OPTIMIZATION OF FEATURES AND CLASSIFICATION OF BRAIN MRI IMAGES USING MAJORITY VOTING

A Dissertation submitted in partial fulfillment of the requirement for the

Award of degree of

MASTER OF TECHNOLOGY

IN

SIGNAL PROCESSING AND DIGITAL DESIGN

Submitted By

PURTI CHOUDHARY

(2K13/SPD/15)

Under the esteemed guidance of

MR. M.S CHOUDHARY

Associate Professor

DEPARTMENT OF ELECTRONICS &COMMUNICATION

DELHI TECHNOLOGICAL UNIVERSITY

DELHI-110042,

Session 2013-2015

CERTIFICATE

This is to certify that the Thesis **"Optimization of features and Classification of Brain MR Images using Majority Voting**" submitted by Purti Choudhary, Roll No. 2K13/SPD/15 student of Master of Technology (M. Tech.) in Signal Processing & Digital Design from Department of Electronics & Communication Engineering, Delhi Technological University, Delhi is a bonafide record of the candidate's own work carried out by him/her under my guidance.

Signature

Mr. M.S Choudhary

Associate Professor
Department of Electronics & Communication
Delhi Technological University

ACKNOWLEDGMENT

I feel privileged in extending my earnest obligation, deep sense of gratitude, appreciation and honor to Mr. P. R. Chadha (HOD), Department of Electronics & Communication Engineering, Delhi Technological University, Delhi, whose benevolent guidance, apt suggestions, unstained help and constructive criticism have inspired me in successful completion of making of this dissertation report.

Sincere gratitude is extended to Mr. M. S. Choudhary, Department of Electronics & Communication Engineering, Delhi Technological University, Delhi, for their valuable suggestions.

I express my appreciation and thanks to all the faculty members and staff of the Department of Electronics & Communication Engineering, Delhi Technological University, Delhi for free exchange of ideas and discussions which proved helpful.

I wish to acknowledge the affection and moral support of my family and friends for being so understanding and helpful during this period.

Finally, I am thankful and grateful to God the Almighty for ushering his blessings on me.

Name: Purti Choudhary

Roll No: (2K13/SPD/15)

ABSTRACT

Magnetic Resonance Imaging is a very popular method for abnormalities detection in the human body. Computer vision systems make this process automatic and more accurate detection so that diagnosis can be done properly. In this thesis a majority voting based system is designed based on the different intensities of texture features in different part of images. Every image is divided into four parts and all the classifiers are applied on every part of the image. The best classifier is found for every part based on the performance parameters and then majority voting concept is applied to get final class results.

One another method that is used in this thesis is feature selection which is used to found the best feature subset among the search space. Total ten texture based features are extracted from the images and represented by string of 0s and 1s. Those which are included are represented by 1s and those which are not included by 0s. The same process is repeated but this time with a combination of feature selection and majority voting.

A method for the classification of MS lesion brain MRI images is proposed. The dataset consists of 35 MRI images is taken from which 20 images are used for training (8 normal images and 12 abnormal images) and rest 15 images are used for testing of the classifier (6 normal and 9 abnormal images). Ten texture based features energy, entropy, variance, correlation, inertia, cluster shade, cluster prominence, IDM function, angular second moment, are extracted from the images, on which feature selection is applied to obtain the best feature subset which will give highest fitness function value. Total five classifiers SVM, LDA, KNN, Decision tree and Naïve Bayesian classifier are used whose performance parameters are compared to get the best classifier for a set of images.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

Now-a-days, Computer technology is becoming popular in a wide range of medical fields such as heart diseases, abnormalities in human body, cancer research and brain diseases. MRI (Magnetic Resonance Imaging) is basically a commonly used medical imaging technique. MRI is operated using radio waves and magnetic field to give high-quality two or three dimensional images of body as a result. It is a non-invasive and pain-free technique for detecting the brain tumors without any human involvement. With the help of technique, detailed information is obtained about normal and abnormal tissues. It is commonly used for medical diagnosis in hospitals.

1.1 COMPUTER VISION SYSTEMS

Computer vision system is basically an application dependent system. They deal with only specific type of detection algorithm or measurement. However there are some basic functions which can be found in many computer vision system. The steps are mentioned below:

i. Image Acquisition
ii. Pre-processing
iii. Feature Extraction
iv. Detection

1.1.1 IMAGE ACQUISITION

Image acquisition is the first stage of any computer vision system. An image is generated by using scanning the outputs of sensors. The resulting image obtained can be represented in two dimensional or three dimensional space which depends on the type of sensors used for image acquisition. The pixel values at various coordinates (gray image or color image) is in accordance with the light intensity at various parts scanned. The intensity at various coordinates in the image can be found with the help of histogram of the image. Various methods such are magnetic resonance imaging, CT scan, X-ray are used by which signals are obtained that are represented in the image form.

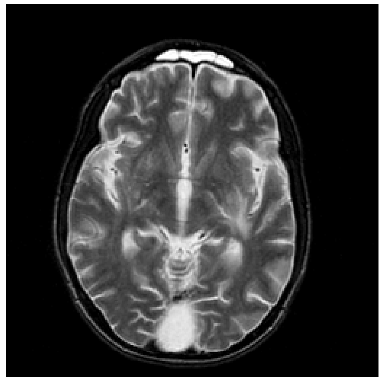

Figure 1.1 Sample of Brain MRI Image *[1]*

1.1.2 PRE-PROCESSING

It is also known as image restoration. When image acquisition is done using sensors, noises are introduced in the image and it may be possible that intensity level at some points may be changes, distorted [2]. Generally the data should be pre-processed before the application of vision system to image data for the purpose of classification or detection, so as to ensure that it satisfies certain assumptions required by the method. Some of the examples of pre-processing methodologies are:

i. To ensure that the image coordinate system is right, re-sampling is performed [3].
ii. Various find of filters are applied to eliminate noise [4] which is introduced due to sensors.
iii. Contrast is enhanced to correct the intensity levels at various points.
iv. Techniques like gamma correction are applied.

1.1.3 FEATURE EXTRACTION

Features are extracted from images at various levels complexity. These features are the basis of information required for classification of two or more types of classes of images. There are various type of feature which can be extracted depending on the type of images and application for which it is to be used. Some examples of features are:

1. Lines, ridges, and edges.
2. Energy, entropy, contrast.

Features are basically texture, shape or intensity based and they are used as information for the purpose of training and testing of classifier.

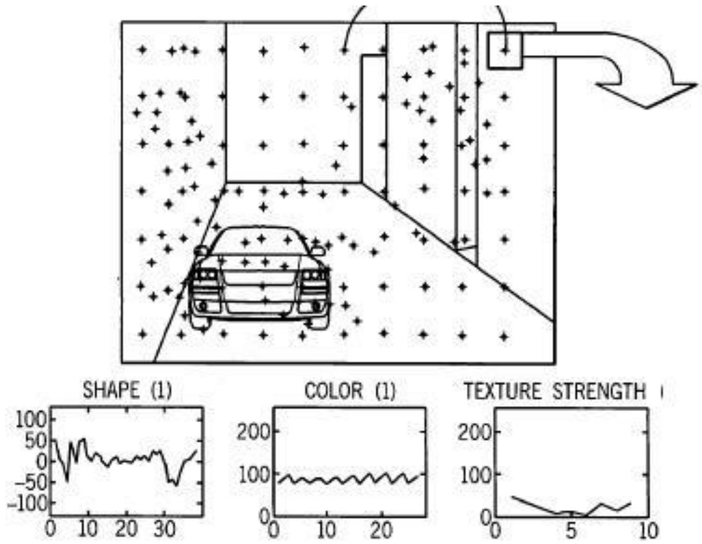

Figure1.1.2 Extraction of features from an image

1.1.4 DETECTION

Then these features are used for the purpose of detection / classification. Classification methods are classified in two types: supervised and unsupervised methods. There are two types of data set one is training data set whose target values are already known and the other one is testing data set whose target values are to be detected. Firstly the classifier are trained using the features of training data set and then this trained classifier is applied on the testing data set for the classification purpose.

1.2 BIOMEDICAL IMAGING

MRI is relatively a new technology. It was invented by Paul C. Lauterbur in September 1971. The first MR image was issued in 1973. Previously it was also known by the name nuclear magnetic resonance imaging (NMRI), magnetic resonance tomography (MRT). The first human MRI image was taken in 1875. Earlier the use of MRI was only limited to neuro-axis, but gradually it became popular and now it is used for every body part. More studies in this field revealed how MRI can be coupled with other methods like EEG [5], ECG [6], to get more depth into diagnostics.

1.2.1 MAGNETIC RESONANCE IMAGING

Magnetic Resonance Imaging (MRI) or Nuclear Magnetic Resonance Imaging (NMRI) is basically an imaging technique which is used by radiologist to get information about physiological and anatomical functioning of the body. MRI makes use of a strong magnetic field, radio waves and a computer to get detailed information about the internal structure of the body. The image produced by MRI provides much more contrast difference between various soft tissues as compared to Computed Tomography (CT). It is generally preferred as compared to CT.

CT makes used of ionizing radiations whereas MRI used strong magnetic field to align the hydrogen atoms of the water in the body in same direction. This magnetization is symmetrically changed by the use of radio frequency (RF) fields. It leads to a rotating magnetic field produced by the hydrogen atoms of water in the body, which can be recognized by the scanner.

1.2.2 WORKING OF MRI

Hydrogen nucleus is used for the imaging purpose as it is found in abundance in water and fat in human body. Hydrogen atom can be considers as like our planet earth, which spin around its axis and also possesses north-south pole. In general, hydrogen atom acts as a bar magnet. In normal conditions the bar magnet like hydrogen atom spins around its axis in random directions in the human body and resulting into net zero magnetic field. As soon as the body is placed in a strong magnetic field, all the hydrogen like bar magnets line up in a same direction along the axis of scanner. There are different strengths in which MRI scanner are available, ranging between 0.5-1.5 tesla. [7]

One other electromagnetic field causes the protons to absorb energy. As soon as the field is turned off, energy is released by the protons which is detected by the scanner. These signals helps in the formulation of MR images. To improve the detection of the emitted signals, receiver coils are applied around the human body. The intensity of the received signal is plotted in grayscale tone and thus MR images are obtained.

