

# Classification of Tumors in MRI Using Supervised Neural Networks

K.Praveena<sup>1</sup>, M.Preetha<sup>2</sup>, S.Vinitha<sup>3</sup>, S.Hariprasath<sup>4</sup>

<sup>1, 2, 3</sup>Final Year ECE, <sup>4</sup>Assistant Professor, Saranathan College of Engineering, Trichy  
hariprasath-ece@saranathan.ac.in

**Abstract:** Manual segmentation of brain tumors by medical practitioners is a time consuming task and has inability to assist in accurate diagnosis. Several automatic methods have been developed to overcome these issues. But Automatic MRI (Magnetic Resonance Imaging) brain tumor segmentation is a complicated task due to the variance and intricacy of tumors; to overcome this problem we have developed a new method for automatic classification of brain tumor. In the proposed method the MRI Brain image classification of tumors is done based on Gray level co-occurrence matrix and support vector machine classifier. By using this MRI we are going to extract the optimal features of brain tumor by utilizing GLCM algorithm with help of k-means Clustering Segmentation. Finally, Support vector machine classifier is utilized to perform two functions. The first is to differentiate between normal and abnormal. The second function is to classify the type of abnormality in benign or malignant tumor.

**Keywords:** Brain Cancer, Magnetic Resonance Image(MRI), k-means clustering, GLCM algorithm, Support Vector Machine(SVM)

## 1. Introduction

Medical imaging constitutes various modalities to create images of the human body such as Xray, Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and Ultrasound. X-RAY, invented by Winhelm in 1895 is the oldest source of electromagnetic radiation used for imaging having wavelength in the range of 0.01 to 10 nm. It emits high radiation to create images of the human body thereby causing side effects such as cancer and eye

cataract. Then Computed Tomography (CT) came into existence. CT scan utilizes X-ray to produce two dimensional images of the structures in the thin section of the body. CT produces a volume of data which can be manipulated through a process known as windowing, in order to demonstrate various bodily structures based on their ability to block the X-ray beam. Although most common in medicine, CT is also used in other fields such as non-destructive materials testing.

Since it has a greater ionizing radiation dose burden, repeated scans must be limited to avoid health effects. MRI or Magnetic Resonance Imaging invented in 1970 is a popular method in medical imaging. A magnetic resonance imaging instrument that is MRI scanner, or Nuclear Magnetic Resonance (NMR) imaging scanner as it was originally known, uses powerful magnets to polarise and excite hydrogen nuclei (single proton) in water molecules in human tissue, producing a detectable signal which is spatially encoded, resulting in images of the body.

MRI uses three electromagnetic fields, a very strong static magnetic field to polarize the hydrogen nuclei, called the static field, a weaker time-varying field for spatial encoding, called the gradient field and a weak radio-frequency field for manipulation of the hydrogen nuclei to produce measurable signals, collected through a Radio Frequency antenna. MRI traditionally creates a two dimensional image of a thin "slice" of the body and is therefore considered a homographic imaging technique. MRI clearly differentiates soft tissues and it produces no side effects like CT and X-rays. One of the challenging tasks in medical imaging is the segmentation of tumors by physicians. A tumor is a mass of tissue that grows as out of control of the normal forces that regulate growth. There are various types of tumors of which brain tumor is the cause

of one quarter of all cancer deaths. The complex brain tumors can be separated into two categories depending on tumor origin as primary and metastatic tumors. Primary brain tumors are tumors that arise from cells in the brain or from the covering of the brain. A secondary or metastatic brain tumor occurs when cancer cells spread to the brain from a primary cancer in another part of the body. Several automatic tumor segmentation methods which are rapid and time preserving have been developed.

The main contributions of our proposed technique are:

- Pre-processing the input MR images using Otsu's binarization.
- Feature extraction for the region is done with the help of GLCM algorithm.
- Selected features are input to Support Vector Machine (SVM) neural network for training, finally classify the tumor and non-tumor images.
- Used evaluation matrices parameters of sensitivity, specificity and accuracy to evaluate the performance of the proposed technique of tumor detection.

## Literature Review

The method proposed by Wankai Deng, Wei Xiao, He Deng, Jianguo Liu is based on gradients and variances along and inside of the boundary curve. MRI brain scans images of patients are obtained. Medical images are often superimposed with some noise. The images are filtered to get smooth homogenous areas without destroying the edges. Anisotropic diffusion filter is applied because they preserve the edge information without destroying it. By designing appropriate diffusion coefficient to control the proliferation of diffusion equation, the image can be smoothed by maintaining or even enhancing the image information.

