**A PROJECT**

**ON**

**SEGMENTATION AND CLASSIFICATION OF BRAIN TUMOR FROM MRI IMAGES USING FUZZY C-MEANS (FCM) AND SUPPORT VECTOR MACHINE (SVM) CLASSIFIER.**

****

This project paper submitted in partial fulfillment of the requirements for the Degree of Bachelor of Science to the Department of Information and Communication Engineering at Islamic University, Kushtia-7003, Bangladesh.

**Supervised By**

Dr. Md. Tariquzzaman

Associate Professor,

Dept. of Information and Communication Engineering

Islamic University, Kushtia-7003, Bangladesh.

**Submitted By**

Sharmen akhter

Roll No: 1118040

Reg. No: 1334

Session: 2011-2012

Dept. of Information and Communication Engineering

Islamic University, kushtia-7003, Bangladesh.

……………………………………………………………………………………….

Dept. of Information and Communication Engineering

Islamic University, kushtia-7003, Bangladesh

**CERTIFICATE**

**……………………………………………………………………………………….**

I am pleased to certify that Sharmen Akhter, Roll No: 1118040, Reg.No:1334 session:2011-2012, is a student of the Department of Information and Communication Engineering .She has performed a project work entitled “**SEGMENTATION AND CLASSIFICATION OF BRAIN TUMOR FROM MRI IMAGES USING FUZZY C-MEANS (FCM) AND SUPPORT VECTOR MACHINE (SVM) CLASSIFIER”** under my supervision in the academic year 2011-2012 for the fulfillment of partial requirement of Bachelor of Science (B.Sc.) Degree. I strongly declare that, this dissertation has not been copied from any other project or submitted to elsewhere prior submission of the department of Information of Communication Engineering, Islamic University, Kushtia-7003.

I wish her every success in life

……………………………

Dr. Md. Tariquzzaman

Associate Professor,

Dept. of Information and Communication Engineering

Islamic University, kushtia-7003, Bangladesh.

# Dedicated

# To

# My parent

**ACKNOWLEDGEMENT**

**………………………………………………………………………………………**

At first I would like to thank the Almighty Allah for completing a big project. Regarding the outcome of my project ,I would like to express my deepest sense of gratitude and respective to my supervisor Dr. Md. Tariquzzaman , Associate Professor, Dept. of Information and Communication Engineering, Islamic University, Kushtia, Bangladesh for his suggestion to select the topic and constant guidance ,great supervision ,advice and other fruitful helps throughout the duration of my project .without his active support and great supervision ,I would not be able to complete the project.

I also wish to acknowledge my parent and some of my close friend and senior sisters and brothers for their continued supports throughout the project .I wish to thank and convey my respect to our Honorable Course Co-coordinator , and all my honorable teacher of the department of information and Communication Engineering, Islamic university for their valuable suggestions and academic help. Finally, I would like to think Almighty Allah the deepest of my heart for making me able to finish this project work.

Author

Sharmen Akhter

12th march ,2018

**Abstract**

Brain tumor is a dangerous threat for human life. Magnetic resonance imaging (MRI) is the very effective technique to detect brain tumor .Using MRI images detection and classification of brain tumor has become possible to be automated as well as effective. In this project data mining methods are used for classification of MRI images. A hybrid technique based on the fuzzy c-means (FCM) for segmentation and support vector machine (SVM) for brain tumor classification is used. The purposed technique is a combination of FCM and SVM (hybrid technique) for prediction of brain tumor. Here suspicious region in MRI images is segmented to detect brain tumor using FCM clustering technique, before that noise is reduced using Median Filter. Feature extraction is done by implementing GLCM (Gray Level Co-occurrence Matrix) features on segmented image to obtain feature vector. At final step obtained feature vector is used to classify brain tumor (Benign or Malignant) using SVM classifier.

**Keywords:** MRI, Brain Tumor, Median Filter, RGB, Binary Image, Otsu,FCM, GLCM, SVM, Supervised Learning

**Table of Contents**

List of Figures8

List of Tables9

Chapter 1: Introduction10

1.1 Introduction11

1.2 Literature Survey 13

1.3 Objective of the Project 15

1.4 Project Organization16

Chapter 2: MRI Images, Brain Tumor and Techniques 17

2.1 Introduction18

2.2 MRI20

2.3 Image processing 22

2.4 Fuzzy C- Means24

2.5 GLCM25

2.6 Features26

2.7 Support Vector Machine (SVM) 27

Chapter 3: Segmentation and Classification29

3.1 Introduction30

3.2 Procedure30

3.2.1 Training The Model…………………………………………...31

3.2.2 Loading Images 31

3.2.3 Image Processing31

3.2.4 Clustering33

3.2.5 Segmentation33

3.2.6 Feature Extraction34

3.2.7 Training38

3.2.8 Testing40

3.3 Classification 40

Chapter 4: Results and Discussion41

4.1 Dataset42

4.2 Image Processing42

4.3 Segmentation43

4.4 Feature Optimization43

4.5 SVM Training44

4.6 Testing the Model45

4.7 Performance Analysis46

Chapter 5: Conclusion and Future Scope47

5.1 Conclusion48

5.2 Future Scope49

Bibliography50

**List of Figures**

|  |  |  |
| --- | --- | --- |
| Figure 1.1 | Brain tumor MRI | 12 |
| Figure 2.1 | Block diagram of training model | 19 |
| Figure 2.2 | Block diagram of testing model | 20 |
| Figure 2.3 | MRI of human brain | 22 |
| Figure 2.4(a) | Noisy image | 23 |
| Figure 2.4(b) | Clear image | 23 |
| Figure 2.5 | GLCM example | 26 |
| Figure 2.6 | Feature matrix example | 27 |
| Figure 2.7 | Support vector machine | 28 |
| Figure 3.1 | Graphical user interface | 30 |
| Figure 3.2(a) | Noisy image | 32 |
| Figure 3.2(b) | Filtered image | 32 |
| Figure 3.3 | Iteration example | 33 |
| Figure 3.4(a) | Original image | 34 |
| Figure 3.4(b) | Using otsu threshold level | 34 |
| Figure 3.4(c) | Using FCM 0 level | 34 |
| Figure 3.4(d) | Using FCM 1 level | 34 |
| Figure 3.5 | Support vector machine | 40 |
| Figure 4.1 | Median calculation | 42 |
| Figure 4.2 | Segmented image using FCM | 43 |
| Figure 4.3 | Feature interface | 43 |
| Figure 4.4 | Linear SVM | 45 |

**List of tables**

|  |  |  |
| --- | --- | --- |
| Table 4.1 | Feature matrix for this project | 44 |
| Table 4.2 | Observed and Predicted output | 45 |

**Chapter 1: Introduction**

* 1. **INTRODUCTION**

Brain tumor is the uncontrolled growth of the brain tissue, which causes abnormalities in the functioning of the brain. Brain tumors are of two type first one is the tumor that is originated at brain tissue itself and another one is started on another part of the body and migrate towards the brain. These are the one of the unavoidable reason of cancer-related death. Types of brain tumor are

1) Benign

2) Malignant

**Benign**

Benign tumors are typically slow-growing and rarely spread to other areas of the body.  They often have well-defined borders, so surgical removal can be an effective treatment.  However, the location of a benign brain tumor can have a significant impact on treatment options and be as serious and life-threatening as a malignant tumor.  Benign brain tumors can be considered malignant if they are located in areas of the brain that control vital functions like breathing.