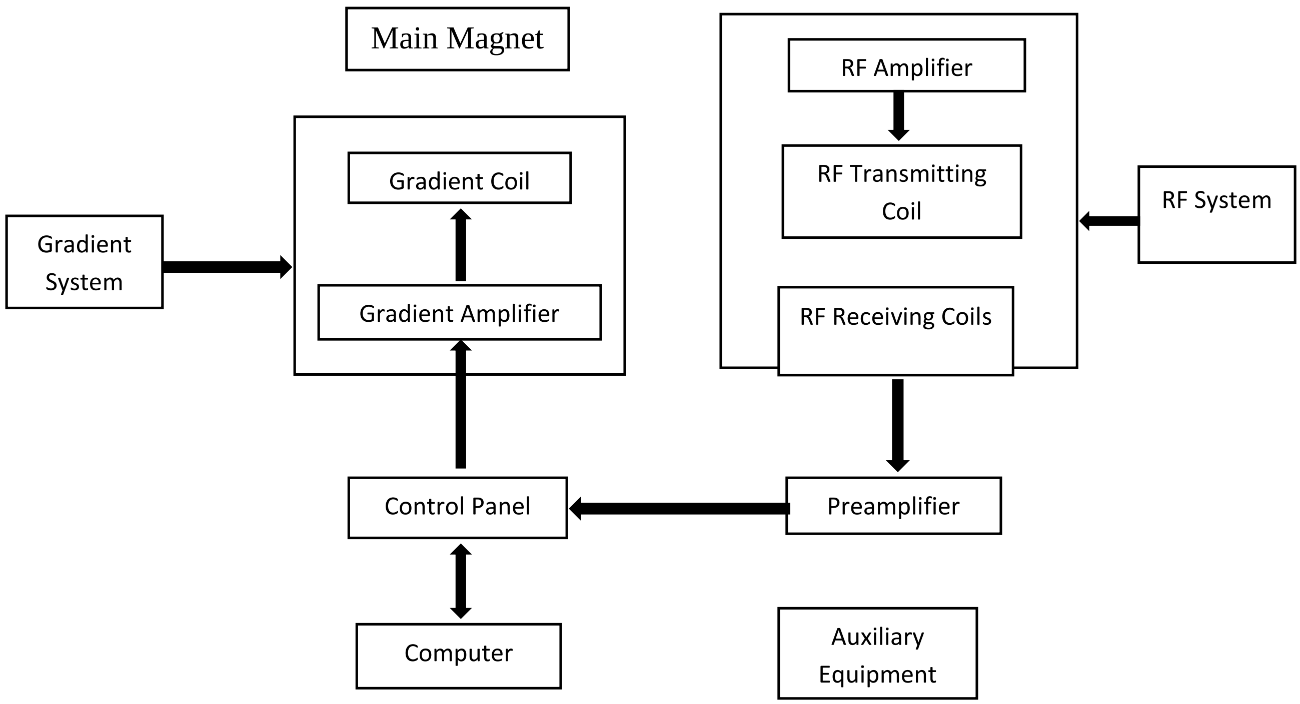

Figure1.3 Structure of MRI Imaging system

Different type of diseases in the human body, such as, various abnormalities, tumors etc. can be detected by inspecting these MR images obtained. As it takes different span of time for different soft tissues in the human body to come to an equilibrium. This thing can be used for changing the contrast of different soft tissues in the MR image by changing the parameters in the scanner in accordance with the difference in the span of time for coming to an equilibrium. To enhance the appearance of tumor, blood vessels or any inflammation, contrast agents are injected into directly into the blood vessels. In case of MR images of joint, contrasting agents are directly injected into the joints. The contrasting agents are based on chelated of gadolinium. These contrasting agents are found to be safe than that used in X-ray radiology or CT, that is ionized contrasting agents.

MR images can be used for every part of the body. It is used especially for checking the neurological conditions, detecting tumors, recognize various abnormalities, disorders in joints and muscles. MRI is a more safe procedure as compared to CT, as it does not used ionized radiations.

Figure 1-1.4 MRI Scanner *[8]*

Patients which are having cardiac pacemakers, or any kind of metallic implants are not allowed to undergo MRI. It is unsafe for them as the strong magnetic field and radio frequency will affect the metals. MR images provides very fine details in the form of features which provides a strong base for the treatment of abnormalities in the human body. However, there are some parameters which put restrictions on computer based MR image observation, that are, surroundings in which body part is scanned, differences in the MRI scanner, noise and some other interference sources, patient's postures and conditions while taking MRI.

1.2.3 IMAGING

To produce an image a number of methods have been discovered for combining the excitation results obtained from the MRI scanner. The image can be reconstructed either in two dimensional or three dimensional space, same as in Computed Tomography. The image can also be constructed point-by- point or line-by-line. Although all of these methods are basically application specific, so now a days most of the MR Images today are produced either by the Two-Dimensional Fourier

Transform (2DFT) analysis, or by the Three-Dimensional Fourier Transform (3DFT) analysis. 2DFT is also known as spin-warp.

1.3 RELATED WORK

Magdi el al [9] introduced a model for the brain tumor detection in MRI images. This method includes a three steps framework that is preprocessing of images, extraction of relevant information from the images and image classification. The purpose of the preprocessing step was to increase the contrast and reduce the noise level in the image and filtering techniques are applied for this process. Texture features are extracted from the MRI images and then Principal Component Analysis (PCA) is applied for dimensionality reduction and improve the complexity of the system. Classification of MRI images is done using Back Propagation Neural Network (BPNN) classifier. Rajesh patil el al [10] discovered a method with the help of MATLAB software to extract and detect tumor from a brain MRI image. There is intensity differences in tumor and its background in brain MRI image. Preprocessing, segmentation and then morphological operations are applied on images for the detection of tumor.

Mehdi Jafri and Raja Shafagi [11] proposed a hybrid approach based on genetic algorithm and support vector machines (SVM) for the detection of tumor in brain MRI images. Using different techniques like FFT, GLCM and DCT features are extracted from the images and then feature selection is done using genetic algorithm. Finally the images are classified as normal and abnormal by applying selected features to the SVM classifier. Nandagopal and Rajamony [12] used a combination of wavelet statistical features (WST) and wavelet co-occurrence texture features (WCT) obtained from a two level discrete wavelet transform applied on the images to classify them as normal or abnormal. Jayachandran and Dhansekharan [13] proposed a hybrid model using statistical and SVM classifier for detection of brain tumor in the brain MRI images. The whole method is divided in four steps: noise reduction, feature extraction, feature selection and classification.

Mustara and Suchalatha [14] in his paper proposed a method for brain cancer detection and classification. Image preprocessing techniques like histogram equalization, morphological operations, image segmentation are used. Feature extraction is done using gray level co-occurrence matrix (GLCM). Ann classifier is first trained and then test for checking the performance. Rathi and Palani [15] in this paper introduced a different method of feature extraction and feature selection. In this method all the three, intensity based, shape based and texture based features are calculated and used for making a distinction between gray matter, white matter and classifying normal and abnormal images.

1.4 ORGANIZATION OF THE THESIS

The first chapter includes about the computer vision system, various steps involved in classifying any test image, basic about MRI images and the working of MRI scanner. It also includes the literature review. The second chapter is about the proposed methodology of this thesis. It is explain in detailed with the help of a flow chart. The third chapter introduces about GLCM (gray level co-occurrence matrix) and how textures features are obtained from this GLCM matrix. Various texture features used in the procedure of this thesis are also explained. The forth chapter is about the theory of fitness function. The fifth chapter includes classification method of images and the details about all the five classifiers. The sixth chapter discusses the results of the proposed methodology and the seventh chapter finally concludes the thesis.

CHAPTER 2 PROPOSED WORK

A method for the classification of MS lesion brain MRI images is being proposed. A dataset of 35 brain MR Images is considered. Out of these these 35 images, 20 images are used for the training purpose of classifiers (8 normal, 12 abnormal) and rest 15 images are used for the testing (6 normal, 9 abnormal) of the trained classifier so as to know the performance parameters that are Sensitivity, Specificity, Accuracy. This whole thesis is basically a two-step framework. In the first step the whole procedure is done without feature selection and in the second step the same procedure is repeated but this time with feature selection application. The whole procedure is explained in detail below:

2.1 IMPLEMENTATION WITHOUT FEATURE SELECTION

Firstly the dataset is taken and feature extraction is done. Here texture features are extracted from images which discriminates one class of images (normal images) from the another class (abnormal images). Total ten texture based features are extracted from the images that are mentioned below:

i. Energy
ii. Entropy
iii. Correlation
iv. Variance
v. Inertia
vi. Angular Second Moment (ASM)
vii. Inverse Difference Moment (IDM)
viii. Cluster Shade
ix. Cluster Prominence
x. Contrast

These ten features are extracted from all the training and testing images dataset. Thus we get two types of feature matrix, one is training feature matrix which contains the features of training dataset and other is testing feature matrix that contains features of testing dataset. Firstly the training feature matrix is used for the training purpose of the supervised classifiers used in this process.

a) Support Vector Machine (SVM) classifier
b) Linear Discriminant Analysis (LDA) classifier
c) Decision Tree classifier
d) K-Nearest Neighbor (KNN) classifier
e) Naïve Bayesian classifier

Now the trained classifier's performance is tested based on the testing feature set matrix. These trained classifiers classifies the testing images as normal or abnormal based on working criterion of each classifier. The performance is based measured based on three parameters:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

As different features can be more pronounced in different part of images. So it may be possible that some classifier gives better result for some part and some other classifier for some different part. This concept is being used in the thesis. The whole image is divided into four equal parts. Each part is being treated as a separate image and the same process of training and testing is applied.

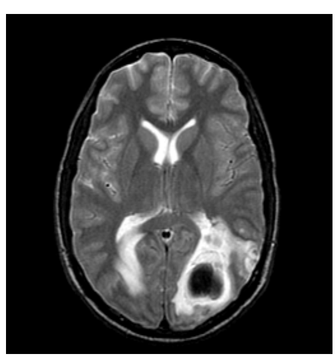

Figure 2.1 Brain MRI Image

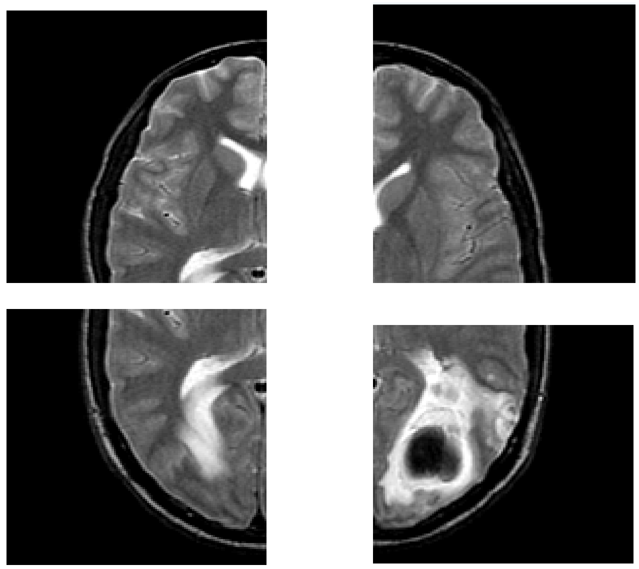

Figure 2.2 Four parts of Brain MRI Image

All the 5 classifiers are applied to the first part of the image. Training and testing of the classifiers is done and the performance parameters are computed. The classifier which is best for the first part is found by another parameter that is *average*, whichever classifier is having the highest *average*, its class results are taken into consideration for the further processing. If any two classifiers are having the same average values than comparison is done on the basis of the *average* of those classifiers when applied on the whole image. Whichever is having the higher *average* value is considered to be finally best classifier for the part 1 of the image.

$$Average = \frac{(Accuracy + Sensitivity + Specificity)}{3}$$

Similar procedure is repeated for rest of the parts of the image that is part 2, 3 and 4. Best classifiers are also found for these parts by following the same procedure and their corresponding class results are stored. Now for finding class result for the image the concept of majority voting is used. In majority voting, the final result is the result of majority. For example, if three classifiers are giving the results that the corresponding three parts of the image belongs to class 1 and the remaining part to class 2. Then according to majority, the image is assigned class 1. In this way from the individual class results of the four different parts of the image, final class for the image is found. After getting the final class results, again performance parameters are calculated. It is shown that performance after majority voting concept is better than that obtained by applying the classifier on the whole image.