The basic formula for selective smoothing algorithm is,

For image:  $I: \Omega \in R^2 \rightarrow R$

$$\frac{\partial I(X, Y, t)}{\partial x} = g(|S|) |\nabla I| \operatorname{div} \left( \frac{\nabla I}{|\nabla I|} \right)$$

$$I(X, Y, 0) = I_0(x, y)$$

Where

- $I(x, y, t)$  represents the continuous change of the image over time,
- $(x, y)$  represents the location in the image,
- $I_0(x, y)$  is the grey level of initial image to be processed,
- The term  $|\nabla I| \operatorname{div} \left( \frac{\nabla I}{|\nabla I|} \right)$  represents a degenerate diffusion term.

Generally diffusion function  $g(|s|)$  is used for the enhancement of the edges, which use the Gaussian function to smooth the image, and then calculate the gradient of processed image. The gradient values determine the smoothness of the image. The gradient decreasing function in different directions is always chosen as the diffusion coefficient of  $g(|s|)$ . In homogeneous region, gray changes little and gradient is small, so the diffusion coefficient is big, which can effectively smooth the noise inside the homogeneous region, while at the edge of the image, gray changes more violently, gradient is larger. So if the diffusion coefficient is smaller the image edge information is preserved. The gradients and variances along and inside of the boundary curve focuses on the consistency of the region and smoothness of the boundary. The mean variance inside of the boundary curve and the reciprocal of the mean gradient along the curve is chosen as the research subjects. There are some features that are similar in same region but they are distinct in different region. The features change smoothly in homogenous region, however in the boundary the changes are intensity relatively. The segmentation of boundary is better and the gradient of boundary is bigger. The more the region is segmented which contains same tissues, the smaller the mean variance of the region. The gray changes are sharp in tumor region. The threshold is increased gradually in region growing process and the set of the coarse contour can be obtained. We can calculate the gradient and the circumference of the contour. Suppose the iteration is  $t$ , the set of the coarse contour  $s = \{s_j\}$ , where  $0 < j < t$  and  $i$  is the pixel of gray-level,  $n_i$  is the number of pixels of  $i$ , we use the parameter  $l_j$  ( $0 < j < c$ ) as the set of coarse contour,  $m_j$  is the number of pixel on the outline  $l_j$ . Through the gradient image, we can get the gradient of each pixel and designate it by  $g_{s_j}$ .

$$g_{sj} = \frac{\sum_{k=1}^{n_j} g_k}{n_j}$$

$$G_{sj} = \frac{1}{g_{sj}}$$

The mean variance  $\sigma_{sj}^2$  of each  $s_j$  can be written as,

$$\mu_{sj} = \frac{\sum_{i=0}^{L-1} i n_i}{\sum_{i=0}^{L-1} n_i}$$

where  $i \in S_j$

$$\sigma_{sj}^2 = \frac{\sum_{i=0}^{L-1} (i - \mu_{sj})^2}{\sum_{i=0}^{L-1} n_i}$$

where  $i \in S_j$

The new objective function is defined using mean gradient and variance as follows:

$$h_{sj} = \alpha G + \beta \sigma$$

where  $\alpha, \beta$  are weight coefficients.

The selection of weight coefficient is done using circle to simulate the region of interest (ROI), the radius of the circle is  $r$ , the inner of circle is the normal distribution of gray, and the mean of the gray is 1, the mean square deviation is  $\delta$ , while outside of the circle, the gray value is 0. we can calculate the gradient and variance of the circle. The mean gradient round the circle is  $1/2\pi r$  and the mean variance is  $\delta/\mu r^2$ , so the  $\alpha$  and  $\beta$  from the equation is:

$$\alpha : \beta = \frac{\delta}{\mu} : 2\mu$$

$$\alpha + \beta = 1$$

Through optimizing the model, the optimal segmentation result can be obtained from the set of contours.

