Brain MR Image Classification Using Least Squares Support Vector Machine

[#1] Mrs.Sapna N.Gaikwad, [#2] Prof. Dr. D. S. Bormane,

[#1] *Department of Electronics Engineering, J.S.P.M. College of Engineering, University Of Pune, Pune , India*

[#2] *Principal, J.S.P.M. College of Engineering, University Of Pune, Pune , India*

Abstract— This research paper proposes an intelligent classification technique to identify tumor. The manual interpretation of tumor based on visual examination by Radiologist/physician may lead to missing diagnosis when a large number of data are analyzed. To avoid the human error, an automated intelligent classification system is proposed which caters the need for classification of medical image after identifying the volume normal and abnormal images for tumor identification. In this research work, advanced classification techniques based on Least Squares Support Vector Machines (LS-SVM) are proposed and applied to medical image classification using features derived from images.

Index Terms—. Classification**,** LS-SVM, SVM, RBF Kernel, K means, PCA,Single Value Decomposition

———————————— ◆ ————————————

1 INTRODUCTION

THIS field of medical imaging gains its importance with increase in the need of automated and efficient diagnosis in a short period of time. Computer and Information Technology are very much useful in medical image processing, medical analysis and classification. In, medical images there are normal and abnormal data of images are use for classification detecting tumor. Vapnik introduced the SVM (support vector machine). SVM classifier was used for medical images classification with statistical features. The latest development in data classification research has focused more on Least Squares Support Vector Machines (LS-SVMs) because several recent studies have reported that LS-SVM generally are able to deliver higher classification accuracy than the other existing data classification algorithms.LS-SVM introduced by suykens.

In this paper, the potential benefit of using an LS-SVM based approach is investigated for the automated classification of medical images. This is for separating normal and abnormal images from data collection containing tumor. The purpose is to perform segmentation process for tumor detection. Segmentation is done by using K-means algorithm. The categorization of medical images is done using statistical features of images such as mean, variance, and co-occurrence based textural features of images such as energy, contrast and correlation.

The motivation behind this paper is to develop a machine classification process for evaluating the classification performance of LS-SVM classifier with RBF kernel to this problem in terms of statistical performance measure.

The paper is organized as follows. Preliminaries dealing with LS-SVM techniques are presented in Section 2. Section 3 discusses the proposed methodology using LS-SVM for MRI brain images classification. Section 4 highlights the implementation of proposed methodology. Results and outputs after

Implementation of the proposed approach is given in Section 5. Finally, the conclusion is presented in Section 6.

2 REVIEW OF LS-SVM LEARNING FOR CLASSIFICATION

The Support Vector Machine algorithm was first developed in 1963 by Vapnik and Lerner's and Vapnik and Chervonenkis as an extension of the Generalized Portrait algorithm. This algorithm is firmly grounded in the framework of statistical learning theory – Vapnik Chervonenkis (VC) theory, which improves the generalization ability of learning machines to unseen data . In the last few years Support Vector Machines have shown excellent performance in many real-worlds applications including hand written digit recognition, object recognition, speaker identification , face detection in images and text categorization . SVM is a classification algorithm based on linear and non linear classification. Dot-products can be computed efficiently in higher dimensional space. The dominant feature which makes SVM very attractive is that classes which are nonlinearly separable in the original space can be linearly separated in the higher dimensional feature space. Thus SVM is capable of solving complex nonlinear classification problems. Important characteristics of SVM are its ability to solve classification problems by means of convex quadratic programming (QP) and also the sparseness resulting from this QP problem. The learning is based on the principle of structural risk minimization. Instead of minimizing an objective function based on the training samples (such as mean square error), the SVM attempts to minimize the bound on the generalization error (i.e., the error made by the learning machine on the test data not used during training). As a result, an SVM tends to perform well when applied to data outside the training set. SVM achieves this advantage by focusing on the training examples that are most difficult to classify. These "borderline" training examples are called support vectors. A least squares version of SVM (LS-SVM) is introduced by Suy-

kens in with the idea of modifying Vapnik's SVM formulation by adding a least squares term in the cost function. This variant circumvents the need to solve a more difficult QP Problem and only requires the solution of a set of   linear equations. This approach significantly reduces the complexity and computation in solving the problem.