**Malignant**

Unlike benign tumors, the cell structure of a “malignant” brain tumor is significantly different than that of “normal” brain cells. Malignant tumors tend to grow faster and can be more invasive than benign tumors.    Malignant tumors are life threatening. Sometimes malignant brain tumors are referred to as “brain cancer,” though they do not share all of the characteristics of cancer.  Most notably, cancer is characterized by the ability to spread from one organ to another.  It is very rare for a primary brain tumor to spread beyond the brain or spine.

This enhanced the growth in researches on the tumor detection and helped the doctors to rescue lives by detecting the disease earlier and initiates necessary treatment. Currently, radiologists locate the tumors in MRI images by manual and time consuming method, this manual detection process is unreliable, tedious and time-consuming.

Besides, this manual detection technique requires prior domain knowledge and experience on radiology for efficient and accurate tumor detection, classification and to know about the state of the tumor in medical imaging. Automation of tumor detection and then to know about the state of the tumor is required, as there is a shortage of skilled radiologists at a time of great need.

There are a number of image processing techniques are available for tumor detection from MRI image will detect certain features of the tumors such as spatial relationship, area of the tumor, number of tumor, energy, entropy, shape, border, calcification, and texture. These features are responsible the detection processes more accurate and easier, as there are some standard characteristics of each feature for a specific tumor.



Fig: 1.1 Brain Tumor MRI

With the passage of time due to increase in the world population, brain tumor is the growing health problem. As per the statistics, In each year, the population of cancer people is about 12.7 million among them 7.6 million peoples dies because of cancer [1. The basic block diagram for this project consists of five modules:

1. Image preprocessing

2. Feature extraction

3. Segmentation

4. Feature optimization

5. Training and testing

The acquired images may be very noisy because of the physical process of imaging. The presence of the noise can misclassify the images and may degrade the performance of the classifier.

Image preprocessing is the process which enhances the image by filtering approach. Otsu technique and FCM is used for clustering for segmentation by determining threshold and converting images into binary.

Feature extraction is the quantitative measurement of the images. The feature extraction, image data are transformed into some statistical numeric value. There are various features which can be extracted from the image such as:

* Contrast
* Homogeneity
* Correlation
* Energy
* Entropy etc.
* Area.

GLCM is a technique which extracts a features form gray as well as binary images.

Extracted feature is the most import input for the classifiers. Classifier analyses the input data features and classify the images accordingly.

There are a number of learning classifiers such as support vector machine (SVM), k- nearest neighbor KNN, artificial neural network (ANN), Hidden Markov Model (HMM), Probabilistic Neural Network (PNN). Every classifier has its own advantages as well as disadvantages. KNN has advantages when feature sets are less, but when features set increases the KNN performance also degraded. ANN is fast and robust, but computing cost is more hence consuming high CPU’s primary physical memory. SVM show more accuracy than other algorithms [1]. That is why in this project SVM is used to train the model and test the model.

**1.2 LITERATURE SURVEY**

Ivana Despotovi (2013), presented a new FCM-based method for spatially coherent and noise-robust image segmentation. The contribution was

1) The spatial information of local image features is integrated into both the similarity measure and the membership function to compensate for the effect of noise and

2) An anisotropic neighborhood, based on phase congruency features, is introduced to allow more accurate segmentation without image smoothing. The segmentation results, for both synthetic and real images, demonstrate that our method efficiently preserves the homogeneity of the regions and is more robust to noise than related FCM-based methods.

Maoguo Gong (2013), presented an improved fuzzy C-means (FCM) algorithm for image segmentation by introducing a tradeoff weighted fuzzy factor and a kernel metric. The tradeoff weighted fuzzy factor depends on the space distance of all neighboring pixels and their gray-level difference simultaneously. The new algorithm adaptively determined the kernel parameter by using a fast bandwidth selection rule based on the distance variance of all data points in the collection. Furthermore, the tradeoff weighted fuzzy factor and the kernel distance measure are both parameter free. Experimental results on synthetic and real images show that the new algorithm is effective and efficient, and is relatively independent of this type of noise.

Bhagwat et al (2013) they showed that DICOM images produce better results as compared to non medical images. They found that time requirement of hierarchical clustering was least of three and that for Fuzzy C means it was highest for detection of brain tumor. K-means algorithm produces more accurate result compared to Fuzzy c-means and hierarchical clustering.[13]

A.Sivaramakrishnan and Dr.M.Karnan(2013) proposed a novel and an efficient detection of the brain tumor region from cerebral image was done using Fuzzy C-means clustering and histogram. The histogram equalization was used to calculate the intensity values of the grey level images. The decomposition of images was done using principle component analysis which was used to reduce dimensionality of the wavelet co - efficient. The results of the proposed Fuzzy C-means (FCM) clustering algorithm successfully and accurately extracted the tumor region from brain MRI brain images[11]

Jaskiratkaur et al (2012), described clustering algorithms for image segmentation and did a review on different types of image segmentation

techniques. They also proposed a methodology to classify and quantify different clustering algorithms based on their consistency in different applications. They described the various performance parameters on which consistency will be measured.

Roy et al (2012) calculated the tumor affected area for symmetrical analysis. They showed its application with several data sets with different tumor size, intensity and location. They proved that their algorithm can automatically detect and segment the brain tumor. MR images gives better result compare to other techniques like CT images and X-rays.. Image pre-processing includes conversion of RGB image into grayscale image and then passing that image to the high pass filter in order to remove noise present in image.[14]

B. Sathya et al (2011), proposed four clustering algorithm; k mean, improved k mean, c mean and improved c mean algorithm. They did an experimental analysis for large database consisting of various images. They analyzed the results using various parameters

Hui Zhang et al (2008), compared subjective and supervised evaluation methodology for image segmentation. Subjective evaluation and supervised evaluation, are infeasible in many vision applications, so unsupervised methods are necessary. Unsupervised evaluation enables the objective comparison of both different segmentation methods and different parameterizations of a single method.[6]

Martial Heber et al (2005), presented an evaluation of two popular segmentation algorithms, the mean shift-based segmentation algorithm and a graph-based segmentation scheme

* 1. **OBJECTIVES OF THE PROJECT**

The condition of the brain is generally reflected in nature of MRI whether it is tumor affected or not and if so then what the type of this. If MRI is properly analyzed then important information regarding various disorders and diseases related to brain can be resolved. However, clinical observations for analyzing MRI images manually is very time consuming and tedious enough. Visual analysis can be relied upon as the possibility of the analyst missing the vital information is high. Sometimes the misleading information may have negative impact on the human health and may be the reason of death.