Pre-processing is done by filtering. Segmentation is carried out by advanced K-means and Fuzzy C-means algorithm. The feature extraction is done by considering the threshold and finally, approximating the reasoning method to recognize the tumor shape and position in MRI image is done using edge detection method. The proposed method is combinations of two algorithms established for segmentation. But they are not decent for all kinds of the MRI images. Pre-processing step converts the image according to the need for next level. Filtering of noise and other artifacts in the image and sharpening the edges in the image are performed here. RGB to gray conversion and reshaping also takes place here. It includes a median filter for noise removal. The

feature extraction is extracting the cluster, which shows the predicted tumor at the FCM (Fuzzy C-means) output. The extracted cluster is given to the threshold process. It applies a binary mask over the entire image. In the approximate reasoning step the tumor area is calculated using the binarization method making the dark pixel darker and white brighter. In threshold coding, each transform coefficient is compared with a threshold and if it is less than the threshold value, it is considered as zero or else one. In the approximate reasoning step the tumor area is calculated using the binarization method. That is the image having only two values either black or white (0 or 1). Here 256x256 JPEG image is a maximum image size. The binary image can be represented as a summation of total number of white and black pixels. A cluster is a collection of objects which are similar between them and are dissimilar to the objects belonging to other clusters. Clustering is an unsupervised learning method which deals with finding a structure in a collection of unlabelled data. A loose description of clustering could be the process of organizing objects into groups whose members are similar in some way. K-means clustering is an algorithm to group objects based on attributes/features into  $k$  number of groups where  $k$  is a positive integer. The grouping (clustering) is done by minimizing the Euclidean distance between the data and the corresponding cluster centroid. Thus the function of k-means clustering is to cluster the data. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. The Fuzzy C-Means (FCM) clustering algorithm was first introduced by Dunn and later was extended by 25 Bezdek. The algorithm is an iterative clustering method that produces an optimal  $c$  partition by minimizing the weighted within group sum of squared error objective

$$Y_m = \sum_{i=1}^N \sum_{j=1}^c U_{ij}^m \|X_i - C_j\|^2$$

where  $X = \{x_1, x_2, \dots, x_n\} \in R^p$  is the data set in the  $p$ -dimensional vector space,

$n$  is the number of data items,  $c$  is the number of clusters with  $2 \leq c < n$ ,  $u_{ik}$  is the degree of membership of  $x_k$  in the  $i$ th cluster,  $q$  is a weighting exponent on each fuzzy membership  $v_i$  is the prototype of the centre of cluster  $i$ ,  $d(x_k, v_i)$  is a distance measure between object  $x_k$  and cluster centre  $v_i$ . The feature extraction is extracting the cluster,

which shows the predicted tumor at the FCM (Fuzzy C-means) output. The extracted cluster is given to the threshold process. It applies a binary mask over the entire image. Each transform coefficient is compared with a threshold in threshold coding. If it is smaller than the threshold value, then it is taken as zero. If it is larger than the threshold, then it is taken as one. The thresholding method is an adaptive method where only those coefficients whose magnitudes are above a threshold are retained within each block. Let us consider an image 'f' that have the k gray level, an integer value of threshold T, which lies in the gray scale range of k. The thresholding process is a comparison process where each pixel in 'f' is compared to T, based on which, binary decision is carried out that defines the value of the particular pixel in an output binary image.

$$g(n) = \begin{cases} 0 & \text{if } f(n) \geq T \\ 1 & \text{if } f(n) < T \end{cases}$$

The tumor area is calculated using binarization method in approximate reasoning step.

### 3. Proposed System

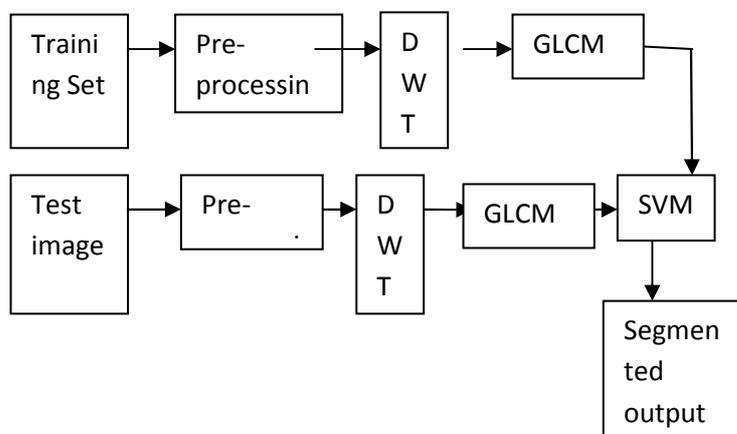


Fig 3.1: Proposed method block diagram

#### A. Preprocessing:

MRI images are altered by the bias field distortion. This makes the intensity of the same tissues to vary across the image. It is the first step in our proposed technique. The purpose of these steps is basically to remove the noise and improve the image quality to get more surety and ease in detecting the tumor. Here, the input image is applied with Otsu's binarization to remove the noise in order to obtain a better image. Also, improves the image quality.

