Feature Extraction and Classification of Brain Tumor using MRI

Ananya S B[1] Lalitha N[2]
[1]M.Tech. Student [2]Assistant Professor
[1,2]Department of Electronics & Communication Engineering
[1,2]VVIET- Mysuru

Abstract— MRI scan of brain tumor gives the detailed information of brain compared to other scans. Brain tumor is an uncontrolled and an unwanted multiplication and growing of cells in the body Image processing in MRI of brain is highly essential due to accurate detection of the type of brain abnormality which can reduce the chance of fatal result. Segmentation technique plays a major role in the brain tumor detection. This paper outlines an efficient image segmentation technique for the different ventricles affected brain tumor images. K-means segmentation is used for the brain tumor detection and extraction. The extraction of texture features in the detected tumor has been achieved by using Gray Level Co-occurrence Matrix (GLCM). By selecting the significant features the Classification has been performed to labeling the images into normal and abnormal (tumor detected) using Support Vector Machine (SVM).

Key words: Brain Tumor, K-means Segmentation, Feature Extraction (GLCM), SVM classifier

I. INTRODUCTION

Tumor is defined as the abnormal growth of the tissues. Brain tumor is an abnormal mass of tissue in which cells grow and multiply uncontrollably, seemingly unchecked by the mechanisms that control normal cells. Brain tumors can be primary or metastatic, and either malignant or benign. A metastatic brain tumor is a cancer that has spread from elsewhere in the body to the brain. Magnetic Resonance Imaging (MRI) is an advanced medical imaging technique used to produce high quality images of the parts contained in the human body MRI imaging is often used when treating brain tumors, ankle, and foot. From these high-resolution images, we can derive detailed anatomical information to examine human brain development and discover abnormalities [1].

Image segmentation is an important and necessary task in image analysis and interpretation. It is a process of partitioning an image into non-overlapping homogeneous regions (e.g. having similar intensity, color, texture). There are different techniques for image segmentation; e.g. thresholding, clustering, classifications, region growing, artificial neural networks (ANNs), edge detection and region extraction, etc. Clustering is a process for segmenting and classifying similar objects or patterns to the same group.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved.

Texture analysis aims in finding a unique way of representing the underlying characteristics of textures and represent them in some simpler but unique form, so that they can be used for robust, accurate classification and segmentation of objects. [2]

Gray level co-occurrence matrix is formulated to obtain statistical texture features. A number of texture features may be extracted from the GLCM.

Classification is one of the most important tasks for different application such as text categorization, tone recognition, image classification, micro-array gene expression, proteins structure predictions, data Classification etc. In this work, Support Vector Machine (SVM) classifier is used. Support Vector Machine, is an effective machine method developed from analytical learning and has made powerful performance in some field. SVM method does not suffer the restrictions of data dimensionality and limited samples [5].

II. RELATED WORKS

Medical image segmentation is a very challenging task due to the complexity of the images. There are many segmentation method had been proposed in the existing system. These segmentation techniques involve thresholding, region based and clustering method. Thresholding typically does not take into account the spatial characteristics of an image [1][3][7]. This causes it to be sensitive to noise and intensity in homogeneities, which can occur in MR images.

The main drawback of region growing is that it requires manual interaction to obtain the seed point [11]. Selection of seed point in this method is very essential for the image segmentation. Image segmentation through clustering techniques includes many methods like c-mean, improved c-mean, k-mean and improved k-mean [4]. K-means segmentation gives best result than the fuzzy c-mean segmentation [12]

III. PROPOSED METHODOLOGY

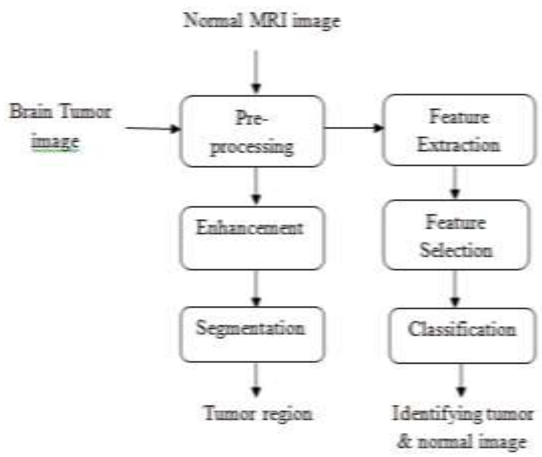


Fig 1: Block diagram of proposed methodology

A. *Preprocessing*

According to the need of the first level the preprocessing step convert the image. It performs filtering of noise in the image. RGB to grey conversion and Reshaping also takes place here. It includes median filter for noise removal.