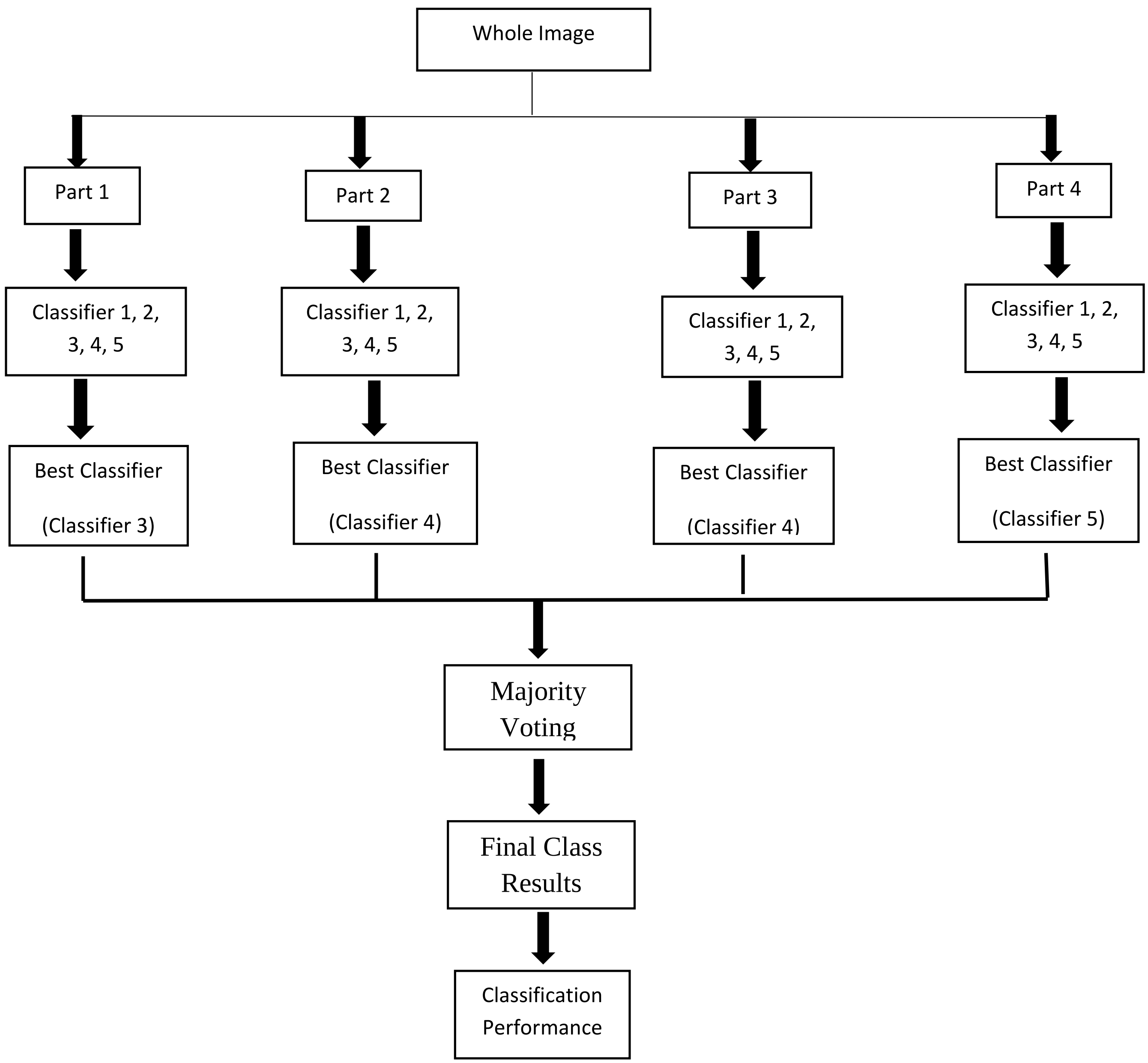

Figure 2.3 Flowchart of project methodology

2.2 IMPLEMENTATION WITHOUT FEATURE SELECTION

This is the second step of the proposed framework. As we all know that as the dimensionality of the data keeps on increasing, the complexity of the classification problem increases. So here with the help of feature selection method the dimensionality of the features is reduced keeping the performance of the classifiers in mind.

The feature set is represented by a binary string of 0s and 1s. The features which are included are represented by 1s and those which are not included are represented by 0s. A fitness function is used for this feature selection process and output obtained by the fitness function is given as a feedback to the system to know whether the stopping criterion is met or we have to keep on generating new generations more. In this way a best feature set which is of reduced dimensions is found for every classifier and corresponding performance parameters are also calculated. From the results we can see that the performance of the classifiers is being increased by using this concept of feature selection.

Now the next stage is the same as that done before. Every image is divided into four different equal parts. On every all the five classifiers are applied but this time it is done with application of feature selection procedure. Best feature subset is calculated for every part of the image and based on the value of *average* parameter, best classifier is defined for every part and their corresponding class results are stored. Based on the majority voting concept, again the final classes of the images are found and accordingly the final performance parameters are calculated.

CHAPTER 3 FEATURE EXTRACTION

Features are basically the most important characteristics of the object of interest. If features are selected then they help in providing relevant information regarding the classification problem. The object and the images are analyzed by the feature extraction methodologies and they the most prominent features are extracted that represents different classes and images. Features are given as the input to the classifiers which assigns them to the class they belong to. The purpose of feature extraction is to reduce the database and take only relevant information into consideration which can differentiate among one class to another class.

The feature space of the proposed method in this thesis is texture based features. Texture features can differentiate between diseased and normal brain MR images. The texture features are calculated using gray level co-occurrence matrices (GLCM) matrices.

3.1 GRAY LEVEL CO-OCCURRENCE MATRIX

Texture features are basically the statistical distribution of the intensities at specified points with respect to each other in the image. The statistics is classified as first order, seconder order and higher, according to pixel values at various points in the image. This approach is used in a number of applications like, eg. Image classification, terrain classification, sandstone reservoir classification, object detection, segmentation of images, [16] [17] [18] [19] [20]GLCM concept is proposed by Haralick in 1970s. It is basically a tabulation of how often different pixel brightness value combinations (gray levels) occur in an image. The elements of matrix that is $P(i, j/\Delta x, \Delta y)$ is basically the frequency with which two neighboring pixels, one at intensity level *i*, and other at intensity level *j*, are separated by a distance of $(\Delta x, \Delta y)$. Similarly the second order statistics of GLCM matrix can be represented by $P(i, j/d, \theta)$, where the pixels values *i* and *j* are separated by a distance *d* at a particular angle θ.

Let us consider an image having a M x N neighborhood and G gray levels ranging between 0 and G-1. Let *f(m,n)* be the intensity at pixel m and line n of the neighborhood.

$$P(i, j/\Delta x, \Delta y) = WQ(i, j/\Delta x, \Delta y)$$

Where

$$W = \frac{1}{(M - \Delta x)(N - \Delta y)}$$

$$Q(i, j/\Delta x, \Delta y) = \sum_{n=1}^{N-\Delta y} \sum_{m=1}^{M-\Delta x} A$$

$$A = 1 \; if \; f(m, n) = i \; and \; f(m + \Delta x, n + \Delta y) = j$$

$$A = 0 \; elsewhere$$

A 5x5 image with 4 gray levels and its gray level co-occurrence matrix $P(i, j/\Delta x, \Delta y)$ is shown below:

IMAGE

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

Figure 3.1 Gray level image having 4 intensity levels

P(i,j;1,0)

	j=0	1	2	3
i= 0	1/20	2/20	1/20	0
1	0	1/20	3/20	0
2	0	0	3/20	5/20
3	0	0	2/20	2/20

Figure 3.2 GLCM matrix of the image above

3.2 TEXTURE FEATURES FROM GLCM

A large number of texture features can be calculated from GLCM matrix (Haralick el al. 1973 [21], Conners el al. 1984 [22]). The following notations are used:

G is the total number of gray levels used.

μ is the mean value of P.

$\mu_x, \mu_y, \sigma_x, \sigma_y$ are the mean and standard deviations of $P_x \; and \; P_y$.

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

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

$$\mu_x = \sum_{i=0}^{G-1} i \sum_{j=0}^{G-1} P(i,j) = \sum_{i=0}^{G-1} i P_x(i)$$

$$\mu_y = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} jP(i,j) = \sum_{j=0}^{G-1} jP_y(j)$$

$$\sigma_x^2 = \sum_{i=0}^{G-1} (i-\mu_x)^2 \sum_{j=0}^{G-1} P(i,j) = \sum_{i=0}^{G-1} (P_x(i)-\mu_x(i))^2$$

$$\sigma_y^2 = \sum_{j=0}^{G-1} (j-\mu_y)^2 \sum_{i=0}^{G-1} P(i,j) = \sum_{j=0}^{G-1} (P_y(j)-\mu_y(j))^2$$

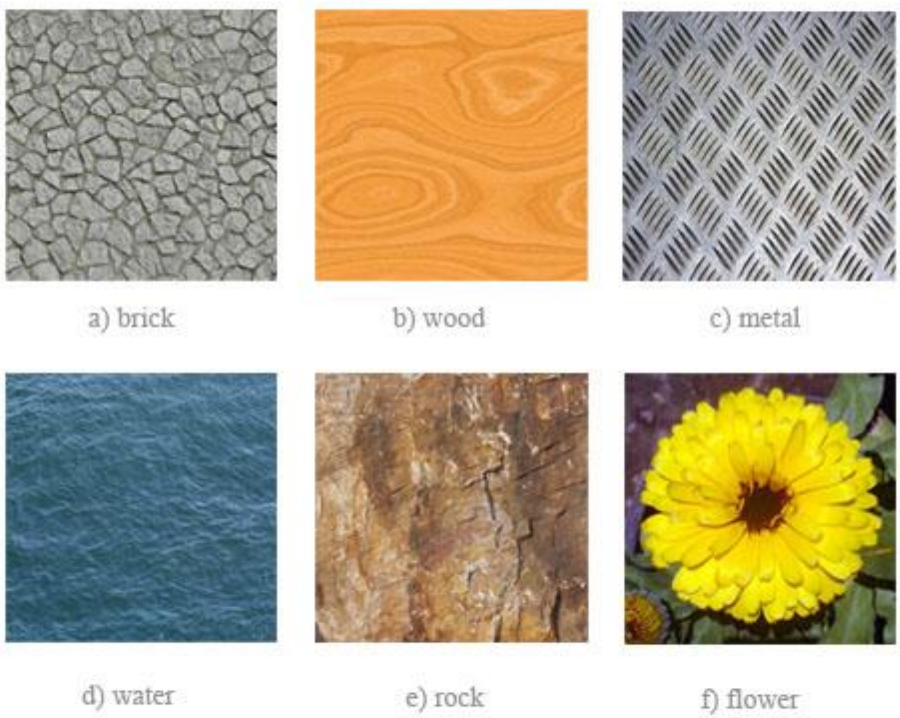

Figure 3.3 Different textures of images

Texture can be categorized as: visual and touch texture. Touch texture relates to the feel when we touch any surface and it range from smoothest to roughest. While visual texture is related to visual perception, when we see any object and the changes occur due to color, brightness, intensity and orientation of an image. Gray level co-occurrence matrix (GLCM) is proved as a potential method of extracting texture features from the images. According to the GLCM matric, fourteen texture features are introduced by Haralick, which define the texture characteristics of the images. In this thesis work ten of these texture features are being used and are explained below [23] [24]:

3.2.1 ENTROPY

Entropy is used to characterize the texture of the image. It is a measure of randomness in the image. Images which are having inhomogeneous scenes are having low value of first order entropy and those having homogeneous scenes are having high value of first order entropy. It is calculated using the formula:

$$ENTROPY = -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P(i,j) * \log(P(i,j))$$

Direct matlab function is also available for calculating entropy, that is, “entropy(I)”. The entropy of image can be calculated from the histogram of the image. It histogram of the image calculates various gray level probabilities and used in the prediction of entropy. Entropy is highest when all the probabilities in P(i,j) are equal and smaller when they are unequal.

3.2.2 CORRELATION

Correlation is parameter of measuring linear dependency of neighboring pixels statistically at particular locations with respect to each other.

$$CORRELATION = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{\{i \times j\} \times P(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}$$

The value of the correlation coefficient lies between -1 and 1. If its value is 1, then it indicates that the variables are positively closely related and if -1, it indicates negatively closely related. But if the value is zero it shows a very weak relationship between the variables.