#### Otsu's Binarization

Otsu's thresholding method involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels each side of the threshold, i.e. the pixels that either fall in foreground or background. The need is to find the threshold value where the sum of foreground and background spreads is at its minimum. The algorithm assumes that the image contains two classes of pixels following bi-modal histogram (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their combined spread (intra-class variance) is minimal, or equivalently (because the sum of pairwise squared distances is constant), so that their inter-class variance is maximal.

#### B. Feature Extraction

Studies have shown that the simple global features such as histogram of edge, gray or colour intensity, can represent images, and give the acceptable performance in image retrieval or recognition research fields. feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative

and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction.

When the input data to an algorithm is too large to be processed and it is suspected to be redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images presented as pixels), then it can be transformed into a reduced set of features (also named a feature vector). Determining a subset of the initial features is called feature selection. The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data.

## **DWT**

A discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information (location in time). Discrete wavelet transforms is the most popular transformation technique adopted for image compression. The wavelet transform has gained widespread acceptance in signal processing and image compression. Recently the JPEG committee has released its new image coding standard, JPEG-2000, which has been based upon DWT. Wavelet transform decomposes a signal into a set of basis functions. Image consists of pixels that are arranged in two dimensional matrix, each pixel represents the digital equivalent of image intensity. In spatial domain adjacent pixel values are highly correlated and hence redundant. In order to compress images, these redundancies existing among pixels needs to be eliminated. DWT processor transforms the spatial domain pixels into frequency domain

information that are represented in multiple sub-bands, representing different time scale and frequency points. One of the prominent features of JPEG2000 standard, providing it the resolution scalability, is the use of the 2D-DWT to convert the image samples into a more compressible form. The JPEG 2000 standard proposes a wavelet transform stage since it offers better rate/distortion (R/D) performance than the traditional DCT.

## **GLCM**

In statistical texture analysis, texture features are computed from the statistical distribution of observed combinations of intensities at specified positions relative to each other in the image. According to the number of intensity points (pixels) in each combination, statistics are classified into first-order, second order and higher-order statistics. The Gray Level Cooccurrence Matrix (GLCM) method is a way of extracting second order statistical texture features. The approach has been used in a number of applications, Third and higher order textures consider the relationships among three or more pixels. These are theoretically possible but not commonly implemented due to calculation time and interpretation difficulty.

## **C. Classification**

Classification is used to classify the tumour as normal, benign, malignant it helps to predict the feature using support vector machine (SVM). The extracted feature image is will be input to the classification system. Classification process is dividing into training phase and testing phase. In the training phase known data are given and the classifier is trained, and the testing phase unknown data are given and the classification is performed using trained classifier. classifications have the assignment to an

unknown pattern of a predefined class, according to the pattern presented in the form of a feature vector.

### SVM-Support Vector Machine

Support vector machine (SVM) is used to classify whether the test image comes under normal, benign and malignant. There are various methods for segmentation. SVM is an attractive and systematic method for two class problems. In this research work we are classifying images into two separate classes such as normal and abnormal using Support Vector Machine. This classifier is usually used in several research areas due to its out performance in pattern recognition and image processing tasks. Support vector machine (SVM) is extensively used in pattern recognition and image processing due to their originality of concepts using strong mathematical base and have strong realistic ability. SVM take intelligence from its training set to classify unknown data in testing phase. SVM suits most excellent for classification troubles with small training dataset and high dimensional feature space.

Like neural networks, SVM also needs two preparation stages; training and testing stage. SVM trains itself by features given as an input to its learning algorithm. The aim of SVM is to choose the suitable margins between two classes during training. Features are labeled according to class associative with particular class. Artificial neural network has a few issues having local minima and number of neurons selection for each problem. In order to resolve this problem SVM occupies no local minima and overhead to neurons selection by initiating the idea of hyper planes. The segmentation results are obtained for the purpose of classifying benign and malignant tumors. Classification is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of

data whose category membership had been defined. There are various algorithms for classification using a feature vector containing image texture contents.