In this paper, we treat slice classification as a two class pattern classification   problem. We apply all the MRI slices to classifier to determine whether the tumor is present or not. We refer to these two classes throughout as "normal" and "abnormal" slices. Let vector $x \in R$ denote a pattern to be Classified, and let scalar y denote its class label (i.e., $y \in +1, 1$) In addition, let $\{(x_i, y_i), i = 1, 2, \ldots, l\}$ denote a given set of l training examples. The problem is now to construct a classifier [i.e., a decision function f(x)] that can correctly classify an input pattern x that is not necessarily from the training set.

2.1 Linear SVM Classifier

Let us begin with the simplest case, in which the training patterns are linearly separable.
That is, there exists a linear function of the form

$$f(x) = w^T x + b$$

such that for each training example xi, the function yield

$$f(x_i) \geq 0 \text{ for } y_i = +1, \text{ and } f(x_i) < 0 \text{ for } y_i = -1.$$

In other words, training examples from the two different classes are separated by the hyperplane $f(x) = w^T x + b = 0$, where w is the unit vector and b is a constant. For a given training set, while there may exist many hyperplanes that maximize the separating margin between the two classes, the SVM classifier is based on the hyperplane that maximizes the separating margin between the two classes (Figure 1). In other words, SVM find s the hyperplane that causes the largest separation between the decision function values for the "borderline" examples from the two classes. Mathematically, this hyperplane can be found by minimizing the cost function:

$$J(W) = \frac{1}{2}W^T W = \frac{1}{2}\|W\|^2$$

subject to the separability constraints

$$W^T X_i + b \geq +1, \text{ for } y_i = +1$$

or

$$W^T X_i + b \leq -1, \text{ for } y_i = -1; i = 1, 2, \ldots, l.$$

In Figure 1, SVM classification with a hyperplane that minimizes the separating margin between the two classes are indicated by data points marked by "X" s and "O"s. Support vectors are elements of the training set that lie on the oundary hyperplanes of the two classes.
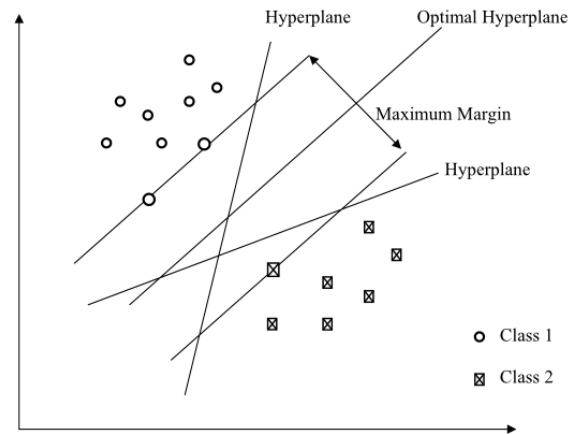

Figure 1. SVM classification.

2.2 Non Linear SVM Classifiers

The linear SVM can be readily extended to a nonlinear classifier by first using a nonlinear operator (φ) to map the input pattern x into a higher dimensional space H. The nonlinear SVM classifier so obtained is defined as

$$f(x) = W^T \phi(X) + b$$

Which is linear in terms of the transformed data Φ(x), but nonlinear rs in terms of the original Data $x \in R$ following nonlinear transformation, the parameters of the decision function f(x) are determined by the following minimization:

$$\min J(W, \xi) = \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{l} \xi_i$$

subject to

$$y_i(w'\phi(X_i) + b) \geq 1 - \varepsilon_i,$$
$$\varepsilon_i \geq 0; i = 1, 2, \ldots, l.$$

The data with linear separability may be analyzed with a hyper plane, and the linearly non separable data are analyzed with kernel functions such as higher order polynomials, Gaussian RBF and tan-sigmoid described as:

$$K(x, y) = x^T y \text{ (Linear SVM)}$$

$$K(x, y) = (\tau + x^T y)^d \text{ (Polynomial SVM of degree d)}$$

$$K(x, y) = \exp(-\|x - y\|^2 / \sigma^2) \text{ (RBF kernel)}$$

$$K(x, y) = \tanh(k_1 x^T y + k_2) \text{ (Tan-sigmoid kernel)}$$

The output of an SVM is a linear combination of the training examples projected onto a high- dimensional feature space through the use of kernel functions.