1) Median Filter

In this paper we are using median filter for removing noise from an image. The median filter is a nonlinear digital filtering technique, is often used to remove noise. Median filtering is very widely used in digital image processing because, under certain conditions, it preserves edges while removing noise. The median filter is normally used to reduce noise in an image, somewhat like the, mean filter. However, it often does a better job than the mean filter.

2) Enhancement:

There are different enhancement methods in preprocessing, but in this work contrast enhancement is used. Contrast enhancement makes the object in the filtered image to looks much brighter than the object which is in the filtered image.

B. Segmentation

Image segmentation methods can be classified into three categories: Edge-based methods, region-based methods, and pixel-based methods [6].The K-means clustering technique is a pixel-based method, it is one of the most simple techniques, it's complexity is relatively lower than other region-based or edge-based methods. Furthermore, Kmeans clustering is suitable for biomedical image segmentation as the number of clusters is usually known for images of particular regions of the human anatomy. This segmentation gives the best result than the other segmentation techniques Combined with the existing methods and aiming to get better results, it is useful to take segmentation method into account.

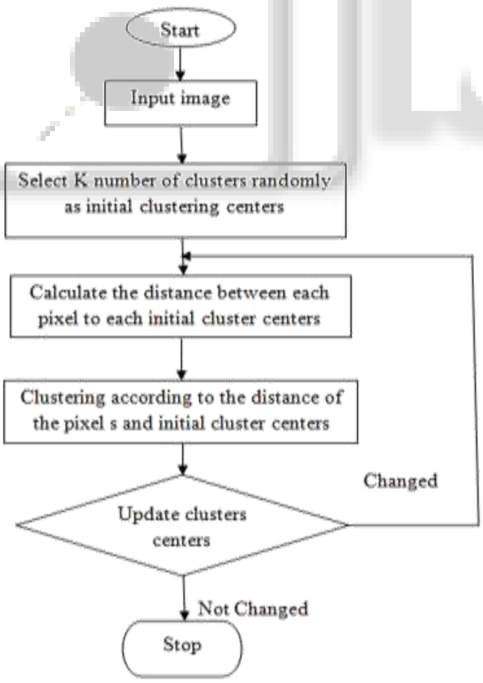


Fig 2: Flowchart of K-means algorithm

Steps for K-Means:

- Give the no of cluster value as k.
- Randomly choose k cluster centres
- Calculate mean or centre of the cluster
- Calculate the distance b/w each pixel to each cluster centre
- If the distance is near to the centre then move to that cluster.
- Otherwise move to next cluster.
- Re-estimate centre.
- Repeat the process until the centre doesn't move.

Fig 2 shows the flowchart of k-means algorithm. Along with these steps, the algorithm is modified by making cluster formation based on the pixel values and also makes the tumor region to be appearing at the final cluster. This modification helps for the elimination of skull region.

C. Feature Extraction

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Transforming the input data into the set of features is called feature extraction [3].

Gray-level co-occurrence matrix (GLCM) is the statistical method of examining the textures that considers the spatial relationship of the pixels. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. The graycomatrix function in MATLAB creates a gray-level cooccurrence 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*. 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 [8,9].

In this paper, we are extracting some of the features using GLCM as follows

1) Mean:

$$\mu_i = \sum_{i=0}^{G-1} i\,(P_{i,j}) = \mu_j = \sum_{i=0}^{G-1} j(P_{i,j}) \qquad (3.1)$$

The left hand equation calculates the mean based on the reference pixels, i. The right-hand equation calculates the mean over the the neighbor pixels, j.

2) Standard Deviation:

$$\sigma_x^2 = \sum_{i=0}^{G-1} P(i,j) = \sigma_j^2 = \sum_{i=0}^{G-1} P(i,j) \qquad (3.2)$$

3) Sum of Squares, Variance:

$$\mathrm{sos} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} (i-\mu)^2 P(i,j) \qquad (3.3)$$

This feature puts relatively high weights on the elements that differ from the average value of P(i, j).

4) Cluster Prominence:

$$Prom = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \{i + j - \mu_x - \mu_y\}^4 \times P(i,j) \qquad (3.4)$$

5) Homogeneity (Angular Second Moment) :

$$Homogenity = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \{P(i,j)\}^2 \qquad (3.5)$$

ASM is a measure of homogeneity of an image

6) Contrast:

$$contrast = \sum_{i,j=0}^{G-1} p_{i,j\,(i-j)^2} \qquad (3.6)$$

This measure of contrast or local intensity variation will favor contributions from P(i, j) away from the diagonal, i.e. $i \neq j$.