The most difficult problem is in automatic MRI analysis and feature computation is false positive outcome and misclassification. The matter here we are working on is very sensitive.

Thus, the objective of the project is to propose and build an efficient model to detect tumor cell and to predict about the condition of the brain tumor. Reducing false positive outcome, increasing efficiency of testing, automatic and effective model building are the motivated objectives of this project are considered.

* 1. **PROJECT ORGANIZATION**

This project document consists of five chapters including this chapter. The following 4 chapters denote the idea of the way the project is organized:

**Chapter 2: MRI Images, Brain Tumor and Techniques**

Introduction, MRI, Image processing, Fuzzy C Means, GLCM, Features, Support Vector Machine (SVM), Different features and considerations of brain tumor, and automatic MRI images detection is described in this chapter. Survey of the related work is carried out in this chapter.

**Chapter 3: Segmentation and Classification**

Concepts of detection and classification are carried in this chapter. Gray scale conversion, binary conversion of images, FCM, clustering, segmentation, detection, feature extraction using GLCM and classification using SVM procedures are clearly described in this chapter.

**Chapter 4: Result and discussion**

In chapter 4 there has been discussed about result and discussion. Performance of the project, accuracy, false positive, ability to meeting the goal of the projects, project analysis and proposed techniques are completely described in this chapter. This chapter consists of the experimental results and discussions.

**Chapter 5: Conclusion and future work**

Analytical remarks to overall achievements and limitations of all the proposed works and scope for further research work study about this project are illustrated in this chapter.

**Chapter 2**

**MRI Images, Brain Tumor and Techniques**

.

**2.1 INTRODUCTION**

This model belongs to supervised machine learning category. This project must propose a technique to build the automatic brain tumor detection and classification using the tools that are efficient enough. The following tools are used to build the project successfully with the expectation of high accuracy and most desired outcome.

Here the project is divided into two segment. They are :

1. Training
2. Testing

**Training**

The training part is to train the model according to the dataset. In this process the following technique is followed:

1. Input MRI from training dataset
2. Converting RGB images into gray images
3. Converting gray images into binary images for segmentation using Otsu and FCM
4. Feature extraction using GLCM feature extraction algorithm
5. Training the model using SVM classifier

The flow chart of the proposed system for training is given below at fig 2.1.

Input MRI from

training dataset

Converting RGB image into Gray Image

Converting Gray image into Binary Image

Clustering using Fuzzy C Means algorithm

Feature extraction using GLCM algorithm

Training the model using SVM

Fig 2.1 Block Diagram of Training the Model

**Testing**

The testing part is to test the model according to the dataset and to determine the condition of the brain tumor. In this process the following technique is followed:

1. Input MRI from training dataset
2. Converting RGB images into gray images
3. Converting gray images into binary images using Otsu and FCM
4. Feature extraction using GLCM feature extraction algorithm
5. Testing the condition of the image according to the trained model

Input MRI from training dataset

training dataset

Converting RGB image into Gray Image

Converting Gray image into Binary Image

Clustering using Fuzzy C Means algorithm

Feature extraction using GLCM algorithm

Testing the tumor condition according to the trained model using SVM

Fig 2.2 Block Diagram of Testing the model

**2.2 MRI**

Magnetic resonance imaging is a [medical imaging](https://en.wikipedia.org/wiki/Medical_imaging) technique used in [radiology](https://en.wikipedia.org/wiki/Radiology) to form pictures of the [anatomy](https://en.wikipedia.org/wiki/Anatomy) and the physiological processes of the body in both health and disease. [MRI scanners](https://en.wikipedia.org/wiki/Physics_of_magnetic_resonance_imaging#MRI_scanner) use strong [magnetic fields](https://en.wikipedia.org/wiki/Magnetic_field), [electric field gradients](https://en.wikipedia.org/wiki/Electric_field_gradient), and [radio waves](https://en.wikipedia.org/wiki/Radio_wave) to generate images of the organs in the body. MRI does not involve [X-rays](https://en.wikipedia.org/wiki/X-rays) and the use of [ionizing radiation](https://en.wikipedia.org/wiki/Ionizing_radiation), which distinguishes it from [CT or CAT scans](https://en.wikipedia.org/wiki/CT_scan). Magnetic resonance imaging is a [medical application](https://en.wikipedia.org/wiki/Nuclear_magnetic_resonance#Medicine) of [nuclear magnetic resonance](https://en.wikipedia.org/wiki/Nuclear_magnetic_resonance) (NMR). NMR can also be used for *imaging* in other [NMR applications](https://en.wikipedia.org/wiki/Nuclear_magnetic_resonance#Applications) such as [NMR spectroscopy](https://en.wikipedia.org/wiki/Nuclear_magnetic_resonance_spectroscopy).

While the hazards of X-rays are now well-controlled in most medical contexts, MRI may still be seen as a better choice than CT. MRI is widely used in hospitals and clinics for [medical diagnosis](https://en.wikipedia.org/wiki/Medical_diagnosis), [staging](https://en.wikipedia.org/wiki/Cancer_staging) of disease and follow-up without exposing the body to [radiation](https://en.wikipedia.org/wiki/Ionizing_radiation). However, MRI may often yield different diagnostic information compared with CT. There may be risks and discomfort associated with MRI scans. Compared with CT scans, MRI scans typically take longer and are louder, and they usually need the subject to enter a narrow, confining tube. In addition, people with some medical implants or other non-removable metal inside the body may be unable to undergo an MRI examination safely.

MRI was originally called 'NMRI' (nuclear magnetic resonance imaging) and is a form of NMR, though the use of 'nuclear' in the acronym was dropped to avoid negative associations with the word.[[1]](https://en.wikipedia.org/wiki/Magnetic_resonance_imaging#cite_note-1) Certain [atomic nuclei](https://en.wikipedia.org/wiki/Atomic_nucleus) are able to absorb and emit [radio frequency](https://en.wikipedia.org/wiki/Radio_frequency) energy when placed in an external [magnetic field](https://en.wikipedia.org/wiki/Magnetic_field). In clinical and research MRI, [hydrogen atoms](https://en.wikipedia.org/wiki/Hydrogen) are most often used to generate a detectable radio-frequency signal that is received by antennas in close proximity to the anatomy being examined. Hydrogen atoms exist naturally in people and other biological organisms in abundance, particularly in [water](https://en.wikipedia.org/wiki/Properties_of_water) and [fat](https://en.wikipedia.org/wiki/Lipid). For this reason, most MRI scans essentially map the location of water and fat in the body. Pulses of radio waves excite the [nuclear spin](https://en.wikipedia.org/wiki/Nuclear_spin) energy transition, and magnetic field gradients localize the signal in space. By varying the parameters of the [pulse sequence](https://en.wikipedia.org/wiki/Pulse_sequence), different contrasts may be generated between tissues based on the [relaxation](https://en.wikipedia.org/wiki/Relaxation_(NMR)) properties of the hydrogen atoms therein.

