50% are related to spine [Jaehan Koh, *et.al*, 2010]. Spinal cord is a tubular shaped organ which carries the sensory signals between brain and nervous system. It contains spinal tracks and neuronal cell bodies. Signals are transmitted through spinal tracks. When any spine injury occur; this track gets disrupted also known as spinal cord injury (SCI) which results in paralysis or stop proper functioning of body. Conventional MRI technique is a widely used technique to diagnose SCI which produce the two dimensional or three dimensional images at microscopic level. It makes use of white and black contrast image to differentiate between hard and soft tissues. Among others; disc herniation, spinal cord compression, tumor and disc displaced are few common spinal disorders.

2. Proposed Work

Our goal is to develop hybrid optimized classifier which accurately detects the disorder and its type based on full protocol MRI of patients. Below figure depicts the systematic overview of proposed hybrid model for automated spine MRI detection and classification. Our method consists of two phases: feature extraction using PCA and GLCM and second phase is classification of

features using SVM. All the images were preprocessed before training and make them of same resolution to reduce biasing effects. Features are extracted from images using PCA and GLCM. These extracted features are then used by SVM.

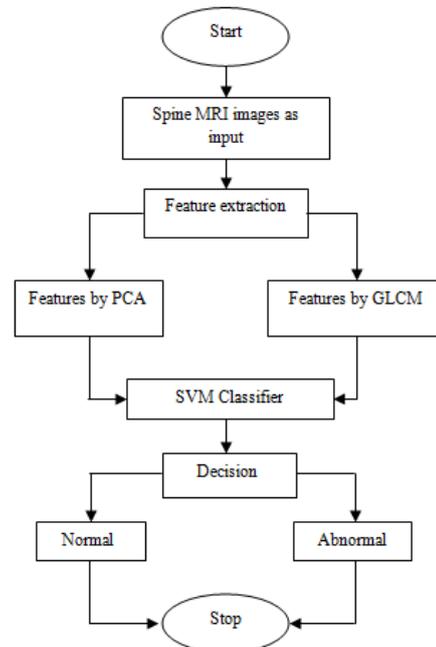


Fig 1: Proposed Methodology

2.1 Image Set

A total of 58 clinical T2 weighted MRI image sets are used. Among them; 18 are normal MRI images and 40 are abnormal MRI images of different patients suffering from

*Corresponding author: Gurjeet Singh

different spine disorders such as disc herniation, spinal cord compression, disc displaced, tumor. All images are of same resolution i.e. 512x512 pixels. Some spine MRI images are shown below.

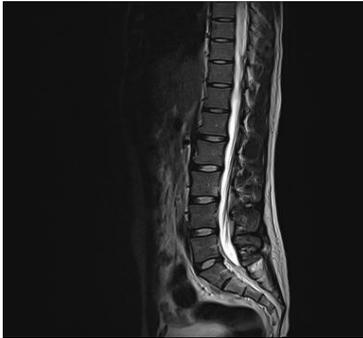


Fig. 2(a) Normal Spine MRI image



Fig. 2(b) Disc displaced



Fig. 2(c) Disc herniation



Fig. 2(d) spinal cord compression

3. Feature Extraction

Features are considered as properties that describe the whole information of an image which helps to solve complex computational tasks for specific applications [K. Kaur *et.al*, 2012]. Feature extraction depicts the process of extracting only valuable data from given data. These extracted features are further used to train the classifier by considering the description of relevant properties of image.

Algorithms used for feature extraction:

3.1 Principal component analysis

It reduces the dataset from correlated variables to uncorrelated variables while discarding redundant information by measuring certain properties to make decision much easier [K. Kaur *et.al*, 2012]. It generates different principal components which are orthogonal to each other and represents different values of variance.

Steps to be kept in mind while using PCA-

Step 1: Reduce the data into two dimensional data to understand the concept behind the operation of PCA.

Arrange the data as a set of N data vectors X_1, \dots, X_N with each X_N representing a single grouped observation of the M variables.

