BRAIN TUMOUR SEGMENTATION USING K-MEANS CLUSTERING

Mohammed Hameedullah Sharief[1], S.Yallamandaiah[2]

[1]*Digital Image Processing, Communication and Signal Processing, VR Siddhartha Engineering College, Andhra Pradesh, India*
[2]*Digital Image Processing, Communication and Signal Processing, VR Siddhartha Engineering College, Andhra Pradesh, India*

Abstract

Medical image processing plays an important role for finding the various types of diseases in real world. Tumour is one of the disease like a cancer can leads to death and it can be detected with the help of image processing. Magnetic resonance imaging MRI image is taken as input image it is resized to 500X500 i.e. five hundread rows and five hundread coloumns and skull stripping is done by morphological operations by converting image to binary based upon ostus thresholding and this image is cleaned using erosion and gaps are filled if any through filling operation. This binary image is overlaid on original image and this image is skull erased image. Then feature extraction is done by calculating the entropy feature by sliding small size window and then feature vector is constructed and given as input to the k-means clustering, it cluster tumour in one region and non tumour in another region. K-means is an example of exclusive clustering in which data belongs to only one cluster. Here two clusters are taken and centroids are obtained from two clusters. Distance is calculated from each pixel value to centroids. After this validation is done with the help of confusion matrix by giving ground truth through region polygon tool which can be used for creating mask from original image then this image is converted to vector along segmented image and both are compared in confusion matrix. This confusion matrix will create a table of different parameters through which accuracy, error rate, specificity and precision are calculated

Keywords: Feature Extraction, K-Means Clustering, Region Polygon, Confusion Matrix.

--***--

1 .INTRODUCTION

Tumour is one an abnormal growth of tissue in human body. There are many number of tumours some are Malignant and some are benign. Malignant types of tumours are cancerous they may lead to death where as benign tumours are not always cancerous but sometimes they are harmful if they are not treated. These tumours can be detected by Magnetic resonance imaging (MRI). It scans whole body is scanned to give a computer recording image technology. There are many other types of scanning technologies but MRI has advantage compared to others and also the effect of radiation is less. The particularly frequency used by MRI ranges from 40-130 MHz. This frequency is in normal range and this range is not at all harmful. MRI is consists of plane axial, coronal, and sagittal. Axial is said to be top view or front view. Coronal is back and sagittal is side view.

The tumour area will be with higher intensity value compared to other tissues in the brain which have very low intensity. Radiologists manually segment the regions of the tumour from MRI films.. The MRI consumes high amount of energy needs huge power. Most of the body of human being is made up of molecules and these consist of atoms like oxygen and hydrogen. At the centre of hydrogen atom a little particle is present known as proton and it is very sensitive to magnetic fields. When a person places him or himself in the scanner the protons in the body line up in a direction in alignment with the magnet. Small amount of radio waves are sent to certain areas of the body, hitting the protons out of order. When waves are turned off, then

protons will realign back to its original position. This will send the radio signals which can be collected by the receivers. Through these signals, information can be known about the exact location of the protons present in the body. This helps to identify many types of tissue present in the body, because the protons in different tissue types are rearranged at distinct speeds and generate different radio signals. From the millions of protons in the body, signals are combined together. Through this a detailed structure of interior image is created. Medical image processing helps in detecting the tumour by segmenting the MR images. Since many years tumours are detected by different types of techniques one of the technique such as thresholding. Others are morphological operations, supervised, unsupervised classification. Support vector machine is supervised classifications it need manual labels however unsupervised doesn't need any human interaction it can automatically classify the regions. Clustering is one of the segmentation techniques and there are many types of clustering such as k-means, self organizing map, hierarchical clustering, fuzzy c-means clustering etc. Are explained briefly in the next section