Since its development in the 1970s and 1980s, MRI has proven to be a highly versatile imaging technique. While MRI is most prominently used in [diagnostic medicine](https://en.wikipedia.org/wiki/Medical_diagnosis) and biomedical research, it also may be used to form images of non-living objects. MRI scans are capable of producing a variety of [chemical](https://en.wikipedia.org/wiki/Diffusion_MRI) and [physical](https://en.wikipedia.org/wiki/Functional_magnetic_resonance_imaging) data, in addition to detailed spatial images. The sustained increase in demand for MRI within [health systems](https://en.wikipedia.org/wiki/Health_system) has led to concerns about [cost effectiveness](https://en.wikipedia.org/wiki/Cost-effectiveness_analysis) and [over diagnosis](https://en.wikipedia.org/wiki/Overdiagnosis).

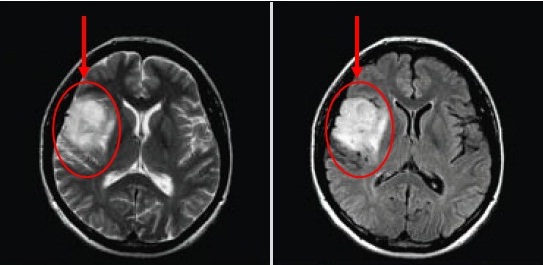


Fig 2.3 MRI of Human Brain

**2.3 IMAGE PROCESSING**

Image processing is the most important part of the machine learning project. Basically noisy images can contain wrong information about the object. In this project there some image processing techniques are used, they are

1. Noise reduction using median filter
2. Conversion of images from RGB to Gray

**Noise reduction using median filter**

The median filter is a nonlinear [digital filtering](https://en.wikipedia.org/wiki/Digital_filter) technique, often used to remove [noise](https://en.wikipedia.org/wiki/Signal_noise) from an image or signal. Such [noise reduction](https://en.wikipedia.org/wiki/Noise_reduction) is a typical pre-processing step to improve the results of later processing (for example, [edge detection](https://en.wikipedia.org/wiki/Edge_detection) on an image). Median filtering is very widely used in digital [image processing](https://en.wikipedia.org/wiki/Image_processing) because, under certain conditions, it preserves edges while removing noise (but see discussion below), also having applications in [signal processing](https://en.wikipedia.org/wiki/Signal_processing).

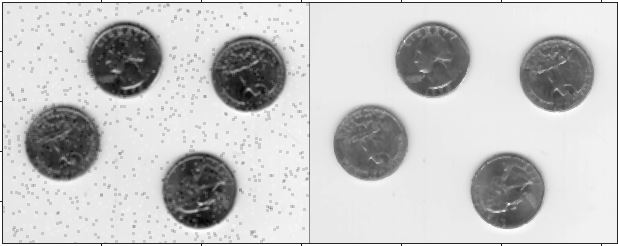


Fig 2.4 a) Noisy image b) Clear image

The median filter follows the following equation

Y[m,n] = median{x[ i,j],( i,j) € ω}……(2.1)

Where ω represents neighborhood definition by the user

**Conversion of images from RGB to Gray**

The basic 'algorithm' is to find three parameters, a,b,c, and set the gray-level value at each pixel, y, as

y=aR+bG+cB ……(2.2)

The naive way to set a,b,c is just an average: a=b=c=1/3. However, in practice this is not the case, and the parameters are based on the human visual systems: we do not see each color as a 'third' of the intensity, but rather we see more green than red and blue.   
Empirical research can get a good estimate of these parameters, and the final value is set to (approximately)

y=0.3R+0.6G+0.1B ……(2.3).

so that the single-channel (gray-level) image best corresponds with our perception of color intensities.

**2.4 FUZZY C- MEANS**

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by [Dunn in 1973](http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html#dunn) and improved by [Bezdek in 1981](http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html#bezdek)) is frequently used in pattern recognition. It is based on minimization of the following objective function:

 here ……..**(2.4)**

where m is any real number greater than 1, uij is the degree of membership of xi in the cluster j, xi is the ith of d-dimensional measured data, cj is the d-dimension center of the cluster, and ||\*|| is any norm expressing the similarity between any measured data and the center.Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership uij and the cluster centers cj by:





This iteration will stop when https://i0.wp.com/home.deib.polimi.it/matteucc/Clustering/tutorial_html/images/image027.gif, where https://i2.wp.com/home.deib.polimi.it/matteucc/Clustering/tutorial_html/images/image002.gif is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of Jm.The algorithm is composed of the following steps:

|  |
| --- |
| 1. Initialize U=[uij] matrix, U(0) 2. At k-step: calculate the centers vectors C(k)=[cj] with U(k) 3. Update U(k) , U(k+1) 4. If || U(k+1) – U(k)||<https://i2.wp.com/home.deib.polimi.it/matteucc/Clustering/tutorial_html/images/image002.gif then STOP; otherwise return to step 2. |

**2.5 GLCM**

To create a GLCM, use the [graycomatrix](https://in.mathworks.com/help/images/ref/graycomatrix.html) function. The function creates a gray-level co-occurrence matrix (GLCM) by calculating how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (i,j) in the resultant glcm is simply the sum of the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image.

The number of gray levels in the image determines the size of the GLCM. By default, grayco-matrix uses scaling to reduce the number of intensity values in an image to eight, but you can use the NumLevels and the GrayLimits parameters to control this scaling of gray levels. See the graycomatrix reference page for more information.

The gray-level co-occurrence matrix can reveal certain properties about the spatial distribution of the gray levels in the texture image. For example, if most of the entries in the GLCM are concentrated along the diagonal, the texture is coarse with respect to the specified offset. You can also derive several statistical measures from the GLCM. See [Derive Statistics from GLCM and Plot Correlation](https://in.mathworks.com/help/images/derive-statistics-from-glcm-and-plot-correlation.html) for more information.