3.2.3 VARIANCE

Variance is the measure of the dispersion of values around the mean of a reference point and all the neighboring pixels. It gives higher weights to the points which differ largely from the mean value. It is calculated using the formula:

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

3.2.4 CLUSTER SHADE

Cluster shade is basically a measure of the skewness of the gray level co-occurrence (GLCM) matrix. It is based upon the concept of symmetry and uniformity. If an image is symmetric then the value of cluster shade will be low and if an image is symmetric then the value of cluster shade will be high. It is calculated using the formula given below:

$$CLUSTER\ SHADE = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i + j - \mu_x - \mu_y)^3 \times P(i,j)$$

3.2.5 CLUSTER PROMINENCE

Similar to cluster shade, cluster prominence is also a parameter for the measure of asymmetry. When image is less symmetric, the value of cluster prominence is very high. When the value of cluster prominence is very high, there exists a peak around the mean value in the GLCM matrix. It is calculated using the formula given below:

$$CLUSTER\ PROMINENCE = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i + j - \mu_x - \mu_y)^4 \times P(i,j)$$

3.2.6 ANGULAR SECOND MOMENT

Angular second moment (ASM) is a measure of homogeneity in the image. It tells about the uniformity in the image. Its value is large when the pixels in the image are very similar. It is calculated using the formula:

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

3.2.7 INERTIA

Inertia is the statistical measure of the rotation inertia of the rigid bodies in the image. It represents how much effort is required to rotate an image, if each pixel intensity is associated with a weighted mass. The higher is the pixel value at a point, the more is the weighted mass value and the harder it is to rotate that point. The colored images are first converted at gray scale for the calculation of the statistical measure. Inertia is calculated using the formula below:

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

3.2.8 ENERGY

Energy is basically the measure of extent to which pixel pairs are repeating. By this the uniformity of image is being measured. The value of energy will be large when there will be too much similarity among the pixels. Energy is calculated using the formula given below:

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

3.2.9 INVERSE DIFFERENCE MOMENT

Inverse difference moment (IDM) is also known as homogeneity and it measures the local homogeneity of the image. The closeness of the GLCM matrix elements to its diagonal elements is measured by the IDM function. By the range of values of IDM function we can know about the texture of an image. IDM function is calculated using the function:

$$IDM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1+(i-j)^2} P(i,j)$$

3.2.10 CONTRAST

In terms of real world, contrast is the difference is color and brightness among two objects. In this context contrast is basically the difference of the intensity values among various point in the image.

$$CONTRAST = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i,j) \right\}, \qquad |i-j| = n$$

CHAPTER 4 FEATURE SELECTION

In machine learning traditional algorithms can handle the data with a large number of instances but as soon as the dimensionality of the data keeps on increasing the complexity of the algorithms increases [25]. The solution to this problem is to find which dimensionality is important and which are not. Evolutionary algorithms are being used for this process of dimensionality reduction [26].

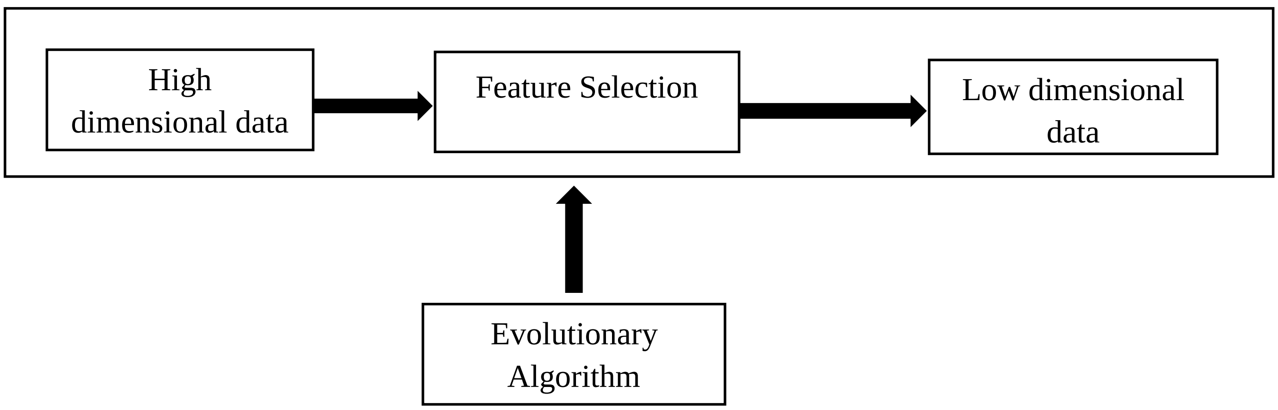

Figure 4.1 General approach of feature selection

In the field of statistics and machine learning, feature selection is also known as candidate selection or search space subset selection. There are commonly three reasons because of which feature selection is used:

i. To simplify the classifications problems so that their interpretation becomes easier.
ii. So as to limit the training time of classifiers.
iii. Over fitting is prohibited to some extent by this method.

4.1 FITNESS FUNCTION

Fitness function is basically an objective function which is used to define how close the designed solutions are to the desired results. The fitness function is useful in selecting the best feature subset among the whole search space of features. We keep on iterating the loop until the maximum number of generations have occurred or the stopping criterion is satisfied and this will give us the best feature set [27].

The choice of fitness function is basically application dependent. It depends on what quantity we want to maximize or minimize. In this thesis, fitness function which is being used is given by the equation shown below:

$$finess\ function = (Accuracy + Sensitivity + Specificity) + (0.05 \times number)$$

The first three parameters, that is, accuracy, sensitivity and specificity are basically the performance parameters [28] which are being used in every classification problem to know the performance of the classification.

$$Accuracy(\%) = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity(\%) = \frac{TP}{TP + FN}$$

$$Specificity(\%) = \frac{TN}{TN + FP}$$

Where

TP: True positive

TN: True negative

FP: False positive

FN: False negative

4.2 PARAMETERS USED

Accuracy is defined as the total number of correct results, from both the categories, among the total number of test samples classified.

Sensitivity is defined as the proportion of positive test samples that are correctly classified. It is also known as true positive rate or recall.

Specificity is defined as the proportion of negative test samples that are correctly classified. It is also called as true negative rate.

The features which are extracted are represented in the form of binary strings of zeros and ones. To find a feature subset, in binary string the 0s represents that the feature is excluded and 1s represent that the feature is included. So here in the fitness function formula the quantity ***number*** represents the total number of 0s in the binary string [29]. We have to look for the maximum value of this fitness function for different feature sets and the formula used will help in this task.

The sum of the performance parameters accuracy, sensitivity and specificity increases as the classification performance increases with the search space of features and will tend to move the value of fitness function to a more positive side. Same thing is done by the parameter *number*, as the number of zeros increases in the binary string that means less number of features are considered, fitness function value increases. So the aim of this fitness function is to get high classification performance with less number of features.

CHAPTER 5 CLASSIFICATION

Classification is the method of identifying the group to which a new observation belongs given a set of groups or classes. This process is based on training data set which contains observation whose class label is known. In terms of machine learning, classification belongs to supervised learning methods [30] [31]. Any algorithm that do classification is a classifier. It refers to the mathematical function, where input data is mapped into a category. Linear classifier include perceptron, logistic regression naive baye's classifier and many more. Classification algorithms are widely used in speech recognition, handwriting recognition, pattern recognition etc.

Classification divides the feature space into several classes depending upon the features extracted from the images.

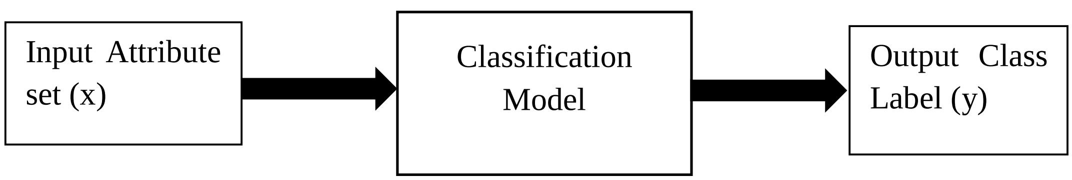

Figure 5.1 Classification as a task of mapping an input attribute set to its class label

Classification methods are basically of two types: Supervised and Unsupervised classification.

i. **Supervised Method:** In case of supervised learning we already have a set of training sample and their target values and inferred to that a function is created which helps in finding target values for the test samples.
ii. **Unsupervised Method:** In case of unsupervised learning no aid is provided by the user and the target values are determined by the analysis of the images. Techniques are used to analyze the images and the pixels which are closely related are assigned to the same class.

5.1 SVM CLASSIFIER

SVM has become popular for its wide use in the fields of pattern recognition and regression in previous years [32]. It belongs to supervised learning methods which run by analyzing data and recognizing patterns. Given a set of training examples with known class labels, SVM develop a model which assigns new examples a class label based on patterns that are recognized. It is a non-probabilistic linear classifier. It not only performs linear classification but also non-linear one. SVM constructs one or a set of hyper-planes in high dimensional space. Basically, a good classification is done by the plane that covers maximum margin as it reduces the generalization error.

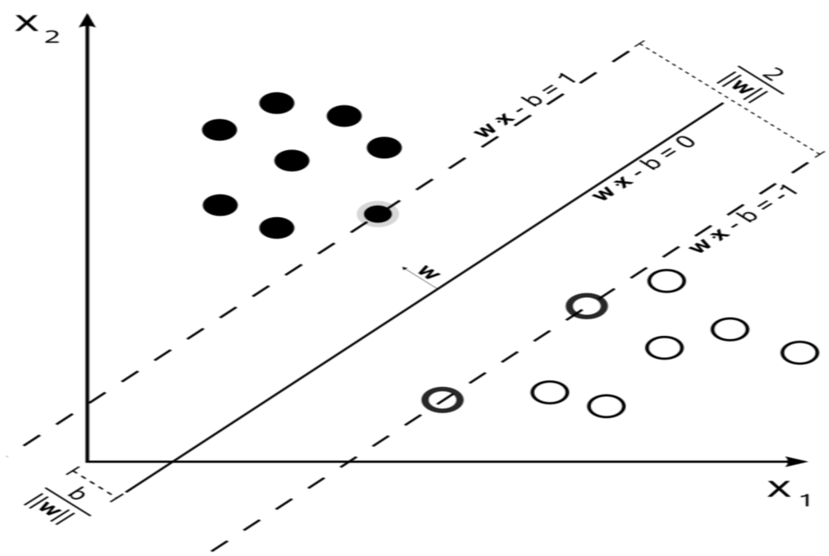

Figure 5.2 Maximal margin hyper-plane and margins for a SVM

A set of n points comprising training data d where y is either 1 or -1 which indicates the class to point x belongs. Any hyper-plane can be represented as the set of points x that satisfy equation:

$$w.x - b = 0$$

It denotes the dot product and w is the normal vector to hyper-plane. Two hyper-planes that separate data are selected if the training data is linearly separable. Then the distance between them called margin is maximized. Hyper-planes are represented as:

$$w.x - b = 1$$

and

$$w.x - b = -1$$

It can be rewritten as:

$$y_i(w.x_i - b) \geq 1, \quad \text{for all } 1 \leq i \leq n$$

The maximum margin classifier, which is the standard SVM, is found by maximizing the margin and is equivalent to the following problem:

Minimize $\quad min_{(w,b)} \quad \frac{1}{2}\|w\|^2$

subject to $y_i(w.x_i - b) \geq 1 \quad$ for all $1 \leq i \leq n$

5.2 LDA CLASSIFIER

Linear discriminant analysis (LDA) is a generalized form of Fisher's linear discriminant. It is used to discriminate or characterize two or more different classes by finding a linear combination of features [33]. It is useful in various fields like pattern recognition, statistics and machine learning. It can be used as either of the two: as a linear classifier or as a dimensionality reduction method. This method guarantees maximal separability by maximizing the ratio of between-class variance to within-class variance in a dataset. This method is also used for the classification problems in the field of speech recognition. The major difference between PCA and LDA is that PCA focuses more on feature classification whereas LDA focuses more on class classification. In PCA, when transformation is done to some other space, the data changes its shape and location, whereas in LDA the shape and location of the data does not change and it tries to give better class seperability by creating a decision region between the two classes.

5.2.1 TRANSFORMATION APPROACHES

There are basically two data transformation approaches in LDA which are explained below:

i. Class-dependent transformation: In this approach the ratio of between class variance to within class variance is maximized. The main objective for this maximization is to get better class seperability.

ii. Class-independent transformation: In this approach the ratio of overall class variance to within class variance is maximized. In this kind of approach each class is considered to be a separate class against all other classes.