2. LITERATURE SURVEY

In medical image processing, segmentation plays a very important role. Manual segmentation can be done by the experts but it takes time and also may consist of errors. Now a day, many kinds of automated segmentation techniques are used. Some are discussed below. Ostus thresholding is one of the segmentation techniques. In this technique image is

converted to binary where it becomes logical. Binary image is obtained based on gray threshold value. The pixels which are greater than the threshold are grouped into one class and remain pixels grouping into another class. The disadvantage of this technique is that it doesn't work on multichannel images. Region growing is another method in which region based segmentation can be done. In this technique selection of seed points is done manually by user choice. It is also said to be iteration based method such as clustering algorithms. The next step is, the pixels in the region of a seeds are inspected thoroughly in order to determine their nature or condition and then added to the region in a manner conforming to a state of being homogeneous standard. This kind of process will be continued until all the pixels belong to the same region. In final step the object illustration is done by growing regions of pixels. The disadvantage of this method is, it needs manual selection of seed points by user. Some of the clustering methods belong are K-means, Fuzzy c means, self organizing mapping, Gaussian mixture model and Hierarchical clustering etc. All these clustering techniques divide the regions into known no of classes. There are other types of methods in which clustering is done based on feature extraction where feature vector is given as input to clustering technique and features are textural, GLCM features and combined all features to form feature vector. All the above techniques are unsupervised classification algorithms, there is no need of training, and they automatically learn the data. The draw back to the clustering techniques is choosing the no of clusters. We can also classify into exclusive and overlapping clustering. In exclusive type of classification, one data belongs to only one cluster only and it never belongs to another cluster. The example of this type of classification is K-means clustering. In overlapping type of classification, one data may belong to two or more clusters. The example of this type of classification is fuzzy c means clustering. In all these techniques morphological processing are used for thinning, cleaning and thickening

3. METHODOLOGY

The tumour segmentation is classified into four blocks skull stripping, feature extraction, classification and validation. Skull stripping is considered to be part of pre-processing stage and validation is post-processing stage. In the first step an MRI image is taken as input and pre-processing is done. The input image is collected from the ABC imaging scanning centre, Vijayawada and also some other data from different websites. The image is in axial view and pathogenic i.e. some type of disease. This image is resized to 500X500. The upper part of the head is skull which is erased in image discussed below.
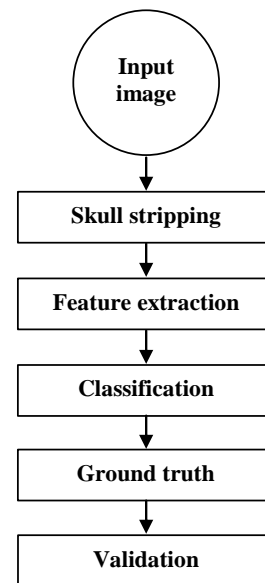

Fig-1: Block diagram

3.1 Skull stripping

First input image is converted to gray scale image and this conversion is conventional for filtering and pre-processing. The skull outline which is the outer part of the head and also boundary is removed in pre-processing stage. It is done by applying many kinds of morphological operations. Gray image is converted to binary image by giving the threshold value 0.1. The binary image is eroded through structuring element. After erosion the image consist of some gaps. To fill gaps, filling region operation is performed on the image now the gaps are filled and the image contains two regions black and white. Where black is empty region and white is brain. Now image is eroded again two times by taking the structuring element such as disk with radius 14 then image is over laid on the input image to skull stripped image.

3.2 Feature extraction

Block processing is performed on the image by sliding a filter. Hence this calculate feature for each block of image. The block can be chosen either 3x3 or 5x5 hence for the larger images increase block size but small windows gives good results [2]. The below matrix example shows the 3x3 block processing
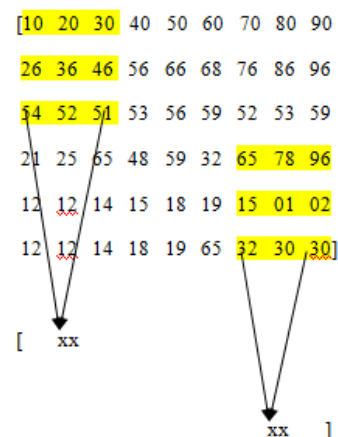
In this paper we use wavelet entropy which is texture feature [4]. This feature is calculated for all blocks. Then feature vector is constructed by combining all the features but here there is only one feature so only one feature will be in feature vector. There are four types of wavelet entropy such as Shannon, norm log energy, sure, threshold. In all these norm entropy is choose as it gives the best result. The power experimented on the image and given accurate results. The power 5 is also said to be non central moment 5th order. This is helpful for analyzing transient features like non stationary images. Wavelet entropy is defined as

$$W(E) = \sum_{i=1}^{m} \sum_{j=1}^{n} |I(m, n)|^5 \qquad (1)$$

Where I is image, i is number of m rows and j is number of n columns and W(E) is wavelet entropy norm .