To illustrate, the following figure shows how graycomatrix calculates the first three values in a GLCM. In the output GLCM, element (1,1) contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively. GLCM(1,2) contains the value 2 because there are two instances where two horizontally adjacent pixels have the values 1 and 2. Element (1,3) in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the values 1 and 3. graycomatrix continues processing the input image, scanning the image for other pixel pairs (i,j) and recording the sums in the corresponding elements of the GLCM.

**Process Used to Create the GLCM**

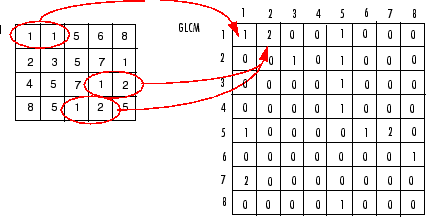


Fig 2.5 GLCM example

**2.6 FEATURES**

Here in this project the following features are considered to train and classify images and tumor types using Gray Level Co Matrix:

1 . Four spatial relationships –

* + - Left diagonal spatial relationship
    - Right diagonal spatial relationship
    - Vertical spatial relationship
    - Horizontal spatial relationship

1. Contrast
2. Homogeneity
3. Correlation
4. Area of tumor cell
5. Number of connected component

**Feature Matrix**

After extracting the features from the images there need to create a feature matrix to train the model. In the matrix each row represents the feature vector for each image and each column represents features. An example of feature matrix is shown in figure 2.3 where each row is a feature vector and each column is any particular feature.

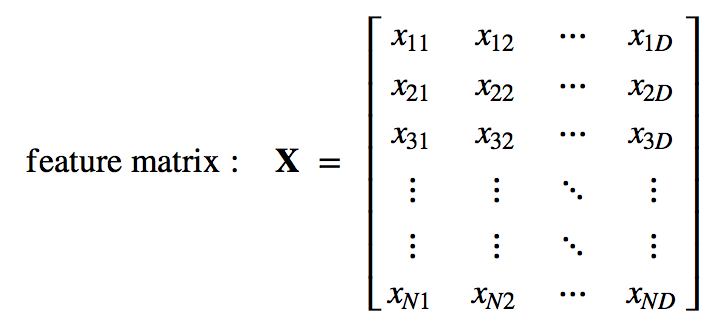


Fig 2.6 Feature Matrix example

**2.7 SUPPORT VECTOR MACHINE**

In [machine learning](https://en.wikipedia.org/wiki/Machine_learning), support vector machines (SVMs, also support vector networks[[1]](https://en.wikipedia.org/wiki/Support_vector_machine#cite_note-CorinnaCortes-1)) are [supervised learning](https://en.wikipedia.org/wiki/Supervised_learning) models with associated learning [algorithms](https://en.wikipedia.org/wiki/Algorithm) that analyze data used for [classification](https://en.wikipedia.org/wiki/Statistical_classification) and [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis). Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-[probabilistic](https://en.wikipedia.org/wiki/Probabilistic_classification) [binary](https://en.wikipedia.org/wiki/Binary_classifier) [linear classifier](https://en.wikipedia.org/wiki/Linear_classifier) (although methods such as [Platt scaling](https://en.wikipedia.org/wiki/Platt_scaling) exist to use SVM in a probabilistic classification setting).

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the [kernel trick](https://en.wikipedia.org/wiki/Kernel_trick), implicitly mapping their inputs into high-dimensional feature spaces.

When data are not labeled, supervised learning is not possible, and an [unsupervised learning](https://en.wikipedia.org/wiki/Unsupervised_learning) approach is required, which attempts to find natural [clustering of the data](https://en.wikipedia.org/wiki/Data_clustering) to groups, and then map new data to these formed groups. The support vector clustering[[2]](https://en.wikipedia.org/wiki/Support_vector_machine#cite_note-HavaSiegelmann-2) algorithm created by [HavaSiegelmann](https://en.wikipedia.org/wiki/Hava_Siegelmann) and [Vladimir Vapnik](https://en.wikipedia.org/wiki/Vladimir_Vapnik), applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data, and is one of the most widely used clustering algorithms in industrial applications.

Support Vectors: Input vectors that just touch the boundary of the margin (street) – circled below, there are 3 of them ( or, rather, the ‘tips’ of the vectors )

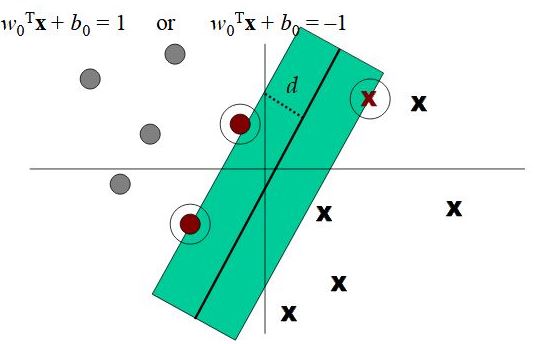


Fig 2.7 Support Vector Machine ( SVM )

**Chapter 3**

**Segmentation and Classification**

**3.1 INTRODUCTION**

Purpose of this chapter is to provide the details explanation of the project. This chapter is to enhance the proposed procedure that has been followed to build the automated and an efficient brain tumor detection and classification from MRI images using the methods that briefly discussed earlier.

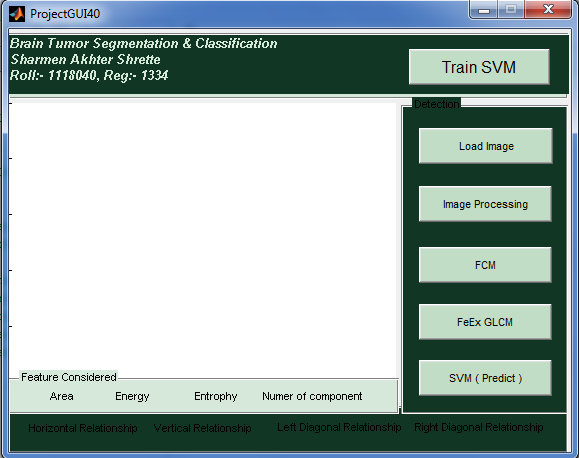


Fig 3.1 Graphical User Interface of the project

**3.2 PROCEDURE**

In this project two procedures are divided into two parts the whole procedure of the project is described in this chapter. The method considered to perform the task.

1. Training the model
2. Testing the model

**3.2.1 Training The model**

To build the system exactly according to the proposed methodology which is discussed in chapter 1 here the procedure has been followed is given below:

1. Loading Images

2. Image Processing

Noise reduction ( Median filter )

Gray Image conversion

3. Segmentation by Clustering

Fuzzy C Means ( FCM )

4. Features Selection

Gray Level Co-occurance Matrix ( GLCM )

5. Training the Model

Support Vector Machine ( SVM )

6. Classification based on Model

**3.2.2 Loading Image**

The dataset contains images to train and test the model and classify images from MRI images there’s need to loading data images and process them. Here the dataset stored in a folder and images are labeled by two distinct values ‘0’ and ‘1’. Here benign class is mapped into ‘0’ and malignant class is mapped into ‘1’. Browsing images folder to load images for processing is the first step of procedure.