Local Homogeneity, Inverse Difference Moment (IDM):

$$IDM = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{1}{1+(i-j)^2} P(i,j) \qquad (3.7)$$

IDM is also influenced by the homogeneity of the image. The value of IDM is low for inhomogeneous images, and a relatively higher value for homogeneous images.

7) Entropy:

$$Entropy = \sum_{i,j=0}^{G-1} -ln(p_{i,j})p_{i,j} \qquad (3.8)$$

Inhomogeneous scenes have low first order entropy, while a homogeneous scene has a high entropy.

8) Correlation:

$$\mathrm{correlation} = \sum_{i,j=0}^{G-1} p_{i,j} \frac{(i-\mu)(j-\mu)}{\sigma^2} \qquad (3.9)$$

Correlation is a measure of gray level linear dependence between the pixels at the specified positions relative to each other.

9) Sum Entropy:

$$Sent = -\sum_{i=0}^{2G-2} P_{x+y}(i)\log(P_{x+y}(i)) \qquad (3.10)$$

10) Difference Entropy:

$$Dent = -\sum_{i=0}^{2G-1} P_{x+y}(i)\log(P_{x+y}(i)) \qquad (3.11)$$

11) Cluster Shade:

$$shade = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \{i+j-\mu_x-\mu_y\}^3 \times P(i,j) \qquad (3.12)$$

Along with these above features twelve more different features have also been calculated.

D. Feature Selection

In this stage, there is a selection of significant number of features from the extracted features, such that it does not make any affect at the classification output. It means, here initially twenty four different features have been extracted in the feature extraction stage. So in the classification process, only significant numbers of features are selected by eliminating the unwanted features. This feature selection provides the same result as the result obtained by using all features in the classification stage.

Five different features are reduced to get the expected result at the output. By using trial and error method feature selection task is performed. Finally twenty features have been selected. The five reducing features are: standard deviation, sum of square, energy, entropy and information measure of correlation.

E. SVM Classification

Classification is one of the most important tasks for different application such as text categorization, tone recognition, image classification, micro-array gene expression, proteins structure predictions, data Classification etc. Support Vector Machine, is an effective machine method developed from analytical learning and has made powerful performance in some field. A special property of SVM is that SVM is responsible to minimize the empirical classification error and maximize the geometric margin simultaneously. SVM mainly classifies the training data into two classes.

Here, training data includes MRI brain tumor image and normal brain image. In training sample, data are arranged as vector such that number of rows in vector indicates the different observations and number of columns represents the set of features. From the training sample data, this classifier differentiates the tumor and the normal brain image.

IV. EXPERIMENTS AND RESULTS

The experiment process involved several preparation operations, as follows:

A. Pre-processing Stage:

This includes the median filtering for the noise elimination. Fig 3,4 and 5 shows the preprocessing stage of different ventricle affected brain image.

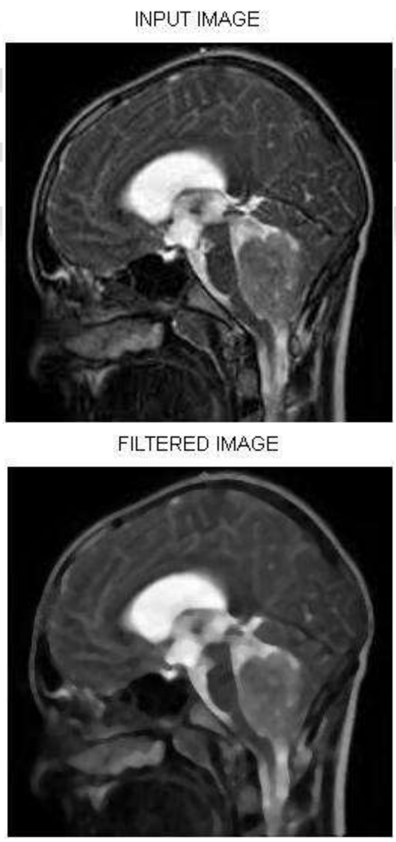
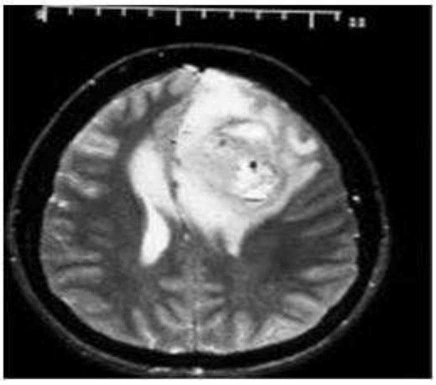