Least Squares SVM classifier The advantage of nonlinear SVM classifiers is the ability to solve classification problems by means of convex quadratic programming, as well as the

sparseness as a result of this QP problem. Suykens proposed the idea of modifying the Vapnik's SVM formulation by adding a least squares term in the cost function, which transformed the problem from solving a QP problem to solving a set of linear equations. This approach significantly reduces the cost in complexity and computation time for solving the problem. Suykens formulated the SVM modification as follows:

$$\min_{w,b,e} J_p = \frac{1}{2} W^T W + \gamma \sum_{k=1}^{N} e_k^2$$

such that

$$y_k[W^T \varphi(X_k) + b] = 1 - e_k, \quad k = 1, \ldots, N.$$

which corresponds to a classifier in the primal space

$$y(x) = \text{sign}[w^T \varphi(x) + b]$$

where $\varphi(\cdot)$ is the mapping to the high dimensional feature space as in the standard SVM. The Vapnik formulation is modified here at two points. First, instead of inequality constraints, equality constraints are used, where the value 1 at the right hand side is considered as a target value instead of a threshold value. Error variable ek allows some tolerance of misclassification in the case of overlapping distributions. This error variables function is similar to the slack variable ξk in SVM formulation. Second, a squared loss function is taken for this error variable. As explained below, these modifications will greatly simplify the problem.

In the case of a linear classifier the primal problem can be solved easily, however w in general might become infinite dimensional. Therefore let us derive the dual problem for the LS-SVM nonlinear classifier formulation. The Lagrangian for the problem is

$$L(w, b, e; \alpha) =$$

$$J_p(w, e) - \sum_{k=1}^{N} \alpha_k \{ y_k [w^T \varphi(x_k) + b] - 1 + e_k \}$$

where the αk values are the Lagrange multipliers, which can be positive or negative, due to the equality constraints.
The conditions for optimality yield

$$\begin{cases} \dfrac{\partial l}{\partial w} = 0 \rightarrow w = \sum_{k=1}^{N} \alpha_k y_k \varphi(X_k) \\ \dfrac{\partial l}{\partial b} = 0 \rightarrow \sum_{k=1}^{N} \alpha_k y_k = 0 \\ \dfrac{\partial l}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k \\ \dfrac{\partial l}{\partial \alpha_k} = 0 \rightarrow y_k [w^T \varphi(x_k) + b] - 1 + e_k = 0 \end{cases}$$

for k = 1, ...,N. These can be written as a linear system

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma^I & -I \\ Z & Y & I & 0 \end{bmatrix} \begin{bmatrix} w \\ b \\ e \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vec{i} \end{bmatrix}$$

Where

$$Z = [\varphi(x_1)^T y_1 \cdots \varphi(x_N)^T y_N]^T,$$
$$Y = [y_1 \ldots y_N]^T, \overline{1 = [1 \ldots 1]^T},$$
$$e = [e_1 \ldots e_N]^T, \alpha = [\alpha_1 \ldots \alpha_N]^T.$$

Elimination of w and e gives

$$\begin{bmatrix} 0 & Y^T \\ Y & Z^T Z + I/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{i} \end{bmatrix}$$

Mercer's condition can be applied to the matrix

$$\Omega = Z^T Z \quad \text{with:}$$

$$\Omega_{kl} = y_k y_l \varphi(x_k)^T \varphi(x_l)$$
$$= y_k y_l K(x_k, x_l), k, l = 1, \ldots N.$$

The classifier in dual space takes the form:

$$y(x) = \text{sign} \left[\sum_{k=1}^{N} \alpha_k y_k K(x, x_k) + b \right]$$

similar to the standard SVM.