**3.2.3 Image Processing**

The images may contain noise which can have negative impact on computation and classification efficiency. Median filter is used in this project to reduce noise of images.

**Median Filter**

The median filter follows the following equation

Y[m,n] = median{x[ i,j],( i,j) € ω}

Where ω represents neighborhood definition by the user

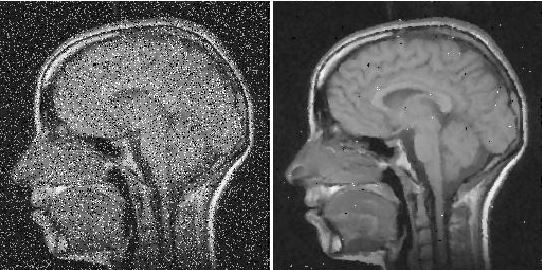


Fig 3.2 a) Noisy image b) Filtered image

In this project the MRI images of dataset may contain noise which can affect the performance of the project.

The median of each 8-pixel is placed by the current position. Also all the nine pixelsare replaced by the calculated median value.

The algorithm for median filter is as follows:

**Algorithm:**

1. [m,n] size of image
2. For each i to m-1 repeat step 2 to
3. For each j to n-1 repeat step 4 to
4. medEight = median(all eight neighboring pixel of im[ i , j ])
5. Update all nine pixel with medEight
6. End of step 3
7. End of step 2

**RGB to Gray Scale conversion**

The dataset contains raw images which are all basically in RGB format. There’s need to convert the RGB images into Gray Images and then binary images. The original RGB images contain tree dimensional pixels but in gray scale images the pixel value is one dimensional. Here the following equation is responsible for converting the RGB images into gray images.

New pixel value = ( max ( R , G , B ) + min( R , G , B) )/2

**3.2.4 SEGMENTATION BY CLUSTERING**

Gray level image is need to convert into binary images by Threshold Selection based on Otsu’s method.

If the new pixel value is greater or equal to the threshold value then the pixel is set to ‘1’ otherwise the pixel value is set to ‘0’. Hence the images is converted into binary images.

**Matlab function**: new\_bin\_image = im2bw(new\_image);

In this method gray image has been clustered by threshold selection based on Otsu method and segmentation and clustering is performed with Fuzzy C Means with different levels.

In this project there’s about 15 - 33 iterations are used to perform clustering using FCM for 0 and 1. The threshold value to perform binary operation is thus determined.

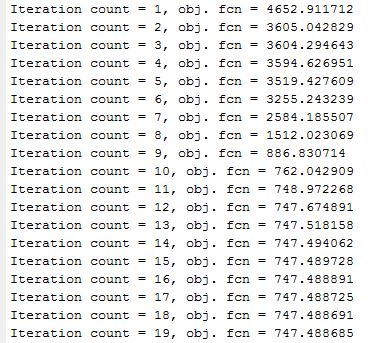


Fig 3.3 Iteration Example from this project

The following figure denotes the result of the clustering output from original images in this project.

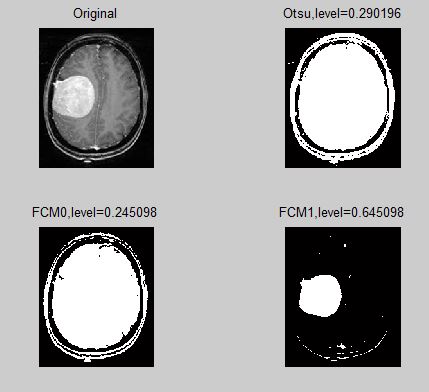


Fig 3.4 a) Original Image b) Using Otsu threshold level c) Using FCM0 and d) Using FCM1

From the figure 3.2 (d) the image denotes the clustered image. The pixels containing the value 1 is treated as tumor cell.

**3.2.5 FEATURE EXTRACTION**

Feature selection is the most important part of any machine learning project. Performance of the project definitely depends on features. To train any machine learning model features should be selected carefully. In this Gray Level Co Matrix method is used to extract features, besides this additional some features are considered. Here four spatial relationships are considered to extract feature from the segmented images is described earlier at article 2.3.

glcm =

46407 675

675 2461

**GLCM**

The glcm matrix is given where each value represent the four spatial relationship of MRI image. From that value the other features are estimated. Feature optimization is performed with the help of some mathematical model.

(i) Contrast

Local variance present in the brain tumor MRI can be measured using contrast. If P (i, j) in the matrix has more variations then the contrast will be high. The contrast value from the matrix can be obtained from

 …. (3.1)

(ii) Homogeneity

The homogeneity of an image can be found from the combination of low and high values of P(i, j) in the co-occurrence matrix. This feature results in spreading the P(i, j) values evenly in the matrix. Mathematically homogeneity can be expressed as

    …. (3.2)

(iii) Entropy

This is a measure of information content and randomness of intensity distribution. When all the entries of the matrix are of same magnitude then entropy is higher otherwise, it is smaller. The entropy can be measured using the equation (13).

 …  (3.3)

(iv) Correlation

Intensity of an image is measured using Correlation through the equation (14). If an image contains a considerable amount of linear structure then the correlation value will be higher.

…. (3.4)

Where and are the mean and standard deviation in the horizontal spatial domain and  are the mean and standard deviation in the vertical spatial domain.

(v) Energy

The Texture energy is measured by

… (3.5)

(vi) Maximum Probability

This feature corresponds to the strongest response. This can be expressed mathematically as

 …. (3.6)

(vii) Local Homogeneity, Inverse Difference Moment (IDM)

This is influenced by the feature homogeneity of an image. IDM values are low for the inhomogeneous images and high for homogeneous images. It can be measured as

 … (3.7)

(viii) Sum of square, variance

The element whose values differ greatly from the P(i,j)’s average value then for such elements the feature values are relatively high value. It can be computed as

….. (3.8)

(ix) Auto correlation

Coarseness and regularity of an images texture can be analyzed through Kaizer. Spatial relationship among the primitives can be measured using

 …. (3.9)

(x) Directionality

Total degree of directionality is calculated for the neighbours that are non-overlapping using

 …. (3.10)

(xi) Mean

Mean (M). The mean of an image is calculated by adding all the pixel values of an image divided by the total number of pixels in an image.

 …. (3.11)

**Feature Matrix:**

The feature matrix of this project for our train dataset is the matrix by combining the feature vector of each image. The feature vector of each image is the individual row of the feature vector. Each column of feature matrix is defined as the feature. The example feature vector is given below:

Feature vector:[1,1,1,0,0,1,0,1].