Fig 3: Preprocessing stage of lateral ventricle affected tumor

INPUT IMAGE



FILTERED IMAGE

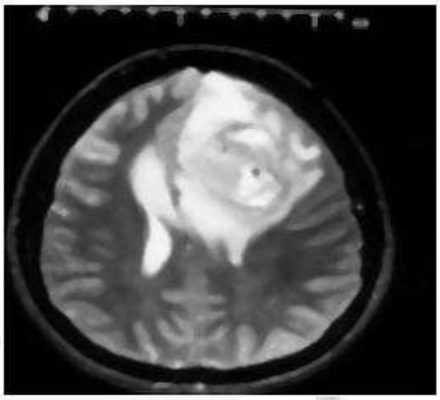


Fig 4: Preprocessing stage of third ventricle affected tumor

INPUT IMAGE

FILTERED IMAGE

Fig 5: Preprocessing stage of fourth ventricle affected tumor

B. K-Means Clustering Algorithm:

Tumor region can be identified and extracted in this stage. The result of this stage is shown in figure 6,7 and 8.

K - MEANS CLUSTERING

Actual tumor

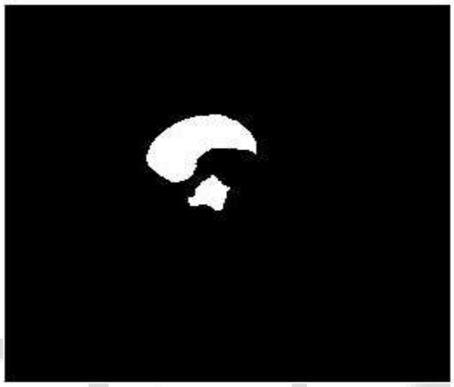
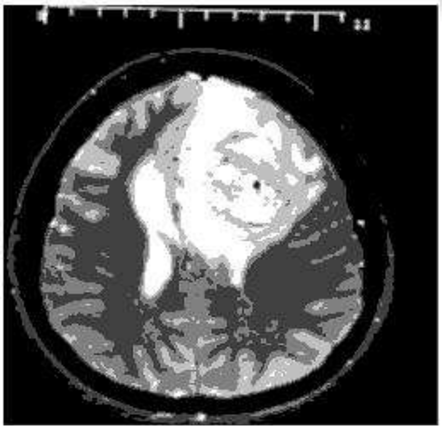


Fig 6: Segmentation stage of Lateral ventricles affected tumor MRI image

K - MEANS CLUSTERING



Actual tumor

Fig 7: Segmentation stage of third ventricle affected tumor MRI image

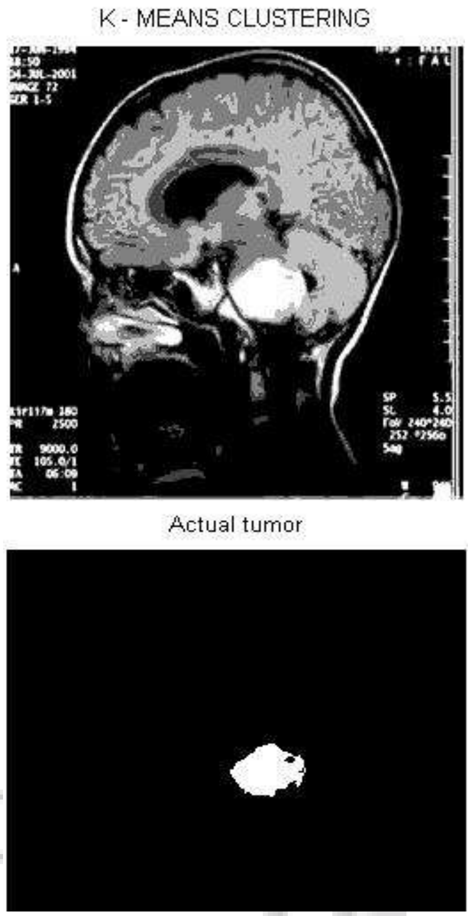


Fig 8: Segmentation stage of fourth ventricle affected tumor MRI image

C. Table 3.1: Tumor Area and Brain mask Calculation

Name	Area of tumor (sq.mm)	Area of brain with tumor (sq.mm)	Area of brain without tumor(sq.mm)
Image 1	1647	46872	45225
Image 2	2314	36436	34122
Image 3	1858	34695	32837
Image 4	569	15895	15326
Image 5	933	27249	26316
Image 6	307	30515	30208
Image 7	665	30446	29781
Image 8	3872	41143	37271
Image 9	1172	42352	41180
Image 10	950	36015	35065