A chosen kernel function should be a positive definite function and satisfy the Mercer condition. All comments on the kernel functions can be equally applied to the use of kernels in the LS-SVM context. However, in this work we focus on the use of linear and RBF kernels.

LS-SVM can be efficiently estimated using iterative methods. In order to make an LS-SVM model, we need two extra parameters: γ(gama) is the regularization parameter, determining the trade-off between the fitting error minimization and smoothness. In the common case of the RBF kernel, $\sigma^2 (\text{sig}^2)$ is the bandwidth.

3 PROPOSED METHODOLOGY

The proposed methodology of classifying MR image slices of human brain is shown in Figure 2. The method uses the steps of feature extraction and classification. Significant difference between tissue types, observed in variety of textural meas-

urements in MR image, is used for this classification. The various measurements based on statistical and co-occurrence matrix textural features from the MR images are given as input to the classifiers for training. If the features of new slices are given as input, the trained classifier can able to classify it. The results of Least Squares Support Vector Machine are analysed to detect the tumour type.

3.1 Medical Image / MR Image (Input Image Data)

Magnetic Resonance Imaging (MRI) uses magnetic energy and radio waves to create images ("slices") of the human body. MR imaging measures the magnetic properties of nuclei within the body tissues. The energy absorbed by the nuclei is then released, returning the nuclei to their initial state of equilibrium and this transmission of energy by the nuclei is ob

served as the MRI signal.MR images are generated by the resonating nuclei for each spatial location. The image gray level in MRI mainly depends on three tissue parameters viz., proton density (PD), spin-lattice (T1) and spin-spin (T2) relaxation time. Generally, for most of the soft tissues in the body, the proton density is very homogenous but may exhibit higher intensity for gray matter. T1 and T2 are sensitive to the local environment; they are used to characterize different Tissue types. T1, T2 and PD type images are mostly used by different researchers for different MR applications. Recently, the FLAIR sequence has replaced the PD image . FLAIR images are T2 weighted with the CSF signal suppressed. T1 shows higher intensity for white matter, T2 presents higher intensity for cerebrospinal fluid. In practice, MR images represent all these three properties with variable weightings, these images are produced as T1 weighted, T2 weighted and T2 FLAIR weighted for each spatial plane. Though several orientations of magnetic resonance imaging are possible, the axial orientation is frequently used in segmentation.

3.2 Segmentation

Segmentation plays an important role in medical image processing. Segmentation of medical image is the division or separation of the image into disjoint regions of similar attribute. We proposed a methodology that integrates K-Means clustering segmentation algorithm for medical image segmentation.K-means algorithm proposed for large datasets and to find initial centroid. An algorithm is described for segmenting medical image (MRI Images) into K different tissue types, which include gray, white matter and CSF, and other abnormal tissues. Medical images considered can be either scale- or multivalued. Each scale-valued image is modeled as a collection of regions with slowly varying intensity plus a white Gaussian noise. The proposed algorithm is an adaptive K-means clustering algorithm for 3-dimensional and multivalued images.Each iteration consists of two steps:estimate mean intensity at each location for each type, and estimate tissue types, Its performance is tested using patient data.

K-Means Algorithm
K-means algorithm is under the category of Squared Error-Based Clustering (Vector Quantization) and it is also under the

category of crisp clustering or hard clustering. K-means algorithm is very simple and can be easily implemented in solving many practical problems. Steps of the K-means algorithm are given below.