5.2.2 MATHEMATICAL OPERATIONS

All the mathematical operations applied in case of LDA are explained here and for this explanation a two-class classification problem is taken into consideration.

i. We have a data set and test vectors for which classification is to be performed. The data set is represented by a set of features shown in a matrix form below:

$$set\ 1 = \begin{bmatrix} a_{11} & a_{12} \\ \vdots & \vdots \\ a_{m1} & a_{m2} \end{bmatrix} \qquad set\ 2 = \begin{bmatrix} b_{11} & b_{12} \\ \vdots & \vdots \\ b_{m1} & b_{m2} \end{bmatrix}$$

ii. Mean of each data set and mean of the whole data set is computed. Let us consider $\mu_1\ and\ \mu_2$ be the mean of class 1 and 2 respectively, and μ_3 be the mean of the whole data set which is obtained by:

$$\mu_3 = p_1 * \mu_1 + p_2 * \mu_2$$

where $p_1\ and\ p_2$ are the apriori probabilities.

iii. In LDA, the between-class variance and the within-class variance are calculated using the following formulae:

$$S_w = 0.5 * cov_1 + 0.5 * cov_2$$

And the covariance matrices are calculated using the formula:

$$cov_j = (x_j - \mu_j)(x_j - \mu_j)^T$$

The between-class scatter is calculated using the following equation

$$S_b = \sum_j (\mu_j - \mu_3) * (\mu_j - \mu_3)^T$$

The optimizing factors in class dependent type of transformation is calculated as:

$$criterion_j = inv(cov_j) * S_b$$

And the optimizing factor for class independent transformation is calculated as

$$criterion = inv(S_w) * S_b$$

iv. The decision region in the transformed space is represented by a solid line seperating the transformed data set.
For class dependent LDA,

$$transformed\_set\_j = transform\_j^T * set\_j$$

For class independent LDA,

$$transformed\_set = transform_{spec}{}^T * data_{set}{}^T$$

Similarily transformation of the test vectors is done and classified using the Euclidean distance from the mean as a criterion.

v. Once the data is transformed Euclidean distance is used as a criterion for the classification. If there are n classes, then n Euclidean distances will be obtained for each test point.

vi. The test point is assigned to that class for which the Euclidean distance is minimum among all the n Euclidean distances.

The type of LDA to be used depends on the application and the classification problem for which it is to be used.

5.3 DECISION TREE CLASSIFIER

Decision tree is used as a predictive model for decision tree classifier in which samples are mapped to its target values [34]. This method is used in various fields like statistics, pattern recognition, machine learning [35]. The decision tree has three types of nodes:

i. **Root node:** It is a node which has no incoming branches and zero or more outgoing branches.

ii. **Internal node:** It has one incoming branch and two or more than two outgoing branches.

iii. **Leaf or terminal nodes:** It has exactly one incoming branch and no outgoing branch.

All the non-terminal nodes, that is root node and internal nodes are basically used to separate samples which have different characteristics based on some attribute test. While the leaf or terminal nodes contain the class results or the target values.

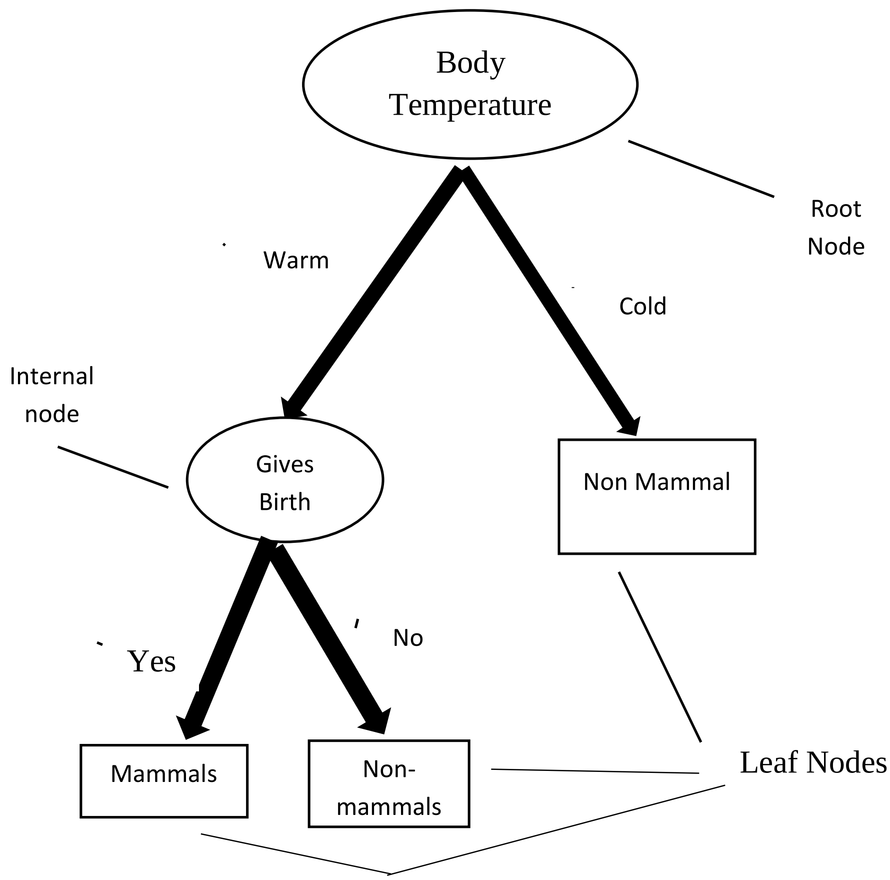

Figure 5.3 Decision Tree for mammal classification system

As it can be seen in the figure above that here body temperature is used as an attribute for the root node to separate out warm blooded vertebrates from cold blooded vertebrates. In this way the tree is proceeds. A leaf node is created at the right side as non-mammal as we know that all cold-blooded vertebrates are non-mammal. Now further, for warm-blooded vertebrates, Gives birth, is used as another attribute for distinction and finally we get the leaf nodes as the target values.

So classifying using a decision tree is quite simple. Once a decision tree is constructed, starting from the root node test conditions are applied and moves to further branch. This will lead to either a terminal or leaf node or to another internal node where again test conditions are to be applied to move further in the decision tree till the end, where at the end we get the target values for the test samples.

The advantage of decision tree classifier is that the computational time is less whereas the disadvantage is that the accuracy is wholly dependent upon the decision tree constructions, the testing attributes selection and feature selected. When two only two outcomes are provided by a decision tree at each stage, then such type of decision tree is called Binary decision tree classifier (BDT).

5.4 KNN CLASSIFIER

KNN classifier is used for the purpose of classification and regression. It is basically a non-parametric method. KNN classification is considered to be the simplest of all the classification problems. As all the approximations are done only locally, that's why this type of learning is also called instance-based learning or lazy learning. Each characteristic in the training set is considered to be a different dimension in the space. The value of any training sample for that characteristic is considered to be shown in the form of a coordinate in that dimension space. The similarity between two samples is determined by the distance between the coordinates of those samples. Smaller is the distance more is the similarity.

As a new test sample comes for the classification, it is also plotted in the same coordinate system accordingly. For choosing the class to which it belongs, its k nearest training samples are consider. K nearest training samples are identified by the distance measure. The class to which the majority of training samples belong among the k nearest samples, that class is assigned to the testing sample. This is the reason why this learning method is called k nearest neighborhood method.

As we can see the figure below there are two types of training samples, one is blue rectangles belonging to class 1 and another is red triangles belonging to class 2. Now the green circle is the test sample whose class we have to identify. If we consider k=3, then the three nearest neighbor to the test sample will be the three samples under the bold circle. As we can see majority belongs to the class 2, so class 2 is assigned to the test sample.

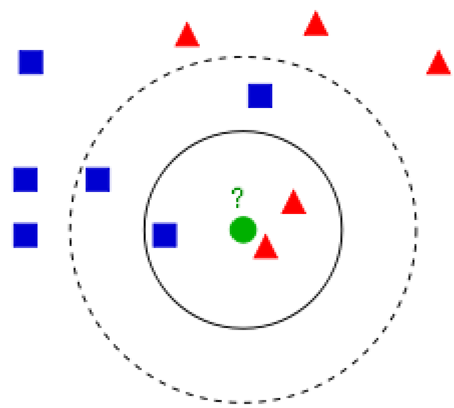

Figure 5.4 Example of KNN classification

The algorithm can be summarized in following steps:

i. A new test sample is introduced, along with specifying the user defined value k.
ii. K samples are chosen from the training samples that are closest to the test sample.
iii. Majority class is found out among the k nearest neighborhoods.
iv. That class is assigned to the test sample.

5.5 NAÏVE BAYES CLASSIFIER

Naïve bayes classifier is a very simple classifier based on baye's theorem and it assumes very strong independence among the features. Initially in 1960s it was introduced under a different name and was a popular method for text classification regarding problems such as, text categorization, personal email sorting, email spam detection etc. Although the functionality of the naïve bayes classifier is quite simple, but it can be applied to very complex real time problems with high classification accuracy.

It gives superior results in terms of CPU utilization time and memory requirement. Naïve Bayes classifier is based on conditional probabilistic model. If some test samples are given for classification, let us assume they are represented by $x = \{x_1, x_2, \dots., x_n\}$ which represents the n features. Then the condition probability for the k classes outcome will be represented by $p(C_k|x_1, \dots, x_n)$. Using the Baye's theorem, the conditional probability can be represented by:

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)}$$

In generalized form, this conditional probability can be written as below using baye's theorem:

$$posterior = \frac{prior \times likelihood}{evidence}$$

The conditional probability $p(C_k|x_1, \dots, x_n)$ using the chain rule can be rewritten as:

$$p(C_k|x_1, \dots, x_n) = p(C_k)p(x_1, \dots, x_n|C_k$$

$$= p(C_k)p(x_1|C_k)p(x_2, \dots, x_n|C_k, x_1)$$

$$= p(C_k)p(x_1|C_k)p(x_2|C_k, x_1)p(x_3, \dots, x_n|C_k, x_1, x_2)$$

$$= p(C_k)p(x_1|C_k)p(x_2|C_k, x_1). \; . \; . \; .p(x_n|C_k, x_1, x_2, x_3, \dots, x_{n-1})$$

Now comes the Naïve's conditional independence assumption: according to it all the feature are independent of each other. That is F_i is conditionally independent of F_j given that $i \neq j$. This means that:

$$p(x_i|C_k, x_j) = p(x_i|C_k)$$

$$p(x_i|C_k, x_j, x_k) = p(x_i|C_k)$$

$$p(x_i|C_k, x_j, x_k, x_l) = p(x_i|C_k)$$

And keeps going on for all $i \neq j, k, l$. Thus the joint model can be represented as:

$$p(C_k|x_1, \dots, x_n) \propto p(C_k, x_1, \dots, x_n)$$

$$\propto p(C_k)p(x_1|C_k)p(x_2|C_k)p(x_3|C_k) \ldots\ldots$$

$$\propto p(C_k)\prod_{i=1}^{n} p(x_i|C_k)$$

Thus the conditional probability of the class variable C is given by:

$$p(C_k|x_1, \dots, x_n) = \frac{1}{Z} p(C_k)\prod_{i=1}^{n} p(x_i|C_k)$$

CHAPTER 6 RESULTS

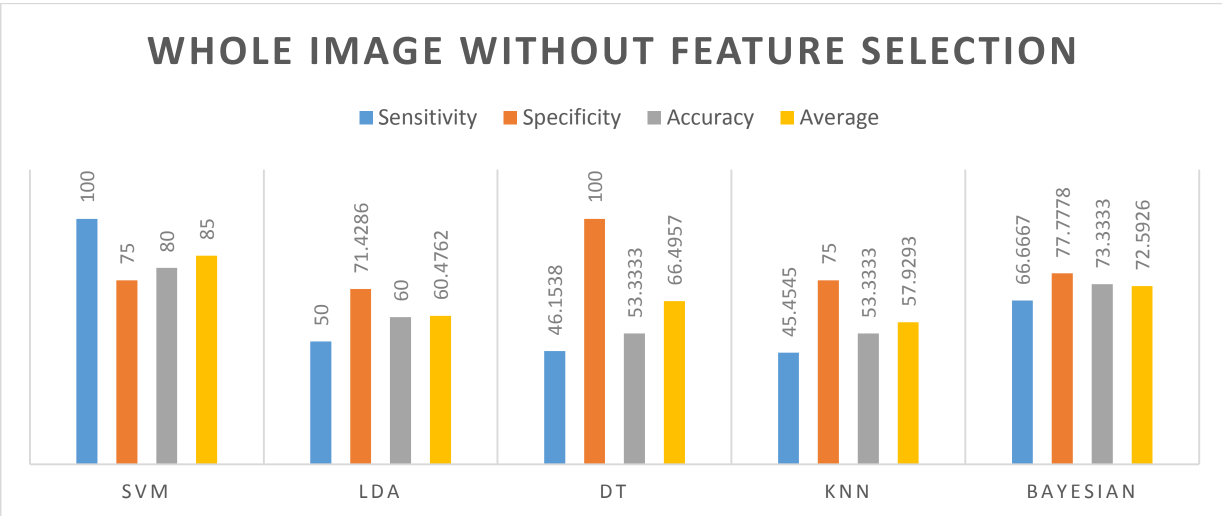

Figure 6.1 Performance parameters of whole image

The figure 6.1 shown above represents the percentage values of performance parameters when classification is performed on whole image without any feature selection. The best classifier among all the classifier is SVM classifier with Accuracy 80%, Sensitivity 100%, Specificity 75% and average value 85%.