- Write X_1, \dots, X_N as column vectors, each of which has M rows.
- Place the column vectors into a single matrix X of dimensions $M \times N$.

Step 2: Calculate the Mean

For PCA to work properly on data; subtract the mean from each of the data dimensions. Subtracted mean is the average across each dimension.

- Calculate the mean along each dimension $m=1 \dots M$.
- Store the calculated mean values into vector 'u' of dimensions $M \times 1$.

$$u[m] = \frac{1}{N} \sum_{n=1}^N X(m, n)$$

Step 3: Calculate Zero Mean data

- Subtract the empirical mean vector 'u' from each column of the data matrix X .
- Store zero mean data in the $M \times N$ matrix B .

$$B = X - uh$$

Step 4: Calculate the Covariance Matrix 'C'.

- Find the Covariance Matrix C from matrix B with itself.

As given data is two-dimensional thus covariance matrix is also two-dimensional. If the non-diagonal elements of co-variance matrix are positive, both variables increase together. If the non-diagonal components are zero; then variable are independent of each other and are uncorrelated.

Step 5: Find Eigenvectors and Eigen values of Covariance Matrix C which is a square matrix.

- Compute the matrix V of Eigen vectors which diagonalizes the covariance matrix C.

Eigenvectors provide the information about patterns in data and are perpendicular to each other.

Step 6: Rearrange the Eigenvectors and Eigen values in decreasing order.

Step 7: Derive a new dataset.

3.2 Grey Level Co-Occurrence Matrix (GLCM)

GLCM is a tabulation how often different combinations of pixel brightness value occur in an image [D. Singh et.al, 2012]. It considers the relation between two pixels at a time; reference pixel and neighbor pixel. This generated matrix is used to extract statistical measures of an image while characterizing the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image. Below example shows how GLCM works on an image.

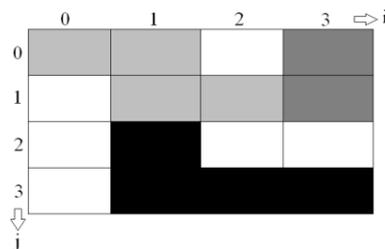


Fig 3: Test Image

2	2	3	1
3	2	2	1
3	0	3	3
3	0	0	0

Fig 4: Pixel values corresponding to Test Image

Now the pixel pair (0, 0) in figure 4 repeats two time hence, in GLCM matrix in figure 5 (0, 0) element has value 2. For pair (0, 1) there is no pixel for (0, 1) in figure 4, so element value is 0. Note that we have used default offset value [0 1] so it will check pairs in right hand direction only. Similarly GLCM matrix is created by same methodology by checking pair values over the whole image.

2	0	0	1
0	0	0	0
0	1	2	1
2	1	1	1

Fig 5: Gray Level Concurrence Matrix

In this paper, authors make use of the features extracted from images using PCA and GLCM algorithms to classify the images between normal and abnormal.

3.3 Features extracted from images

Mean: It gives the average value of all the data in an image along with that it depicts the brightness of an image.

$$\frac{\sum_{i,j}^{N-1} P(i, j)}{N - 1}$$

Energy: It returns the sum of squared elements of an image.

$$\sum_{i,j=0}^{N-1} P(i, j)^2$$

Entropy: It is a measure of randomness.

$$\sum_{i,j=0}^{N-1} P(i, j)[-ln(P(i, j))]$$

Contrast: It provides the measure of how sharp the structural variations in the image are.

$$\sum_{i,j=0}^{N-1} |i - j|^2 P(i, j)$$

Homogeneity: It returns the value that measures the closeness of the distribution of the elements in the GLCM to the GLCM diagonal.

$$\sum_{i,j} \frac{p(i, j)}{1 + |i - j|}$$

Correlation: It depicts the value that how a pixel is correlated with its neighbor pixel for the whole image. For constant image; its value is Nan. Its range is [-1,1].

$$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j}$$

4. Classification

In this phase; differences between extracted features from normal and abnormal images are identified and classifier decides on the basis of these differences. Classification means to examine the different numerical properties of an image and re-organize the data in two different classes. Classification is based upon learning techniques. Different learning techniques are:

- Supervised learning
- Unsupervised learning

This process employs two phases: training and testing. In training phase, characteristic properties of image features are isolated and a unique description of each classification category is created. In testing phase, these features space partitions are used to classify image features.