1. Choose k cluster centres to coincide with k randomly chosen patterns inside the hyper volume containing the pattern set.(C)

2. Assign each pattern to the closest cluster center. (Ci, i = 1,2,. . . . C)

3. Recompute the cluster centers using the current cluster memberships.(U)

$$u_{ij} = \left\{ 1 \text{ if } \| x_j - c_i \|^2 \leq \| x_j - c_k \|^2, \text{ for each } k \neq i, 0 \text{ otherwise} \right|$$

4. If a convergence criterion is not met, go to step 2 with new cluster centers by the following equation, i.e., minimal decrease in squared error.

$$Ci = \frac{1}{|G_i|} \sum_{k, x k \in Gi} x_k$$

Where, $|G_i|$ is the size of G_i
or,

$$|G_i| = \sum_{j=1}^{n} u_{ij}$$

The performance of the K-means algorithm depends on the initial positions of the cluster centres. This is an inherently iterative algorithm. And also there is no guarantee about the convergence towards an optimum solution. The convergence centroids vary with different initial points. It is also sensitive to noise and outliers. It is only based on numerical variables.

3.3 Feature extraction

The purpose of feature extraction is to reduce the original data set by measuring certain properties ,or features, that distinguish one input pattern from another.

The extracted features provide the characteristics of the input type to the classifier by considering the description of the relevant properties of the image into a feature space. Most of the tumor is heterogeneous tissues and the mean values of relaxation times are not at all sufficient to characterize the heterogeneity of the different tumor types .

An alternative approach, which is being investigated within the framework of this study, is to apply texture analysis to the T1, T2 and T2 FLAIR images to describe quantitatively the brightness and texture of the images. Texture analysis covers a wide range of techniques based on first- and second order image texture parameters.

In the present study the statistical features based on image intensity like mean & variance and features from gray level co-occurrence matrices (GCMs) such as contrast, energy, and
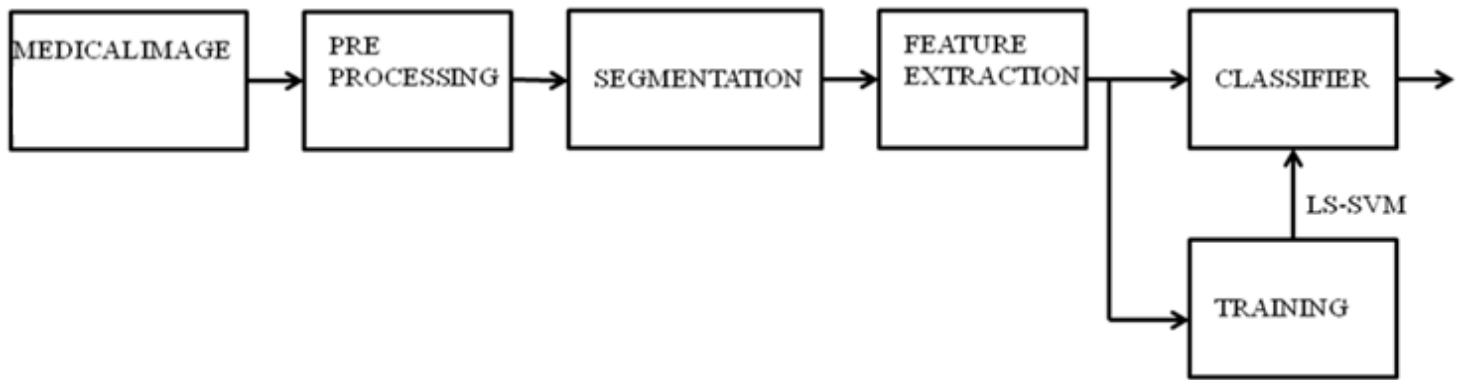
Figure II. Medical Image Classification – Proposed Methodology

Correlation are used to investigate the adequacy for the disrimination of normal and abnormal patient.

The following statistical features are computed as described in the following equation . Let x(i,j)be the image intensity for the location (i, j).