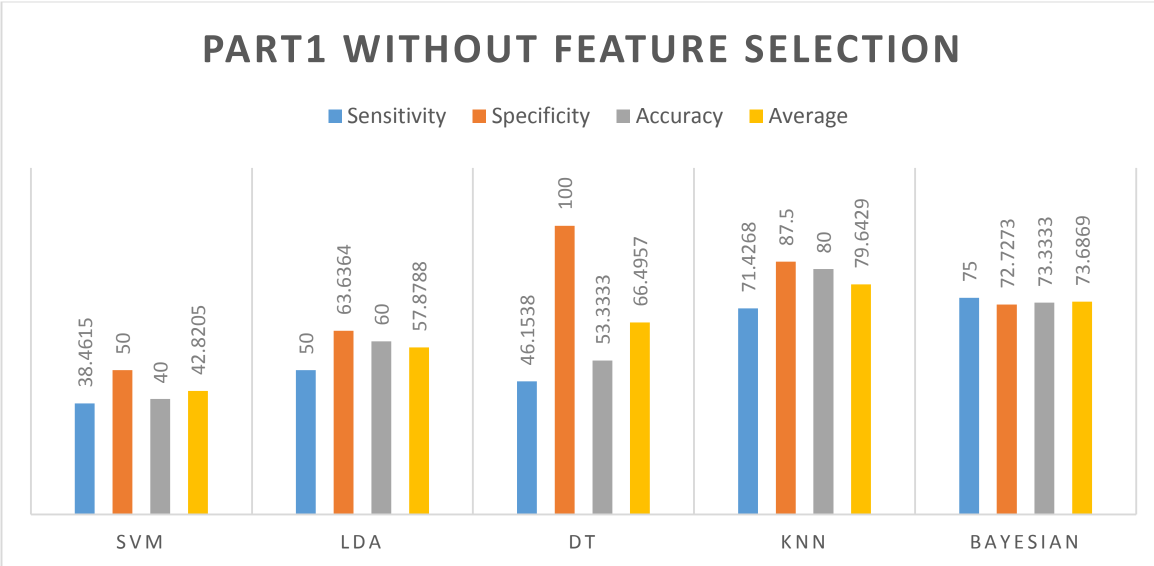

Figure 6.2 Performance parameters of first part of image

The figure 6.2 shown above represents the performance parameters of different classifiers when applied on the first part of the brain MRI images among the four equal parts in which each image is divided without feature selection. The best classifier among all of these classifiers is KKN classifier with the percentage performance parameter values Sensitivity 71.4268%, Specificity 87.5%, Accuracy 85% and average value 79.6429%.

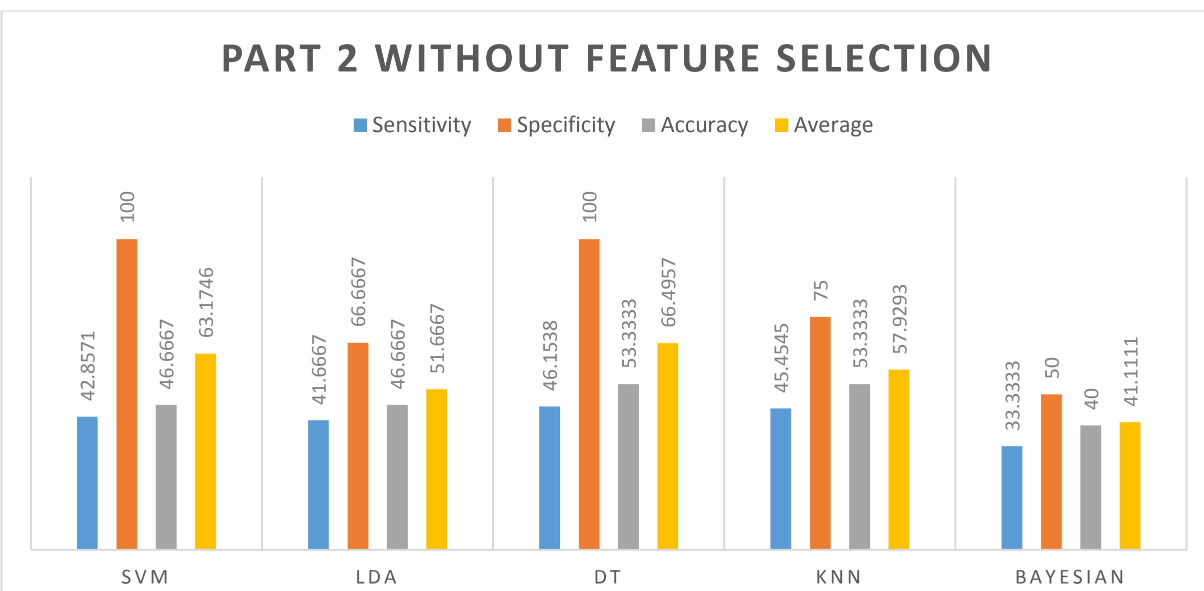

Figure 6.3 Performance parameters of second part of image

The figure 6.3 shown above represents the performance parameters of different classifiers on second part of the image among all the four parts of the brain MRI images without feature selection. The best classifier among all the five classifiers is Decision Tree classifier with percentage performance parameter values Sensitivity 46.1538%, Specificity 100%, Accuracy 53.3333% and average value 66.4957%.

The figure 6.4 shown below represents the performance parameters of different classifiers on third part of the image among all the four parts of the brain MRI images without feature selection. Here there are two classifiers LDA and Bayesian classifier having the highest average value. To get best out of we go one step and find the best average value when these classifiers are applied on whole image that is 60.4762 and 72.5926 respectively. Hence we get the best classifier among all the five classifiers is Bayesian classifier with percentage performance parameter values Sensitivity 85.7143%, Specificity 100%, Accuracy 93.3333% and average value 93.0159%.

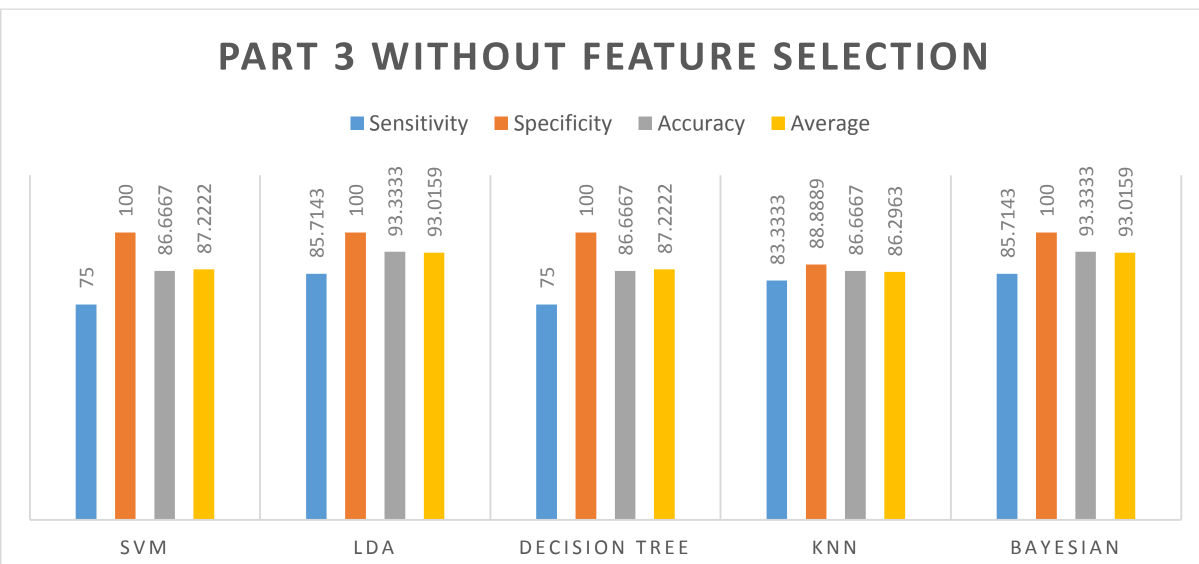

Figure 6.4 Performance parameter of third part of image

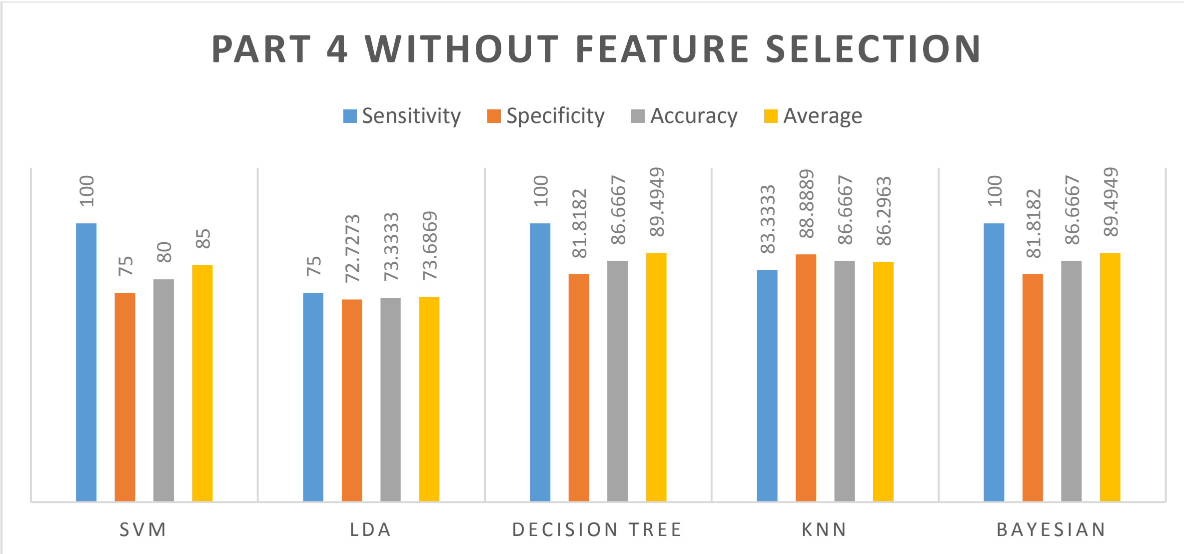

Figure 6.5 Performance parameter of forth part of image

The figure 6.5 shown above represents the performance parameters of different classifiers on third part of the image among all the four parts of the brain MRI images without feature selection. Here there are two classifiers Decision Tree and Bayesian classifier having the highest average value. To get best out of we go one step and find the best average value when these classifiers are applied on whole image that is 66.4957 and 72.5926 respectively. Hence we get the best classifier among all the five classifiers is Bayesian classifier with percentage performance parameter values Sensitivity 100%, Specificity 81.8182%, Accuracy 86.6667% and average value 89.4949%.

Combining the class results of all the classifiers and applying majority voting we get the performance parameters as Sensitivity 85.7143%, Specificity100%, Accuracy 93.3333% and Average value 93.0159%.