4.1 Support Vector Machine

As discussed above; SVM is also a supervised learning based binary classifier which performs classification only between two classes by constructing a hyper-plane in high-dimensional feature space. As SVM classifies between two classes 0 and 1, Class 0 is for normal images and class 1 is for abnormal images. The reason behind using SVM is that it gives better performance among other binary classifiers where only two classes are to be classified.

Hyper-plane can be represented by equation-

$$w \cdot x + b = 0 \tag{1}$$

w is weight vector and normal to hyperplane.
 b is bias or threshold.

A. Linear Separable Binary Classifier

Consider N training points, where each input x_i has ‘ A ’ attributes and is in one of two classes [C.Cortes et.al, 1995]. $y_i = +1$ or -1 , i.e. training data is of the form: (x_i, y_i) , $i=1, 2, 3 \dots N$; $y_i \in \{+1, -1\}$.

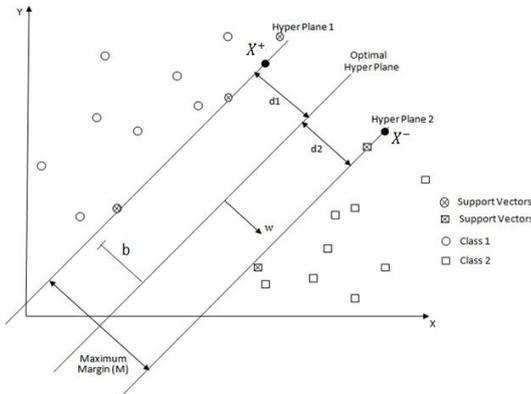


Fig 6: Linear separable binary classification

The main purpose of using Support Vector Machine (SVM) is to orientate hyperplane in such a way as to be as far as possible from the closest members of both classes. Training data can be described by:

$$(w \cdot x_i + b) \geq l \text{ for } y_i = +1 \tag{2}$$

$$(w \cdot x_i + b) \leq -l \text{ for } y_i = -1 \tag{3}$$

Equation (2) and (3) can also be written as:

$$y_i (w \cdot x_i + b) - l \geq 0 \quad \forall i \tag{4}$$

Consider the points as shown in fig.6 that lie closest to the separating hyper-plane, i.e. the Support Vectors; then the two planes; hyper-plane 1 and hyper plane 2 lie on points which can be described as:

$$\text{For Hyper Plane 1, } w \cdot x_i + b = +1 \tag{5}$$

$$\text{For Hyper Plane 2, } w \cdot x_i + b = -1 \tag{6}$$

Mathematics geometry analysis defined a distance from point P (m, n) to a line $ax + by + c = 0$ as:

$$\frac{(am + bn + c)}{\sqrt{a^2 + b^2}}$$

In the same way, the distance from a point X^+ on HyperPlane1 to Optimal HyperPlane is given by $d1$ as

$$d1 = \frac{1}{||w||} \tag{7}$$

Similarly from a point X^- on HyperPlane2 to Optimal HyperPlane is given by $d2$ as:

$$d2 = \frac{1}{||w||} \tag{8}$$

By adding (7) and (8) we will get margin width

$$M = \frac{2}{||w||} \tag{9}$$

Simple vector geometry in equation (9) shows that the margin is equal to $2/||w||$ and maximizing it subject to the constraint in (4) is equivalent to finding:

$$\begin{aligned} \min ||w|| \\ \text{Subject to } y_i (w \cdot x_i + b) - l \geq 0 \quad \forall i \end{aligned} \tag{10}$$