**3.2.6 Training**

Basically training means the parameter estimation of determination. In this project we used linear kernel SVM. In SVM the parameters are the:

1. Margin
2. Weigh matrix W
3. The slop of the line b

After extracting feature vector from each image, there need to make data matrix combining each row together. As we know the level of each row i.e either it is benign or malignant. Then we also have the level vector. In this project SVM is used to train the machine learning model.

**SVM**

Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well

Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/line).

1. Linear kernel:

The linear kernel is the simplest kernel. It is defined by

 …. (3.12)

2. Polynomial kernel:

The polynomial kernel is suited for problems with normalized training data.It is given by

 …. (3.13)

Where α is the slope, c is a constant and d is the polynomial degree.

3. Gaussian Radial Based Function kernel:

The GRBF kernel is defined as

 … (3.14)

where σ is an adjustable parameter. It is a non-linear kernel and is very sensitive to noise.

4. Exponential Radial Based Function kernel:

The ERBF kernel is close to GRBF with only the square of the norm removed. It is defined as

 … (3.15)

5. ANOVA kernel:

The ANOVA kernel is also a radial based function kernel. It is defined as

  …    (3.16)

6. Multilayer Perception kernel:

Multilayer Perception kernel is also called as Hyperbolic Tangent kernel or Sigmoid kernel. It is defined as

  …   (3.17)

where γ is the slope and r is the intercept constant.

7. Fisher kernel: Fisher kernel is defined is

 ….  (3.18)

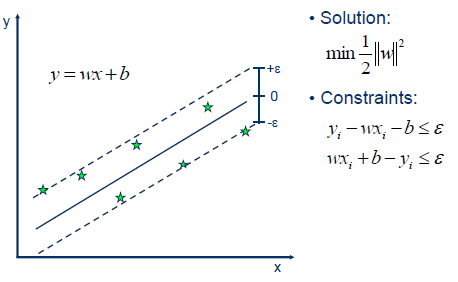


Fig 3.5 Support Vector Machines ( SVM )

In this project the linear kernel is applied to train the model.

**3.2.7 TESTING THE MODEL**

To test the model the whole procedure is repeated and the SVM trained model is used to classify the image or to test the condition of the tumor. The following chapter denotes the result and discussion.

**3.3 Classification**

Aim of SVM classifier is to group items that have similar feature values into groups. Classifier achieves this by making a classification decision based on the value of the linear combination of the features

**Chapter 4**

**Results and Discussion**

In this chapter, all the observed results using the procedures and algorithms are discussed in chapter 3 are explained at different sections. The MRI dataset is collected which contains RGB images the dataset is reliable enough. The effectiveness of image processing portion of this project using rgb2gray function, segmentation using FCM, feature extraction using GLCM and training and testing using SVM have been come to the discussion in this project. In this project all the MRI is in RGB format and contains noise with itself.

**4.1 Dataset**

The dataset is collected and assemble manually by own. There are two types of images at train dataset and test dataset, they are benign and malignant. All the images is affected by tumor. The train dataset is used to train the SVM model by processing image, extracting feature, segmentation and feature optimization. Half of the images are benign, and half of the images are malignant.

**4.2 Image Processing**

Image processing part is the combination of RGB to gray conversion and noise removal technique. RGB to gray conversion is perfumed with satisfactory level.

In noise removal section median filter is used to remove noise. The performance of the median filter is satisfied. This filter removes noise with high accuracy.

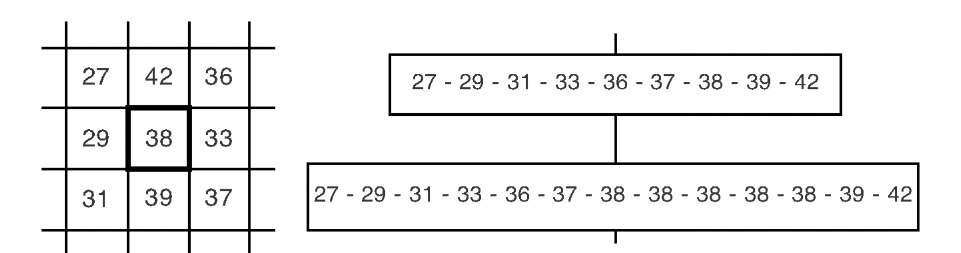


Fig 4.1 Median Calculation

**4.3 Segmentation**

In this section tumor is segmented from the image by removing background with the help of threshold level otsu and FCM clustering. The tumor portion is detected from the image. Determination of threshold level is an iterative process and after selecting threshold the image is converted into binary form to segment the tumor.

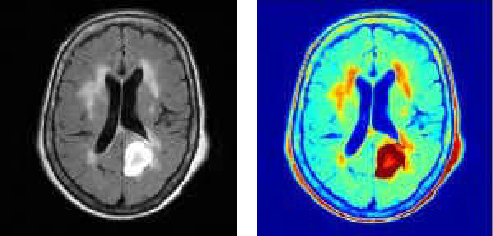
.

Fig 4.2 Segmented Image using FCM

The segmentation is done by converting the image into binary image. The main part of the segmentation process is threshold. In this project for each image the threshold value is determined by FCM algorithm. To determine the threshold for images some number of iteration is performed until the centroid point being unchangeable.

**4.4 Feature optimization**

After segmenting the tumor from MRI images, it’s needed to extract features from the segmented image as there is need feature matrix to train the model. The features are considered in this project is as follows:



Fig 4.3 Feature Interface

Feature is calculated using GLCM algorithm. The feature matrix is built by combining each row into one matrix. The feature matrix is given below:

Table 4.1 : Feature Matrix for this project

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of Tumor | Energy  10e9 | Entropy  10e5 | Horizontal  relation | Vertical  relation | Left  diagonal | Right  diagonal | Area | Level |
| 67 | 1.24e9 | 3.99e5 | 35045 | 188 | 188 | 3601 | 3789 | 1 |
| 47 | 2.01 | 5.22 | 44661 | 861 | 861 | 3927 | 4788 | 2 |
| 197 | 1.94 | 5.201 | 43822 | 1243 | 1246 | 4089 | 5342 | 3 |
| 78 | 1.97 | 5203 | 44094 | 927 | 927 | 4316 | 5243 | 4 |
| 74 | 1.84 | 5.16 | 42410 | 891 | 891 | 5968 | 6859 | 5 |
| 65 | 4827 | 2.39 | 21888 | 623 | 623 | 1672 | 2295 | 6 |
| 45 | 1.79 | 5.15 | 41785 | 956 | 956 | 6493 | 7688 | 7 |
| 64 | 1.9 | 5.19 | 43655 | 872 | 872 | 4791 | 5662 | 8 |
| 161 | 2.00 | 5.24 | 44525 | 620 | 620 | 4635 | 5255 | 9 |
| 99 | 1.61 | 5.08 | 39129 | 1248 | 1248 | 8469 | 9117 | 10 |
| 27 | 3479 | 1.96 | 18582 | 165 | 165 | 1599 | 1764 | 11 |
| 28 | 2.27 | 5.32 | 47608 | 445 | 445 | 1770 | 2215 | 12 |
| 121 | 6832508 | 3.08 | 25752 | 771 | 771 | 4347 | 5118 | 13 |
| 112 | 1.72 | 5.12 | 40868 | 1222 | 1201 | 6857 | 8100 | 14 |
| 37 | 1.92 | 5.23 | 44158 | 559 | 559 | 5052 | 5611 | 15 |
| 86 | 1.49 | 5.06 | 37119 | 1421 | 1421 | 10329 | 11985 | 16 |

**4.5** **SVM Training**

In this project linear SVM classifier is used to classify images. It classify. First the model is trained using feature matrix.