Mean (M) $\quad \overline{x} = \dfrac{1}{X*Y}\sum_{i=1}^{X}\sum_{j=1}^{Y} x(i,j)$

Variance (V) $V = \dfrac{1}{X*Y}\sum_{i=1}^{X}\sum_{j=1}^{Y}\left(x(i,j)-\overline{x}\right)^2$

Where X & Y are number of pixels available in rows and columns of images respectively. The spatial coordinates and intensities of   the extracted pixels were preserved for constructing the Gray Level Co-Occurrence Matrix (GCM). The choice of Haralick features based on GCMs was made considering their proven applicability to analyze objects with irregular outliners. The GCMs are constructed by mapping the gray level co-occurrence probabilities based on spatial relations of pixels in different angular directions.

A GCM P(i, j) reflects the distribution of the probability of occurrence of a pair of gray levels (i,j) given the spacing between the pixels is x∆ and y∆ in the x and y dimensions respectively. The symmetric co-occurrence matrixes are estimated by averaging the matrixes $\quad P(i,j/\Delta x,\Delta y)\quad$ and $P(i,j/-\Delta x,-\Delta y)$,   so eliminating the distinction between opposite offset directions. Four angles namely 0, 45, 90, 135 as well as a predefined offset distance of one pixel in the formation of symmetric co-occurrence matrices are considered. A pixel offset distance of one is preferred to ensure a large numbers of co-occurrences derived from images. From these four GCM matrices, set of features are computed (called feature vectors), and four texture   measures in the present work as in were computed as described below,

Difference moment: A measure of contrast

$$\text{Contrast} = \sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2\left(\frac{P(i,j)}{R}\right)$$

where N is the number of gray levels, equal to 256 for images in the present study. R is equal to the total number of pixel pairs used for the calculation of texture features in the specified angular direction.

Energy: A measure of homogeneity

$$\text{Energy} = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{P(i,j)}{R}\right)^2$$

Inverse Difference Moment: A measure of local homogeneity
Inverse Difference Moment

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{[P(i,j)/R]}{1+(i-j)^2},\quad i\neq j.$$

Correlation: A measure of linear dependency of brightness.

$$\text{Correlation} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}[ijP(i,j)/R]-\mu_x\mu_y}{\sigma_x\sigma_y}.$$

where,

$$\mu_x = \sum_{i=1}^{N}i\sum_{i=1}^{N}P(i,j)/R$$

$$\mu_y = \sum_{j=1}^{N}j\sum_{i=1}^{N}P(i,j)/R$$

$$\sigma_x^{\ 2} = \sum_{i=1}^{N}(i-\mu_x)^2\sum_{j=1}^{N}P(i,j)/R$$

$$\sigma_y^{\ 2} = \sum_{i=1}^{N}(j-\mu_y)^2\sum_{i=1}^{N}P(i,j)/R$$

In the above expressions, μ_x, σ_x, μ_y and σ_1 are the mean and standard deviation values of GCM values accumulated in the x and y directions, respectively.

4   IMPLEMENTATION OF PROPOSED METHODOLOGY

4.1 MR image data

Each axial brain slice used in this work consists of three feature images: T1 weighted, T2 weighted and T2 FLAIR weighted.The 43 image slices (29 abnormal slices and 14 normal slices) of 03 abnormal patient volumes were considered in this work.

Training & Testing Data

The MRI image slices were grouped into two classes, namely normal and abnormal depending on the tumor present in the slice. The MRI data set contains 43 slices (29 abnormal slices and 14 normal slices) and from which all 43 slices are grouped to have training of classifier and 20 images are grouped to have testing of classifier(13 abnormal and 07 normal images).

4.2 Feature extraction

In this paper, for every medical image of each patient, various statistical features and features for the four GCMs described above are computed. Instead of using angularly dependent features directly in GCM, the average (which are invariant under rotation) of each four co-occurrence features over the four angular directions are used, thus resulting in a total of 5features per image. The features need to be normalized so that no one feature dominates the others. To facilitate training, the feature vectors are normalized to have zero mean & unit variance. The normalization is done by using

$$x_i' = \frac{x_i - \mu}{\sigma_x}$$

where μ -mean and σ_x -standard deviation of that feature x_i has been used

4.3 Classifier

The MRI slices were classified using KULeuven's MATLAB/C LS-SVMlab toolbox for LS-SVM classification with both linear and RBF kernels. Various parameters for the classifiers are selected through pilot runs.