Minimizing $||w||$ is equivalent to minimizing $\frac{1}{2} ||w||^2$ (the factor $\frac{1}{2}$ being used for mathematical convenience) and the use of this term makes it possible to perform Quadratic Programming (QP) optimization later on. We therefore need to find:

$$\begin{aligned} \min \frac{1}{2} ||w||^2 \\ \text{st } y_i (w \cdot x_i + b) - 1 \geq 0 \quad \forall i \end{aligned} \tag{11}$$

In order to cater for constraints in this minimization [D.Singh et al, 2012], we need to allocate them Lagrange’s multiplier (LM) α , where $\alpha_i \geq 0 \quad \forall i$; the primal form of the function is:

$$L_P = \frac{1}{2} ||w||^2 - \alpha [y_i (w \cdot x_i + b) - 1] \quad \forall i \tag{12}$$

To get the optimal hyper-plane, the resulting classifier and objective functions are:

$$f_\alpha(x) = \text{sgn}[\sum_{i=1}^N y_i \alpha_i (x_i \cdot x) + b] \tag{13}$$

The equation (12) gives the new formulation which being dependent on α , thus need to maximize it:

$$\begin{aligned} L_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \\ \text{Subject to } \alpha_i \geq 0 \quad \forall i \end{aligned} \tag{14}$$

This is new formulation referred to as the dual form of the primary L_P .

Having moves from minimizing L_P to maximizing L_D , find:

$$\begin{aligned} \max_\alpha [\sum_{i=1}^N \alpha_i - \frac{1}{2} \alpha^T H \alpha] \\ \text{st } \alpha_i \geq 0 \quad \forall i, \quad \sum_{i=1}^N \alpha_i y_i = 0 \end{aligned} \tag{15}$$

This is convex quadratic optimization problem thus need to run a quadratic programming solver which will return α and by differentiating equation (12); it gives ‘ w ’:

$$w = \sum_{i=1}^N y_i \alpha_i x_i$$

B. Linear non-separable binary classifier

When the training dataset is not linearly separable as shown in fig.7; optimal separating hyper-plane is found by solving an optimization problem relaxed. By introducing a set of slack variable ξ_i and penalization for cases that are misclassified or inside the margin.

The task for finding the optimal hyper plane is to minimize the following objective function,

$$\min \frac{1}{2} ||w||^2 + C \sum_{i=1}^N \xi_i, \quad i=1, 2, 3, \dots, N$$

subject to $y_i (w \cdot x_i + b) \geq 1 - \xi_i$

$$\xi_i \geq 0 \quad i=1, 2, \dots, N$$

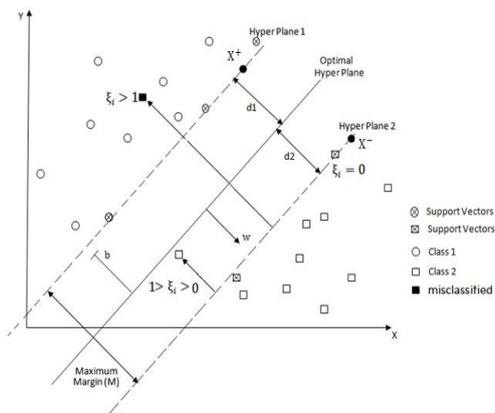


Fig 7: Linearly non-separable binary classification

‘C’ is regularization parameter and controls the trade-off between the slack variable penalty and the size of the margin.

1. Small C allows constraints to be easily ignored i.e. wide margin. It allows a lot of samples that are not in ideal position.
2. Large C allows constraints hard to ignore i.e. narrow margin. It allows very few samples that are not in ideal position.
3. C=∞ enforces all constraints i.e. hard margin.

C. Non-Linear SVM

In both above discussed cases of SVM classifier also shown in fig.6 and fig.7, straight line or hyper-plane is used to distinguish between two classes. But datasets or data points are always not separated by drawing a straight line between two classes. For example the data points in the fig.8 can't be separable by using both SVM's discussed cases.