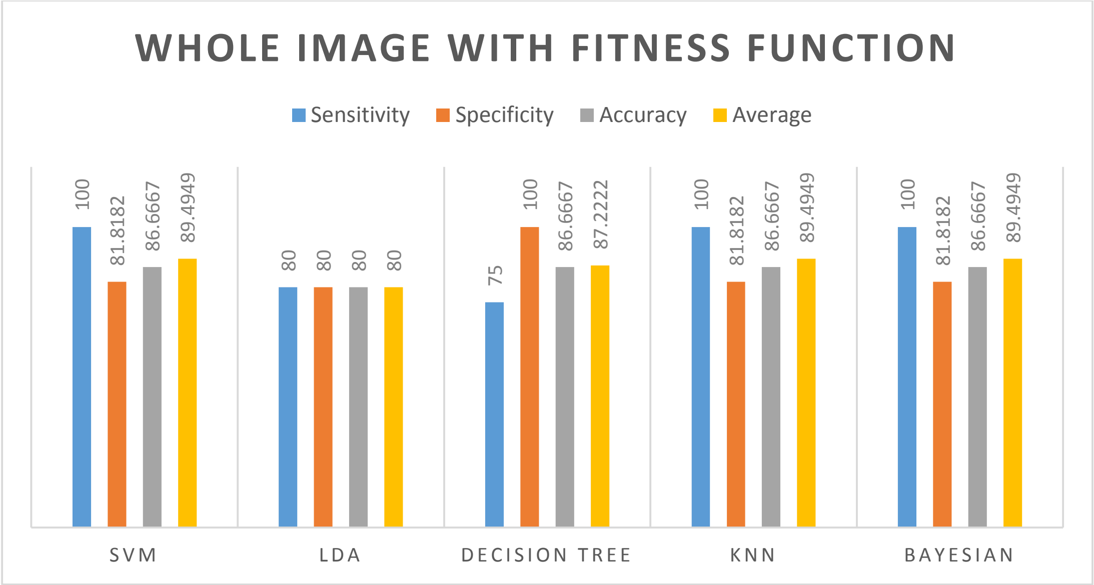

Figure 6.6 Performance parameter of whole image with feature selection

The figure 6.6 shown above represents the performance parameters of whole image on applying feature selection. Here three classifiers that is SVM, KNN and Bayesian classifier are having the same average value. The best classifier among these is decided based on length of the feature set as shown in the table below. SVM classifier is having the least number of features in the subset and is assigned the best classifier with Sensitivity 100%, Specificity 81.8182%, Accuracy 86.6667% and Average 89.4949%

CLASSIFIER	FITNESS VALUE	FEATURE SUBSET
SVM CLASSIFIER	2.8848	10
LDA CLASSIFIER	2.75	2,5,7
DECISION TREE	3.0167	1,4
KNN CLASSIFIER	3.0848	2,8
BAYESIAN	3.0848	1,3

Table 6.1 Fitness value and Feature subset for all classifiers on whole image

The table 6.1 shown above represents the value of fitness function when applied on whole image taking different classifiers into consideration and the optimized feature subset selected.

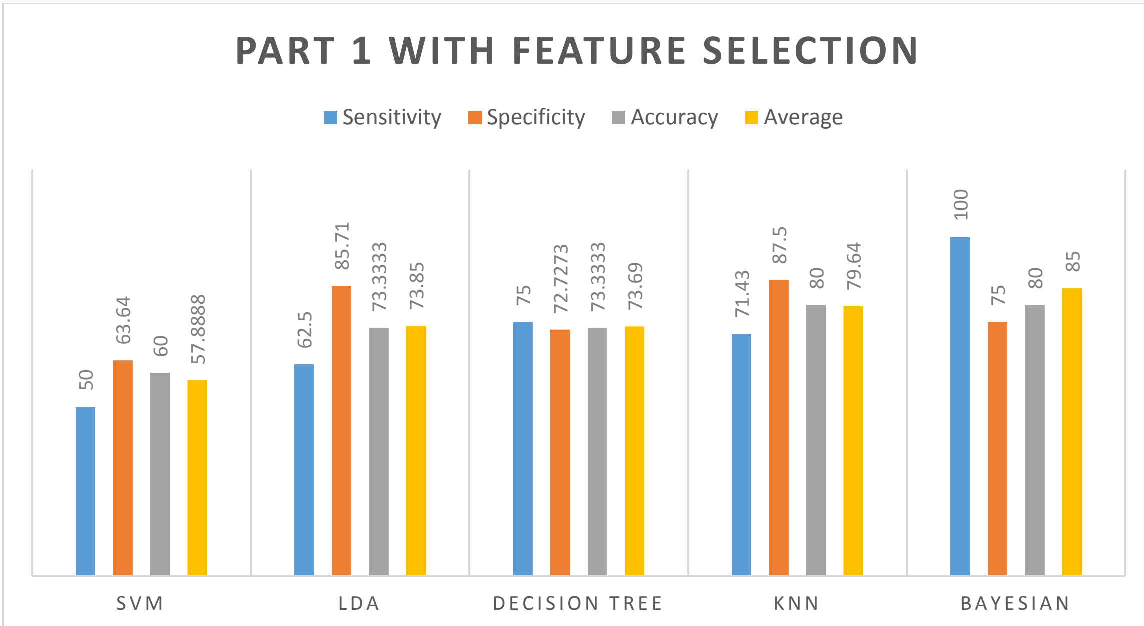

Figure 6.7 Performance parameter of first part of image with feature selection

The figure 6.7 shown above represents the performance parameters of first part of image on applying feature selection. Among all the classifiers Bayesian classifier is having the highest average value and hence considered to be the best classifier with Sensitivity 100%, Specificity 75%, Accuracy 80% and Average 85%.

CLASSIFIER	FITNESS VALUE	FEATURE SUBSET
SVM	2.1364	6,9
LDA	2.6155	2,7
Decision Tree	2.6606	6
KNN	2.7393	4,6,8
Bayesian	2.95	1,9

Table 6.2 Fitness value and Feature subset of all classifiers on first part of image

The table 6.2 shown above represents the value of fitness function when applied on first part of the image among the four parts taking different classifiers into consideration and the optimized feature subset is selected.

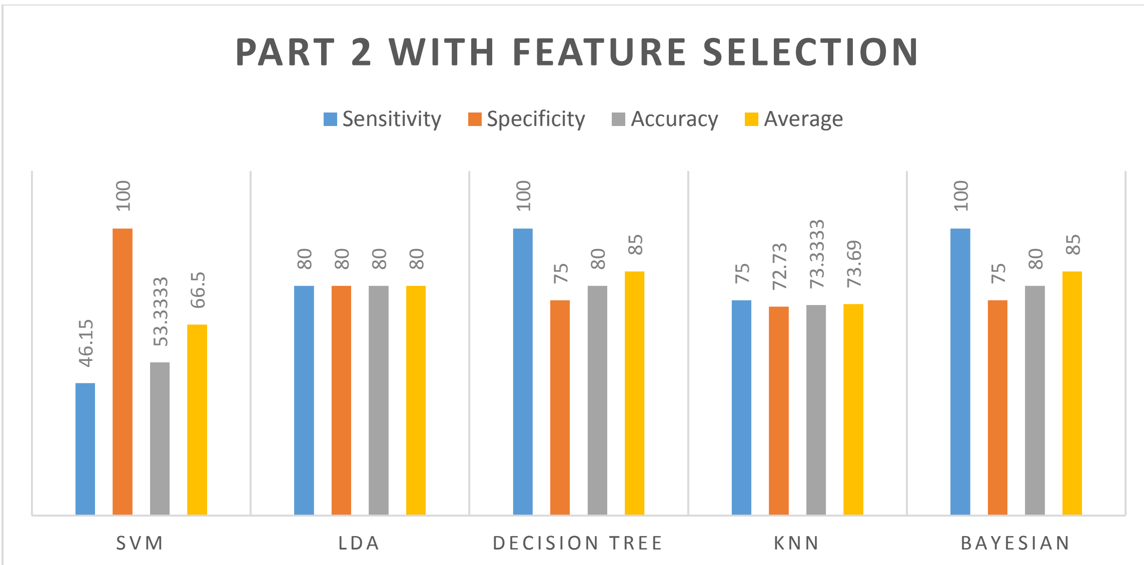

Figure 6.8 Performance parameters of second part of image with feature selection

The figure 6.8 shown above represents the performance parameters of second part of image on applying feature selection. Among all the classifiers Bayesian and Decision tree classifier are having the highest average value and to find best out of both we go one step back by comparing performance parameters of these classifiers on whole image. Finally Bayesian classifier is found to be the best classifier with Sensitivity 100%, Specificity 75%, Accuracy 80% and Average 85%.

CLASSIFIER	FITNESS VALUE	FEATURE SUBSET
SVM	2.3949	4,8
LDA	2.8	5,9
Decision Tree	3	9

KNN	2.6106	5,9
Bayesian	2.9	1,3,5

Table 6.3 Fitness value and Feature subset of all classifiers on second part of image

The table 6.3 shown above represents the value of fitness function when applied on second part of the image among the four parts taking different classifiers into consideration and the optimized feature subset is selected.

The figure 6.9 shown below represents the performance parameters of second part of image among all the parts on applying feature selection. In this case all the classifiers have the same average value. To find best among all average value at previous step for whole image is compared. It is found that SVM classifier is the best as it used only one feature as its feature subset. The SVM is the best classifier with Sensitivity 85.71%, Specificity 100%, Accuracy 93.3333% and Average 93.012%.

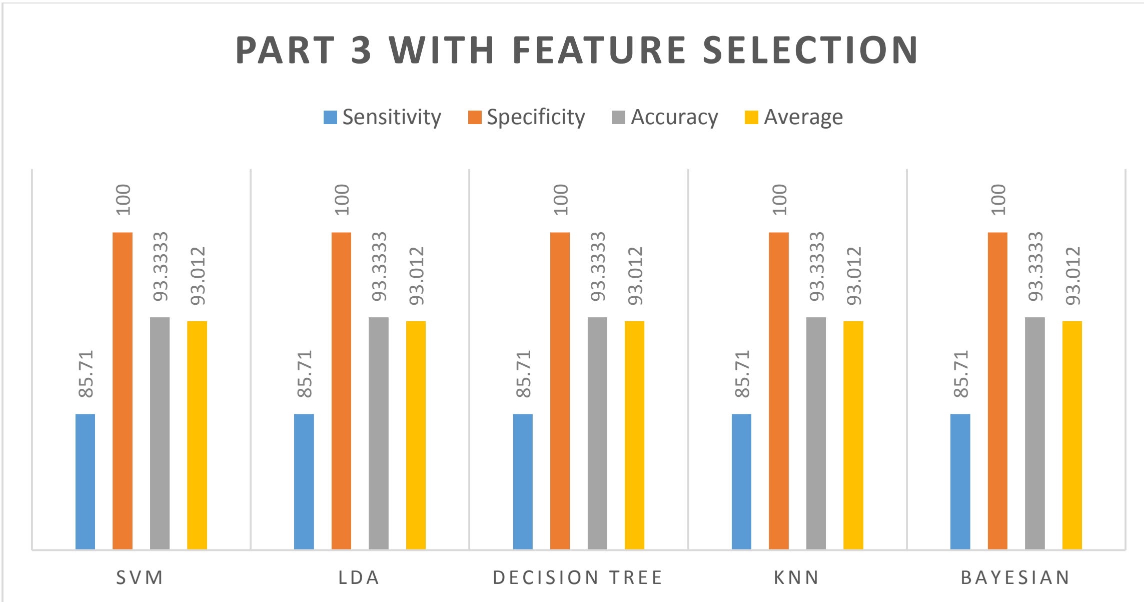

Figure 6.9 Performance parameter of third part of image with feature selection

CLASSIFIER	FITNESS VALUE	FEATURE SUBSET
SVM	3.2405	10
LDA	3.1905	2,6
DECISION TREE	3.2405	8
KNN	3.1905	1,8
BAYESIAN	3.2405	4

Table 6.4 Fitness value and Feature subset of all classifiers on third part of image

The table 6.4 shown above represents the value of fitness function when applied on third part of the image among the four parts taking different classifiers into consideration and the optimized feature subset is selected.