The parameters i.e weighted matrix and bias is determined according to the data and the equation is used here is

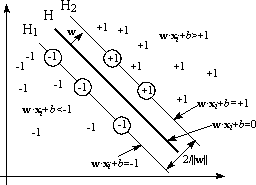


Fig 4.4 Linear SVM

**4.6 Testing the Model**

Testing is performed according to the model which is defined earlier. The parameter is used to determine the classification of the images. The test set data for the images is given below.

Table 4.2 Observed and predicted output

|  |  |  |
| --- | --- | --- |
| Serial No. | Observed | Predicted |
| 1 | Benign | Benign |
| 2 | Benign | Benign |
| 3 | Benign | Malignant |
| 4 | Malignant | Malignant |
| 5 | Malignant | Malignant |
| 6 | Malignant | Malignant |
| 7 | Malignant | Benign |
| 8 | Benign | Benign |
| 9 | Malignant | Malignant |
| 10 | Benign | Benign |

**4.7 Performance Analysis**

**Accuracy**

According the classification matrix

False positive = 2

False negative = 8

True positive = 8

True negative =2

Accuracy (%) = %

Accuracy (%) =

=80%

**Chapter 5**

**Conclusion and Future Scope**

**5.1 Conclusion**

In this project algorithm for brain tumor detection and classification is implemented using Median Filter, RGB to Gray conversion, segmentation using FCM and Otsu algorithm, Feature extraction using GLCM, feature optimization, training, testing and performance analysis.

The project is started with loading MRI images and end up with performance analysis.

The algorithms are used in this project is to enhance the efficiency of the model built earlier by other researchers.

The limitations and advantages of this project is described below:

There’s many filters to remove noise from MRI images, but median filter is used here for it’s high efficiency. Noise removal using median filter performs well as expected. But the result is not accurate. The result is optimal.

RGB to Gray conversion is done by using an algorithm described earlier.

Binary conversion of the images using Otsu threshold and FCM is done. Here a number of iteration is executed to determine the threshold of the image.

Segmentation is done using the binary images. Then the GLCM algorithm is used to extract features and the features are optimized. Here the most important features are used so that the model can be trained well. Experimental result and performance shows that the feature selection and optimization is done. The features are not perfectly accurate because the binary images may contain noises. But the result is highly acceptable.

Feature matrix is created with the feature vector of each images and the label of each feature vector is considered.

The model is trained according to the trainset i.e feature matrix. The parameters value are determined but there’s should be maintained high precision than that. But the accuracy of this project is optimal and expected.

The performance analysis denotes that there is a few scopes to enhance the performance of this project in future.

**5.2 Future Scope**

In this project the algorithms and techniques are implemented are very popular and examined. But there’s also scope to improve the performance of this project. In the following fields there are some scope to do better:

1. Noise removal performance can be enhanced
2. Determination of threshold for binary conversion and segmentation performance may be increased.
3. Feature selection process can be improved.
4. Optimizing the parameters for learning the model is done well but there’s also scope to increase the performance.
5. Training process can be make accurate with compare this project with increasing the size of dataset.

**Bibliography**

[1] Prakash Mahindrakar and Dr. M.Hanumanthappa, “Data Mining In Healthcare: A Survey of Techniques and Algorithms with Its Limitationsand Challenges”, Int. Journal of Engineering Research and Applications,ISSN : 2248- 9622, Vol. 3, Issue 6, Nov-Dec 2013, pp.937–941

[2] Kailash Sinha, G.R.Sinha, “Efficient Segmentation Methods for Tumor Detection in MRI Images”, 2014 IEEEStudent’s Conference on Electrical, Electronics andComputerScience,978 – 1 – 4799 – 2526-1/14/$31.00©2014 IEEE

[3] R.S.RajKumar and G.Niranjana, “Image Segmentationand Classification of MRIBrain Tumor Based on CellularAutomata and Neural Networks”, IJREATInternationalJournal of Research in Engineering & AdvancedTechnology, Volume 1, Issue 1, March, 2013ISSN: 2320–8791.

[4] H. B. Nandpuru, Dr. S. S. Salankar, Prof. V. R. Bora, MRI brain cancerclassification using support vector machine.2014 IEEE Students'Conference onElectrical, Electronicsand Computer Science, 978-1-4799-2526-1/14/$31.00 ©2014IEEEInternational Research Journal of Engineering and Technology

(IRJET)e-ISSN: 2395 – 0056Volume: 04 Issue: 05 | May – 2017[www.irjet.net](http://www.irjet.net)p-ISSN: 2395 – 0072©2017,IRJET |Impact Factor value: 5.181|ISO9001:2008 Certified Journal| Page796

[5] A. Padma and R.Sukanesh, “SVM based classification ofsoft tissues in brain CTimages using wavelet based dominant gray level run length texture features”, middle-east journal of scientific research, 2013, 13(7): 883-888.

[6] S.H.S.A. Ubaidillah, R. Sallehuddinand N.A. Ali, “Cancerdetection using artificialneural network and support vector machine:AComparative study”, jurnalteknologi(science & engineering),2013, 65:1.

[7] O.P. Verma, M. Hammandlu, S. Susan, M. Kulkami andP.K. Jain, "A simplesingle seeded region growingalgorithm for color image segmentation using adaptive

thresholding," 2011InternationalConference on Communication Systems and Network Technologies,©2011 IEEE.

[8] G.V. Kumar and Dr G.V. Raju, “Biological early brain cancer detection using artificial neural network”, International Journalon Computer ScienceandEngineering Vol. 02, No. 08, 2010, 2721–2725

[9] B. Gupta and S. Tiwari, “Brain Tumor Detection using Curvelet Transform and Support Vector Machine”,International Journal of Computer Science and Mobile Computing, Vol.3 Issue.4, April-2014, pg. 1259-1264, ISSN2320–088X 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN) 102.