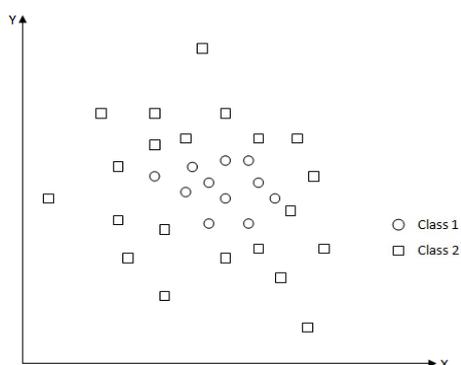


Fig 8: Non-linear data points

So, Kernel functions are used with SVM classifier. Kernel function provides the bridge between from non-linear to

linear. Basic idea behind using kernel function is to map the low dimensional data into the high dimensional feature space where data points are linearly separable. There are many types of kernel function but Kernel functions used in this research work are given below:

1. Radial basis function (RBF)
2. Linear
3. Quadratic

5. Results

After classification it had been concluded that features extracted by PCA when classified with SVM using kernel functions Linear, Quadratic and RBF; then out of 15 images only 12,13 and 13 were accurately classified resp. and rest are misclassified.

Similarly, features extracted by GLCM when classified with SVM using same kernel functions; all images were accurately classified. In spite of these kernels, other functions were also used; but here authors discuss only those functions which give better performance Results in tabular form are given below.

Table I Range for extracted features

Features	Normal range	Abnormal range
Contrast	0.00-0.0125	0.0182-0.1277
Homogeneity	0.0- 0.02	-0.0002-0.05
Energy	0.9995-1.000	0.9947-0.9993
Co-relation	0.9998-1.0000	0.9979-0.9997
Entropy	0.0-0.00045	0.0025-0.0149

6. Graphs

This section represents the graphical results of proposed technique for both PCA and GLCM using SVM and its different kernel functions. Results of both algorithms are compared on the basis of accuracy and execution time.

In below graphs; red points denotes data points for normal images, green points denotes data points for abnormal images and blue points denotes data points for images in test dataset. Black circles are support vectors and lines drawn are hyper-planes.

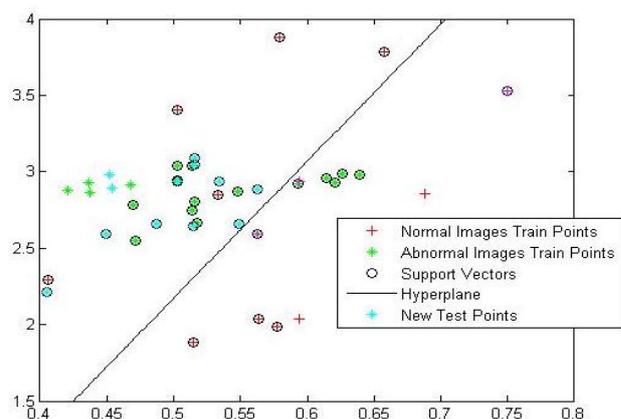


Fig 9: Classification with contrast and energy by PCA using SVM-linear kernel

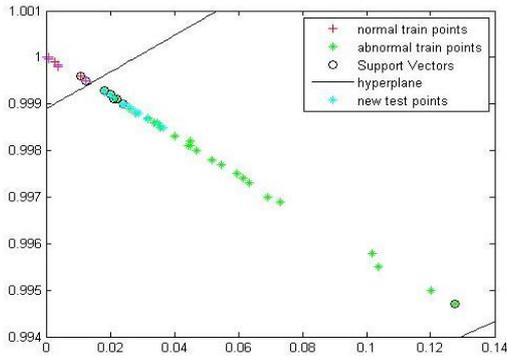


Fig 10: Classification with contrast and energy by GLCM using SVM-Quadratic kernel function