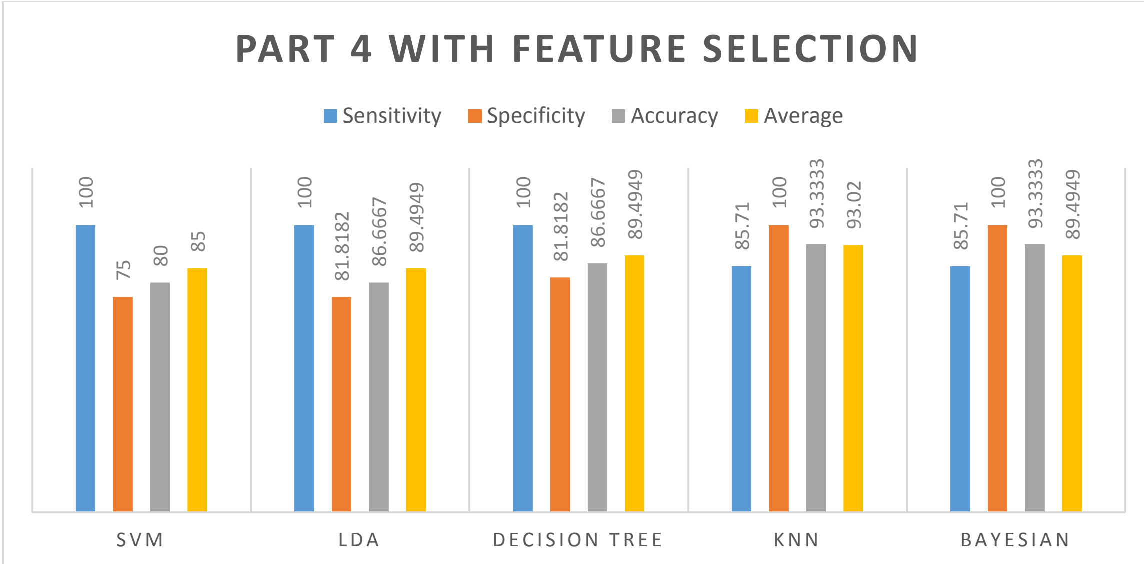

Figure 6.10 Performance parameter of forth part of image with feature selection

The figure 6.10 shown above represents the performance parameters of forth part of image among all the parts on applying feature selection. KNN classifier is having the highest average value among all the classifiers with performance parameters Sensitivity 85.71%, Specificity 100%, Accuracy 93.3333% and Average 93.012%.

CLASSIFIER	FITNESS VALUE	FEATURE SUBSET
SVM CLASSIFIER	3	10
LDA CLASSIFIER	3.0848	2,6
DECISION TREE	3.2405	6
KNN	3.1405	3,6,7
BAYESIAN	3.0848	3,8

Table 6.5Fitness value and Feature subset of all classifiers on forth part of image

The table 6.5 shown above represents the value of fitness function when applied on forth part of the image among the four parts taking different classifiers into consideration and the optimized feature subset is selected.

Results	Sensitivity (%)	Specificity (%)	Accuracy (%)	Average (%)
Without majority voting and fitness function	100	75	80	85
With majority voting	85.7143	100	93.3333	93.0159
With fitness function	100	81.8182	86.6667	89.4949
With fitness function and majority voting	100	90	93.3333	94.444

Table 6.6 Comparison between the performance parameters using different techniques

The table 6.6 shown above represents the comparison between the various performance parameters on applying different methods on the image database. When the classification is done on applying both fitness function and majority voting best results are obtained.

CHAPTER 7 CONCLUSION

Magnetic Resonance Imaging (MRI) is a commonly used method for detecting abnormalities in human body. By the MR image obtained from the MRI scanner doctors can know about the stage of the disease and can diagnose him / her properly. With the help of computer-aided systems this diagnosis can be done automatically on computers and it will reduce the work load of the doctors and will give more accurate results about the abnormalities in the brain MR Images.

In this thesis such type of a system is designed with the help of feature selection algorithm to reduce the dimensionality of the feature set and get more optimal results. As we know that different parts of the image have different part of features more pronounced in their area and in this way the performance parameters of the different classifiers also changes with this change of features. So a methodology is developed to collect the best of all and define a more strong system. To achieve this every image is divided in four parts. Individually each classifier is applied on every part and best of them is chosen for every part. Majority voting is the second step in which the majority rule is being used to get the final class results and the performance parameters accuracy, sensitivity and specificity. Better results are being obtained after the proposed methodology as compared to the traditional methodology.

Another advancement is done in the above mentioned method and that is feature selection. As we know that complexity of the system increases with the dimensionality increment and even sometimes it leads to over training of the classifier. Among all the features extracted it may be possible that only few are contributing towards the better results and rest are either overtraining the system or just overlapping the results already obtained. In such a case to obtain the best feature subset is very much important task. This task is performed with the help of feature selection method. A fitness function is used for this process of feature selection. The output that is the fitness value is given as a feedback to the classification system. The feature subset corresponding to the highest fitness value is considered for the performance parameter estimation. It can be concluded from the results that the combination of the two proposed methods, feature selection and majority voting lead to a more accurate and efficient system.

CHAPTER 8 REFERENCES

[1] "BrainWeb," [Online]. Available: http://brainweb.bic.mni.mcgill.ca/brainweb/.

[2] N. F. Ishak, M. J. Gangeh and R. Logeswaran, "Comparison of Denoising Techniques Applied on Low-field MR Brain Images," in *Fifth International Conference on Computer Graphics, Imaging and Visualization*, Korea, 2008.

[3] S. H. Joshi, A. Marquina, S. J. Osher, I. Dinov, J. D. Van Horn and A. W. Toga, "MRI Resolution ehancement using total variation regularization," *IEEE,* 2009.

[4] H. Tiwari, V. Bhateja and A. Shrivastava, "Estimation Based Non-Local Approach for Pre-Processing of MRI," in *2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2015.

[5] T. Adali, V. Calhoun and J. Liu, "A feature based approach to combine functional MRI, structural MRI and EEG brain imaging data," in *Engineering inMedicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, New York, NY, 2006.

[6] A. Gupta, A. R. Weeks and S. M. Richie, "Simulation of Elevated T-Waves of an ECG Inside a Static Magnetic Field (MRI)," *Biomedical Engineering, IEEE,* vol. 55, no. 7, pp. 1890-1896, 2008.

[7] G. Katti, S. A. Ara and A. Shireen, "Magnetic Resonance Imaging (MRI) – A Review," *INTERNATIONAL JOURNAL OF DENTAL CLINICS,* vol. 3, pp. 65-70, 2011.

[8] M. Kakapurayil, MNT, 2003. [Online]. Available: http://www.medicalnewstoday.com/articles/146309.php.

[9] B. M. A. Magdi, A.-e. Ahmad and I. Walla, "An Intelligant Model for Automatic Brain Tumor Diagnosis Based on MRI Images," *International Journal of Computer Applications,* vol. 72, pp. 21-24, 2013.

[10] R. C. Patil and D. A. S. Bhalchandra, "Brain Tumour Extraction From MRI Images using MATLAB," *International Journal of Electronics, Communication & Soft Computing Science and Engineering,* vol. 2, no. 1, 2011.

[11] A. Kharrat and K. Gasmi, "A Hybrid Approach for Automatic Classification of Brain MRI Using Genetic Algorithm and Support Vector Machine," *Journal of Sciences,* pp. 71-82, 2010.

[12] P. N. Gopal and R. Sukanesh, "Wavelet statistical feature based segmentation and classification of brain computed tomography images," *IET image process,* vol. 7, pp. 25-32, 2013.

[13] A. Jayachandran and R. Dhanasekaran, "Brain tumor detection and classification of MRI using texture features and Fuzzy SVM classifiers," *Journal of Applied Sciences, Engg. and Tech.,* vol. 6, 2013.

[14] Mustara and Suchalatha, "Brain cancer classification using GLCM based feature extraction in artificial neural network," *ELSEVIER, computerized Medical Imaging and Graphics,* vol. 29, pp. 21-34, 2005.

[15] V. P. G. P. Rathi and S. Palani, "Linear Discriminant Analysis for Brain tumor classification using feature selecion," *Int Journel of Communication and Engineering,* vol. 5, 2012.

[16] R. W. Conners and C. A. Harlow, "A theoretical comparision of texture algorithms," *IEEE Trans. on Pattern Analysis and Machine Intell.,* Vols. PAMI-2, pp. 204-222, 1980.

[17] M. Iizulca, "Quantitative evaluation of similar images with quasi-gray levels," *Computer Vision, Graphics, and Image Processing,* vol. 38, pp. 342-360, 1987.

[18] L. H. Siew, R. H. Hodgson and E. J. Wood, "Texture Measures for Carpet Wear Asessment," *EEE Trans. on Pattern Analysis and Machine Intell.,* Vols. PAMI-10, pp. 92-105, 1988.

[19] J. S. Weszka, C. R. Dyer and A. Rosenfeld, "A comparative Study of Texture Measures for Terrain Classification," *EEE Trans. on Systems, Man and Cybernetics,* Vols. SMC-6, pp. 269-285, 1976.

[20] D. C. He, L. Wang and J. Juibert, "Texture Feature Extraction," *Pattern Recognition Letters,* vol. 6, pp. 269-273, 1987.

[21] R. W. Conners, M. M. Trivedi and C. A. Harlow, "Segmentation of a High-Resolution Urban Scene Using Texture Operators," *Computer Vision, Graphics, and Image Processing,* vol. 25, pp. 273-310, 1984.

[22] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. on Systems, Man and Cybernetics,* Vols. SMC-3, pp. 610-621, 1973.

[23] F. Albregtsen, "Statistical Texture Measures Computed from Gray Level Coocurrence Matrices," Image processing laboratory, University of Oslo, 2008.

[24] R. M. Haralick and K. Shanmugam, "Textural features for Image classification," *IEEE transactions on systems, man and cybernetics,* Vols. SMC-3, pp. 610-621, 1973.

[25] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research,* vol. 3, pp. 1157-1182, 2003.

[26] H. Liu, H. Motoda, R. Setiono and Z. Zao, "Feature Selection: An Ever Evolving Frontier in Data Mining," in *The Fourth Workshop on Feature Selection in Data Mining*, 2010.

[27] P. Joshi, A. Selwal and A. Sharma, "An effective algorithmic approach for fitness function in clustering," *International journal of computing and corporate research,* vol. 3, no. 3, 2013.

[28] F. Keyvanfard, M. A. Shoorehdeli and M. Teshnehlab, "Feature selection and classification of breast MRI lesion based on Multi classifier," *IEEE,* 2011.

[29] B. Bhanu and Y. Lin, "Genetic algorithm based feature selection for target detection in SAR images," *Elsevier Science,* vol. 21, pp. 591-608, 2003.

[30] T. Selvakumar and D. L. S. Jayashree, "C;assification of dermoscopic images for studying cancer and non cancer," *Inernational journal of innovative research in computer and communication engineering,* vol. 3, no. 5, 2015.

[31] R. C. Patil and D. A. S. Bhalchandra, "Brain tumor extraction from MRI images using MATLAB," *International Journal of Electronics, Communication & Soft Computing Science and Engineering,* vol. 2, no. 1, 2000.

[32] U. Javed, M. M. Riaz, A. Ghafoor and T. A. Cheema, "MRI Brain classification using texture features, fuzzy weighting and support vector machine," *Progress In Electromagnetics Research B,* vol. 53, pp. 73-88, 2013.

[33] V. P. G. P. Rathi and D. S. Palani, "Brain tumor MRI image classification with feature selection and extraction using Linear Discriminant Analysis," *IEEE,* 2000.

[34] M. E. Sweety and G. W. Jiji, "Detection of Alzheimer Disease in Brain Images using PSO and Decision Tree approach," in *2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, 2014.

[35] H. R. Bittencourt and D. A. d. Oliviera Moraes, "A binary decision tree classifier implementing logistic regression as a feature selection and classificiation method and its comparision with Maximum Likelihood," *IEEE,* 2013.