For the LS-SVM classifier, the value of gamma is chosen as 10 and value of sig2 is chosen as 0.2.

4.4 Performance measure

All classification result could have an error rate and on occasion will either fail to identify an abnormality, or identify an abnormality which is not present. It is common to describe this error rate by the terms true and false positive and true and false negative as follows:

True Positive (TP): the classification result is positive in the presence of the clinical abnormality.

True Negative (TN): the classification result is negative in the absence of the clinical abnormality.

False Positive (FP): the classification result is positive in the absence of the clinical abnormality.

False Negative (FN): the classification result is negative in the presence of the clinical abnormality.

Table 1 is the contingency table which defines various terms used to describe the clinical efficiency of a classification based on the terms above and

Sensitivity = TP/ (TP+FN) *100%

Specificity = TN/ (TN+FP) *100%

Accuracy

= (TP+TN)/ (TP+TN+FP+FN)*100 %

are used to measure the performance of the classifiers.

Table 1. Contingency table

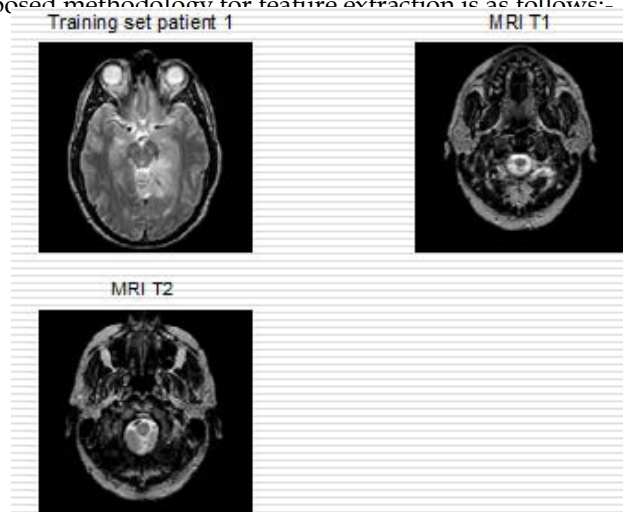| Actual Group	Predicted Group	
	Normal	Abnormal
Normal	TN	FP
Abnormal	FN	TP

5   RESULT AND OUTPUT

The results of segmentation by using K-Means clustering algorithm on medical image (Brain MRI image) is as follows:-
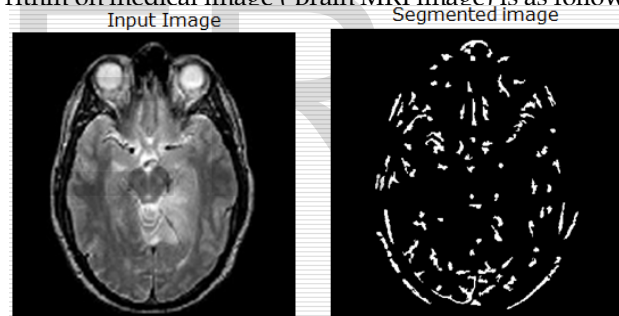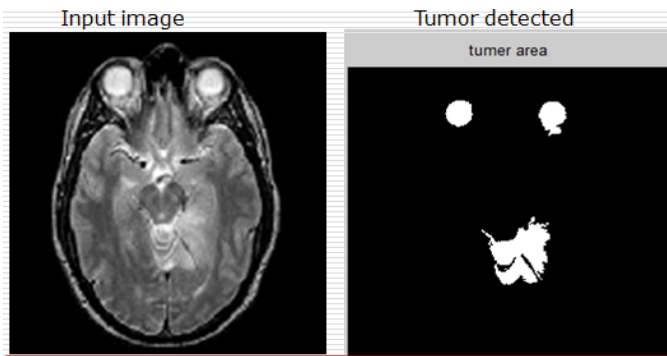

The results of feature extraction after implementation of proposed methodology for feature extraction is as follows:-

The results of Medical Image (Brain MRI image) classification

using LS-SVM as per the proposed methodology is as follows.