Table 3.1: Tumor Area and Brain mask Calculation

The table shows the result of ten different brain tumor images, where calculating the area of the tumor region and also calculating the area of the brain with tumor region. From these two calculations it can be possible to find the remaining area of the brain by removing its tumor region.

D. Feature Extraction and Selection:

Fig 9 shows an output result of feature extraction of the tumor affected brain image. Here twenty four different features have been extracted and fig 10 shows the result of selected features from the output of feature extraction. Here feature extraction is applied to only one image. This process is same for the other ventricle affected image.

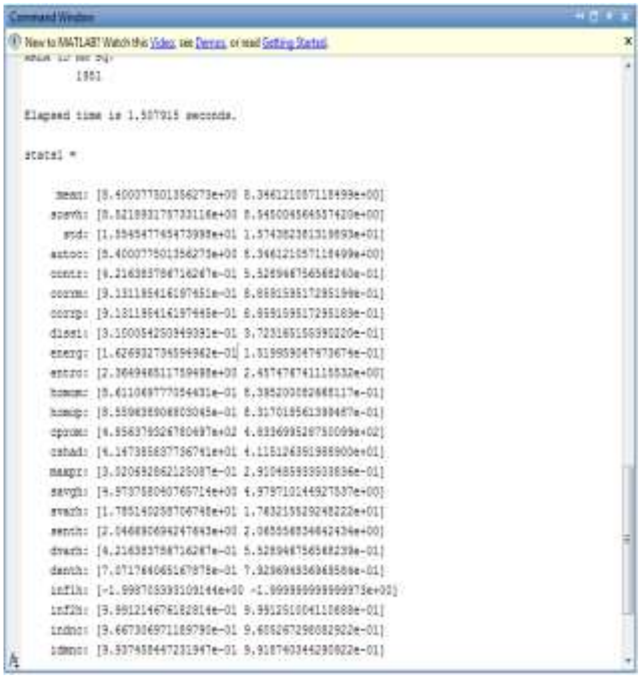


Fig 9: Feature extraction of Lateral ventricles affected tumor

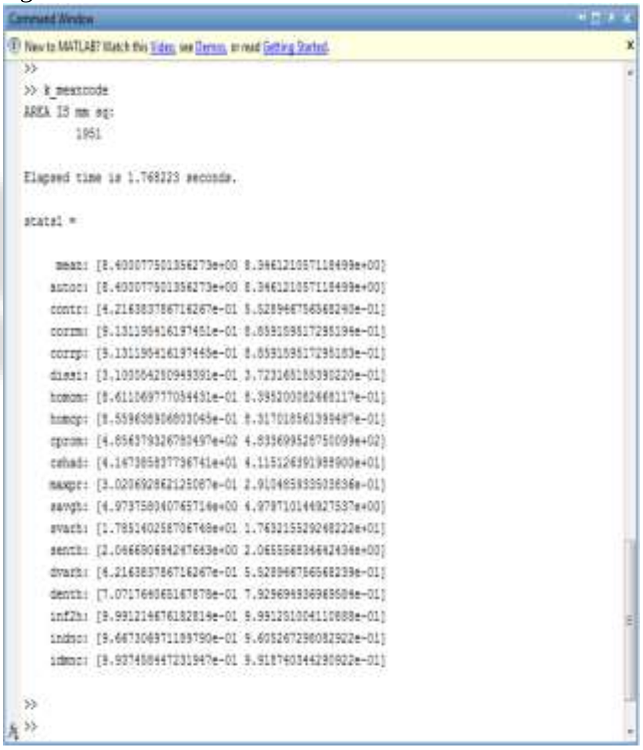


Fig 10: Feature Selection of Lateral Ventricle affected tumor

E. SVM Classification

Figure 11 shows an output of SVM classification with training samples. Here, five tumor & five non tumor images are taken as trained samples. SVM classifier makes classification of the training samples. Here, 1 indicates tumor image and 2 indicates non tumor image.