SVM is a supervised learning method. It is a good tool for data analysis and classification. SVM classifier has a fast learning speed even in large data. SVM is used for two or more class classification problems. Support Vector Machine is based on the conception of decision planes. A decision plane is one that separates between a set of items having different class memberships. The Classification and detection of brain tumor was done by using the Support Vector Machine technique. Classification is done to identify the tumor class present in the image. The use of SVM involves two basic steps of training and testing.

## 4. Methodology

The test image is first resized into image of size 200\*200 for comparison with data base images. In Otsu's model operates directly on the gray level histogram. We have assumed that histogram is bimodal. Otsu's method chooses the threshold to minimize the intraclass variance of the black and white pixels.

$$\sigma_{\omega}^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t)$$

where class probabilities are

$$q_1(t) = \sum_{i=1}^1 \frac{iP(i)}{q_1(t)} \quad q_2(t) = \sum_{i=t+1}^I \frac{iP(i)}{q_2(t)}$$

class means are given by

$$\mu_1(t) = \sum_{i=1}^1 \frac{iP(i)}{q_1(t)} \quad \mu_2(t) = \sum_{i=t+1}^I \frac{iP(i)}{q_2(t)}$$

class variance are

$$\begin{aligned}\sigma_1^2(t) &= \sum_{i=1}^t [i - \mu_i(t)]^2 \frac{P(i)}{q_1(t)} \sigma_2^2(t) \\ &= \sum_{i=t+1}^l [i - \mu_i(t)]^2 \frac{P(i)}{q_2(t)}\end{aligned}$$

For this output, a single-level two-dimensional wavelet decomposition is performed. It computes the approximation coefficients matrix cA and details coefficients matrices cH, cV and cD (horizontal, vertical and diagonal respectively), obtained by wavelet decomposition of the input matrix.image. The matrix element  $P(i, j | \Delta x, \Delta y)$  is the relative frequency with which two pixels, separated by a pixel distance  $(\Delta x, \Delta y)$ , occur within a given neighborhood, one with intensity 'i' and the other with intensity 'j'. The matrix element  $P(i, j | d, \theta)$  contains the second order statistical probability values for changes between gray levels 'i' and 'j' at a particular displacement distance d and at a particular angle  $(\theta)$ . Using a large number of intensity levels G implies storing a lot of temporary data, i.e. a  $G \times G$  matrix for each combination of  $(\Delta x, \Delta y)$  or  $(d, \theta)$ . Due to their large dimensionality, the GLCM's are very sensitive to the size of the texture samples on which a GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G, in the they are estimated. We have extracted 13 features as

Contrast, Correlation, Energy, Homogeneity, Mean, Variance, Standard Deviation, Entropy, RMS, Smoothness, Kurtosis, Skewness, IDM.

Correlation

Measure the joint probability occurrence of the specified pixel pairs.

$$\sum_{m=0}^{G-1} \sum_{n=0}^{G-1} \frac{mnP(m, n) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

Where,

$$\mu_x = \sum_{m=0}^{G-1} m \sum_{n=0}^{G-1} P(m, n), \mu_y = \sum_{n=0}^{G-1} n \sum_{m=0}^{G-1} p(m, n)$$

$$\sigma_x = \sum_{m=0}^{G-1} (m - \mu_x)^2 \sum_{n=0}^{G-1} P(m, n)$$

Energy

Provide the sum of squared in the GLCM. Also known as uniformity or angular second moment.

$$\sum_{m=0}^{G-1} \sum_{n=0}^{G-1} P(m, n)^2$$

These features are used to classify the tumor type by comparing it with the training set. The input arguments for training algorithm is training and grouping. Training data is in the form of matrix, where each row corresponds to an observation or replicate, and each column corresponds to a feature or variable. For classification the test image features is compared with the training image features.