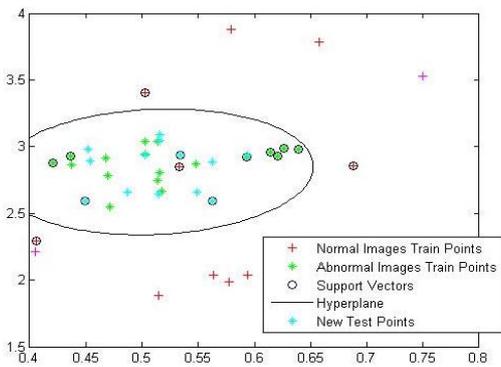


Fig 11: Classification with contrast and energy by PCA using SVM-Quadratic kernel

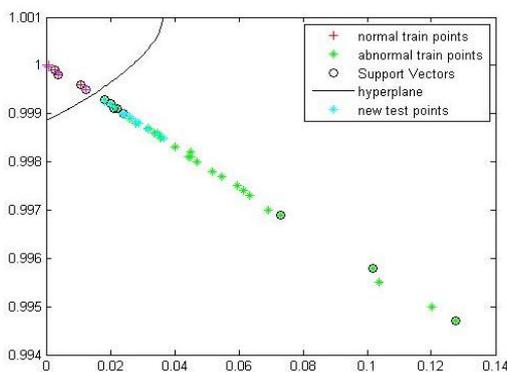


Fig 12: Classification with contrast and energy by GLCM using SVM-RBF kernel function

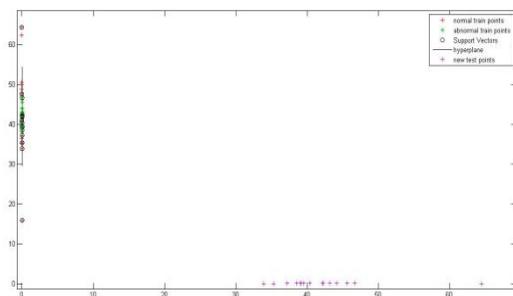


Fig 13: Classification with mean and entropy by GLCM using SVM-RBF kernel function

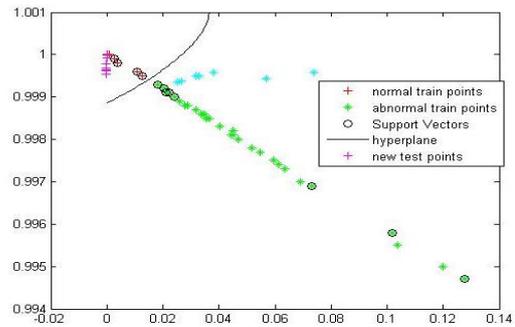


Fig 14: Classification with homogeneity and correlation by GLCM using SVM-RBF

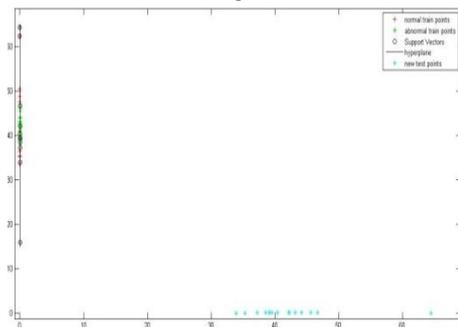


Fig15: Classification with mean and entropy by GLCM using SVM-Linear kernel function

Table II Combined results for different extracted features with accuracy and execution time

Feature extraction technique	Extracted Feature	SVM Kernel	Accuracy (%)	Execution time (sec)
PCA	Contrast & Energy	Linear	80	4.181
		Quad	86.66	4.074
		RBF	86.66	4.234
GLCM	Contrast & Energy	Linear	100	3.763
		Quad	100	3.869
		RBF	100	3.761
GLCM	Correlation & Homogeneity	Linear	46.66	4.88
		Quad	46.66	4.809
		RBF	46.67	5.031
GLCM	Mean & Entropy	Linear	86.66	4.513
		Quad	13.33	4.459
		RBF	13.33	4.46