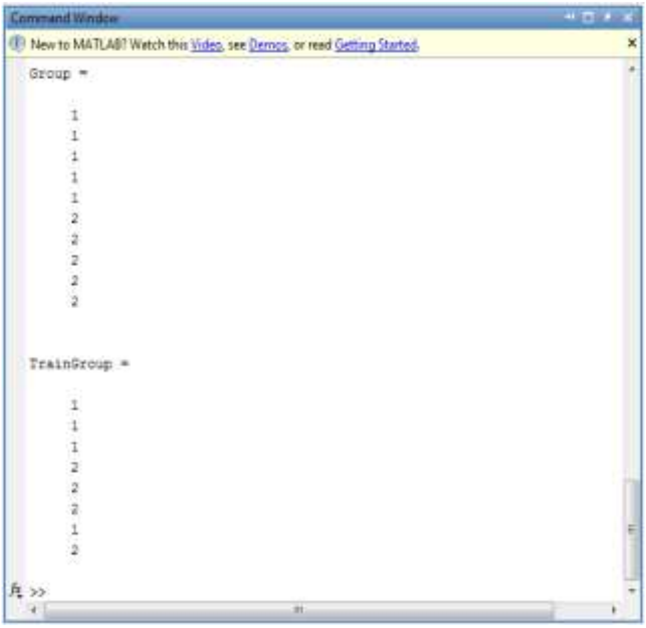


Fig 11: SVM classification with training sample

V. CONCLUSION

Finally the proposed work been implemented using MATLAB. The proposed work gives more accurate result than the existing system. Four different ventricles affected brain tumor MRI images are collected to this work. K-means segmentation algorithm works successfully and gives the actual tumor region to these different brain tumor images. This segmentation has been successfully implemented to 40 different brain tumor images. Feature extraction using GLCM has been successfully performed and classification between tumor and normal image also done from SVM classifier.

In future work, segmentation technique has to be improved for efficient and accurate result.

REFERENCES

[1] Rajesh C. Patil, Dr. A. S. Bhalchandra, "Brain Tumour Extraction from MRI Images Using MATLAB" International Journal of Electronics, Communication & Soft Computing Science and Engineering ISSN: 2277-9477, Volume 2, Issue 1

[2] P. Mohanaiah, P. Sathynarayana and L. Gurukumar, "Image Texture Feature Extraction Using GLCM Approach", International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013

[3] Monica Subashini.M , Sarat Kumar Sahoo , "Brain MR Image Segmentation for Tumour Detection using Artificial Neural Networks" ISSN : 0975-4024 Vol 5 No 2 Apr-May 2013.

[4] B. Sathya and R. Manavalan Monica Subashini.M, "Image Segmentation by Clustering Methods: Performance Analysis", International Journal of Compute Applications (0975 – 8887), Volume 29– No.11, September 2011

[5] Durgesh K. Srivatsa, Lekha Bhambhu, "Data Classification Using Support Vector Machine", Journal of Theoretical and Applied Information Technology, 2005-2009.

[6] H.Costin,"A fuzzy rules-Based segmentation method for medical images Analysis" INT J COMPUT COMMUN, ISSN 1841-98368(2):196-205, April, 2013.

[7] G. Evelin Sujji, Y.V.S. Lakshmi, G. Wiselin Jiji," MRI Brain Image Segmentation based on Thresholding": International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-3 Number-1 Issue-8 March-2013.

[8] Shweta Jain, "Brain Cancer Classification Using GLCM Based Feature Extraction in Artificial Neural Network", International Journal of Computer Science & Engineering Technology (IJCSET).

[9] Nitish Zulpe and Vrushsen Pawar, **"**GLCM Textural Features for Brain TumorClassification", IJCSI International Journal of Computer Science Issues, Vol 9, Issue 3, No3, May

[10] SivaSankari.S, Sindhu.M, Sangeetha.R, ShenbagaRajan, "Feature Extraction of Brain Tumor Using MRI", International Journal of Innovative Research in Science, Engineering and Technology, 2014.

[11] Sonali Wadgure , Prof. Mrs. Pooja Thakre, "Detection of Brain Tumor from MRI of Brain Using Fuzzy C-mean (FCM), International Journal of Science, Engineering and Technology Research, Volume 3, Issue 8, August 2014.

[12] Soumi Ghosh, Soumi Ghosh, "Comparative Analysis of K- Means and Fuzzy C-Means Algorithms", International Journal of Advanced Computer Science and Applications,Vol. 4, No.4, 2013.