## 5. Result and Discussion:

The extracted features from the GLCM are given in table 1. It shows the extracted features for both benign and malignant type of tumors. The validation is performed to identify the accuracy of classification in each of the kernels. The accuracy of used kernel in percentage is shown in table2

Table 1: Features Extracted from the input images

| Image name     | Image1.jpg | Image 2.jpg |
|----------------|------------|-------------|
| Tumor Type     | Malignant  | Benign      |
| Contrast       | 0.3059     | 0.2558      |
| Correlation    | 0.1421     | 0.0895      |
| Energy         | 0.7862     | 0.7557      |
| Homogeneity    | 0.9379     | 0.9314      |
| Mean           | 0.0063     | 0.0025      |
| Std. Deviation | 0.0896     | 0.0898      |
| Entropy        | 3.2051     | 3.0756      |
| RMS            | 0.0898     | 0.0898      |
| Variance       | 0.0080     | 0.0081      |
| Smoothness     | 0.9591     | 0.9040      |
| Kurtosis       | 12.2408    | 7.7971      |
| Skewness       | 1.1048     | 0.5774      |
| IDM            | 1.2156     | -0.2601     |

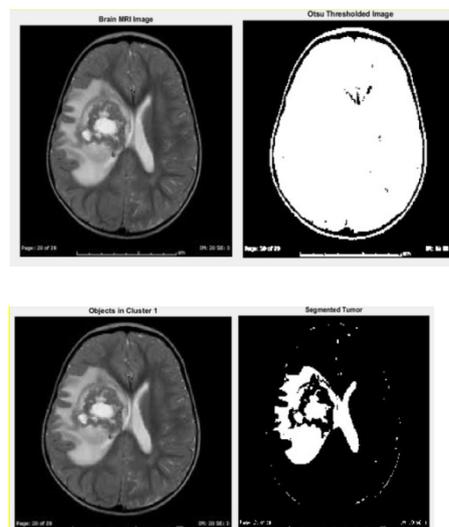


Figure 2. Clustering output of image1.jpg –malignant tumor

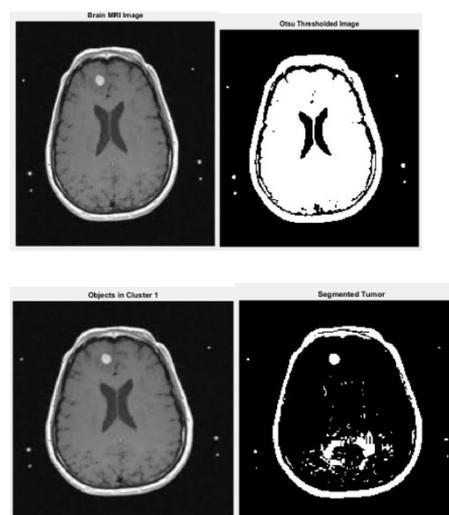


Fig 3: Clustering output of image2.jpg –malignant tumor

Table 2: Accuracy of kernel's in percentage

| Image Name   | Linear | RBF | Polynomial | Classification |
|--------------|--------|-----|------------|----------------|
| 1Perfect.jpg | 50     | 55  | 63.3333    | 100            |
| Perfect.jpg  | 70     | 65  | 63.3333    | 100            |
| 6Perfect.jpg | 40     | 45  | 53.3333    | 100            |
| 4Perfect.jpg | 40     | 50  | 50         | 100            |

The output is shown in figure 2 and figure 3.

## 6. Conclusion

The presence of brain tumor is segmented in an MR images and it is further classified into benign or malignant type of tumors. The input image is converted into gray scale image and by applying Otsu's binarization method binary image is obtained. By applying

DWT and using GLCM method, thirteen features are extracted and fed as input to neural network. The no of neurons in the input layer are 13 and 6 neurons are used in hidden layer and 2 neurons are used in the output layer. Using ann the accuracy for each kernel is calculated.

## References

[1] Alan Jose, Ravi.S ,Sambath.M, “Brain Tumor Segmentation Using K-Means Clustering And Fuzzy C-Means Algorithms And Its Area Calculation”, International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 3, March 2014

[2]SaeidFazli, ParisaNadirkhanlou “A Novel Method for Automatic Segmentation of Brain Tumors in MRI Images”, Research Institute of Modern Biological Techniques University of Zanjan, Iran, 2015.

[3] Wankai Deng , Wei Xiao, He Deng, Jianguo Liu ,“MRI Brain Tumor Segmentation With Region Growing Method Based On the Gradients and Variances Along and Inside of the Boundary Curve”, Biomedical Engineering and informatics(BMEI) 2010 3<sup>rd</sup> international conference on 16-18 Oct,2010.

[4]Wen-Liange, De-Hua Chen, Mii-Shen Yang, “Suppressed fuzzy-soft learning vector quantization for MR Segmentation”, Elsevier Ltd, Vol 52, Issue 1,Pag: 33-43, May2011.