Input image   Tumor detected

Classifier Type	TP	TN	FP	FN	SENSITIVITY %	SPECIFICITY %	ACCURACY %
LS-SVM WITH RBF KERNEL	13	7	1	0	100	87.5	95.23

CLASSIFICATION RESULTS OF LS-SVM CLASSIFIER

6   CONCLUSION AND FUTURE SCOPE

The computer based technique for automatic classification of MRI slices as normal or abnormal with various MR image features using LS-SVM classifier is proposed.

The performance of the classifier in terms of statistical measures such as sensitivity, specificity and classification accuracy are proposed for the analysis.

This proposed methodology suggests that LS-SVM is a promising technique for image classification in a medical imaging application.

This proposed methodology provides the indication that the LS-SVM approach definitely yields the better performance.

This methodology can be used in computer aided intelligent health care systems. This automated analysis system could be further used for classification of images with different pathological condition, types and disease status.

REFERENCES

[1]   [1] H.Selvaraj, S.Thamarai Selvi, D.Selvathi, L.Gewali , "Brain MRI slices classification using least squares support vector machine" , International Journal of Intelligent Computing in Medical Science and Image Processing, IC-MED, Vol. 1, No. 1, Issue 1, Page 21 of 33,2007

[2]   [2] Prof. dr. ir. J. Vandewalle , Prof. dr. ir. S. Van Huffel , "Least squares support vector machines classification applied to brain tumour recognition using magnetic resonance spectroscopy ", D/2003/7515/71,ISBN 90-5682-460-0

[3]   [3] Milan Sonka, Vaclav Hlavac, Roger Boyle, " Image processing ,analysis , and machine vision", Pages 399 – 461,ISBN : 13:978-0-495-24428-7.

[4]   [4] B. Scholkopf, S. Kah-Kay, C. J. Burges, F.Girosi, P. Niyogi, T. Poggio, and V.Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," IEEE trans Signal Processing, vol. 45, pp. 2758-2765,1997.

[5]   [5] Jan Luts, Fabian Ojeda, Raf Van de Plas, Bart De Moor, Sabine Van Huffel, Johan A.K. Suykens," A tutorial on support vector machine-based methods for classification problems in chemometrics", Department of Electrical Engineering (ESAT), Research Division SCD, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Analytica Chimica Acta 665 (2010) 129–145.

[6]   [6] Tony Van Gestel,Johan A.K.Suykens ,Bart Baesens ,Stijn Viaene ,Jan Vanthienen ,Guido Dedene , Bart De Moor,Joos Vandewalle," Benchmarking Least Squares Support Vector Machine Classifiers", Kluwer Academic Publishers. Manufactured in The Netherlands, Machine Learning, 54, 5–32, 2004.

[7]   [7] Arjan Gijsberts , Giorgio Metta , L´eon Rothkrantz ," Evolutionary Optimization of Least-Squares Support Vector Machines ",Italian Institute of Technology, Via Morego, 30 – Genoa 16163, Italy

[8]   [8] Kristiaan Pelckmans, Johan A.K.Suykens,T. Van Gestel, J. De Brabanter, L. Lukas, B. Hamers, B. De Moor and J.Vandewalle," LS-SVMlab: a MATLAB/C toolbox for Least Squares Support Vector Machines", ESAT-SCD-SISTA K.U. Leuven Kasteelpark Arenberg 10 B-3001 Leuven-Heverlee, Belgium.

[9]   L. Lukas, A. Devos, J.A.K. Suykens, L. Vanhamme, S. Van Huffel, A.R.Tate, C. Maj´os, C. Ar´us, "The use of LS-SVM in the classification of brain tumors based on Magnetic Resonance Spectroscopic signals", in Proc. Of the European Symposium on Artificial Neural Networks (ESANN'2000), Bruges, Belgium, 2000, pp. 131-136