Conclusions

In this paper; authors have developed a hybrid technique which performs the classification between normal and abnormal human spine MRI images. Different features are extracted from spine MRI images using two different extraction techniques PCA and GLM resp. Output of these algorithms is further fed to SVM classifier to get it train so that it can accurately classify the test images between

normal and abnormal. The more data is trained; more accuracy will be given. GLCM extracted features when classified with SVM-Linear, SVM-Quad, SVM-RBF gives 100% accuracy with execution time of 3.7635, 3.8691, 3.7617 seconds resp. in spite of it; when PCA is used with SVM it leads to misclassification of some images. It gives the results but the final decision is made after consultation with medical specialist.

References

- C.Cortes, V. Vapnik (1995), in *Machine Learning*, pp. 273-297
- El-Sayed Abdal Wahid et.al, Abraham Al Emam, Ami Badr (2012), Feature selection for Cancer Classification: An SVM based approach, *International Journal of Computer applications* (0975-8887), Volume 46- No.8.
- Isam Abu-Qasmeih (2013), Multiclass multi label classification and detection of lumbar intervertebral disc degeneration MR images using decision tree classifiers, *Computer Engineering and Intelligent Systems*, ISSN 2222-2863, Vol.4 No.9
- N. Kumaran (2013), MRI Image retrieval using Gabor Wavelet Based Texture Features, *International Journal of advanced research in Science and Technology*. Vol.2, Issue 1, ISSN 2320-1126, pp 40-45
- Ryszard S.Choras (2007), Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems, *International Journal of Biology and Biomedical Engineering*. Issue 1, Vol.1.
- EL-Sayed, EL- Dahshan, Abdul- Badeeh M. Salem, Tamer H.Yousin (2009), A Hybrid Technique for Automatic MRI Brain Images Classification. *Studia Univ. Babeş.Bolyai*, Volume LIV, Number 1.
- El-Sayed Abdal Wahed, Abraham Al Emam, Ami Badr (2012), Feature selection for Cancer Classification: An SVM based approach, *International Journal of Computer applications* (0975-8887), Volume 46- No.8.
- Pranati Das, Sudipta Muhodhyapay (2010), Cervical Spine image retrieval with semantic shape features. *International Journal of computer vision and robotics*. Volume 1 Issue 2, pp. 136-146
- Bing-Yu Sun, Xiao-Ming Zhang, Ru-Jing Wang (2008), Training SVMs for Multiple Features Classification Problems, *Fourth International Conference on National Computation*, 978-0-7695-3304-9/08. IEEE.
- Ghosh.S (2011), Computer aided diagnosis for lumbar MRI using heterogenous classifiers, *IEEE International Symposium on Biomedical Imaging*, ISSN:1945-7928 , pp 1179-1182.
- Ghosh.S (2011), Composite features for automatic diagnosis for intervertebral disc herniation from lumbar, *Engineering in Medical and Biological society, IEEE International Conference*, ISSN: 1557-170X, pp 5068-5071.
- D.Singh, K.Kaur (2012), Classification of abnormalities in Brain MRI Images using PCA, GLCM and SVM, *International journal of engineering and advance technology*. ISSN: 2249-8958, Vol.X Issue X, pp- 243-248.
- K.Kaur, D.Singh (2012), Brain CT-Scan Images Classification Using PCA, Wavelet Transform and K-NN, *International Journal of Advance Research in Computer Science and Electronics Engineering*, ISSN: 2277-9043, Vol. I, Issue I, pp- 109-115.
- Jeahan Koh.(2010), Diagnosis of Disc Herniation Based on Classifiers and features generated from Spine MRI images, *Medical Imaging; Computer aided diagnosis*. Vol. 7624.
- Bunheang Tay (2014), A Machine Learning Approach for Specification of Spinal Cord Injuries Using Fractional Anisotropy Values Obtained From Diffusion Tensor Images, *Computational and Mathematical Methods in Medicine*, Vol.2014, Article ID: 276589.