3.3 Classification

Classifications are of two types supervised and unsupervised. Supervised need labels where as unsupervised is automatic classification. It can perform without any knowledge of labelling. K-means clustering [3] is a technique which is used very often. It is easy to implement and can be applicable even on large data set. It is used in various topics like marketing segmentation, computer vision, Geostatistics, Astronomy and Agriculture and it is also used as a pre-processing step for many other algorithms. During the cluster analysis the k-means used to partition the data set into no of clusters. But the problem is, hard to choose the no of k clusters. The feature vector is taken as input and then we need to identify clusters. Taking no of clusters depends upon the need of user. Here one cluster is tumour and other is non tumour. The feature vector gives the data calculating the wavelet entropy and the data present in this vector is given as input to k-means clustering then it groups the tumour region as one cluster and non tumour as other. The below algorithm is proposed method of k-means clustering

Algorithm
i)   Read image
ii) Initialize no of clusters c=2
iii) Now find all unique values and sort it then select maximum value
iv) Now multiply maximum value with no of clusters divide it to the no of clusters by adding one

$$c_t = \frac{k_c \times max}{k_c + 1}$$

Where c_t is centroid and we get two centroids
v)   Now calculate the difference from every data point to centroid

$$d = \sum_{i=1}^{m} \sum_{j=1}^{n} I(i, j) - c_t \qquad (2)$$

vi) Find the minimum or maximum value in the distance and then fill the data with this value

$$c_i = \sum_{i=1}^{m} \sum_{j=1}^{n} I(i, j) = minimum(d)$$

or (3)

$$c_i = \sum_{i=1}^{m} \sum_{j=1}^{n} I(i, j) = maximum(d)$$

where c_i is clustered image

3.4 Ground Truth

This is necessary for comparing to the obtained result to compute the accuracy. This ground truth can be obtained by trace the target region from the image. For example cutting tool from image. In matlab we can get ground truth from image by. It is used to cut the particular part from image in polygons. Selecting the tumour region using region of polygon tool after this selection mask is created. In this only the selected part is visible and other region will be dark as shown in figure-4. For evaluating confusion matrix is applied on ground truth and tumour segmented image and accuracy is calculated

3.5 Confusion Matrix

It is also called matching matrix, plays a very important role in describing the model. It is a table consists of rows and columns through parameters are analyzed. This gives detailed Information about the analysis than making insignificant proportion of correct guesses. It gives a measure of how does the classification performed. The below table is output of confusion matrix

Table-1 validation of tumour segmentation

| Total no of pixel values 250000	Predicted Tumour no	Predicted Tumour yes	
Actual Tumour no	TN:248894	FP:251	249145
Actual Tumour yes	FN:75	TP:780	855
	248969	1031	

3.6 Discussion

From the above table validation starts. True positive (TP) cases are those in which we predicted tumour yes and actually tumour exists. True negatives (TN) are those cases in which we predicted no tumour, and they are not having the tumour. False positive (FP) are those cases in which we predicted tumour is present, yes but tumour don't exists and it is one type of error. False negatives (FN) are those cases in which we predicted no there is no tumour, but tumour exists and this is another type of error.

Accuracy tells about how much is our classification is right? (TP+TN)/total= (780+248894)/250000 =0.9987 and Misclassification Rate or error rate tells about how much

error is present in classification? (FP+FN)/total=(251+75)/250000= 0.0013.Specificity is that case in which When we say actually no, how much does it predict no tumour? TN/actual no = 248894/250000 = 0.9990. Precision is when it predicts yes tumour is present and how much is it right? TP/predicted yes=780/1031=0.7565. Prevalence is how much tumour present yes condition actually occur in our sample? Actual yes/total = 855/250000 = 0.0034. The below section shows the results where figure one is input image and morphological operation are performed on this image to erase the outline of the brain image shown in figure-2. For figure-2 feature extraction is performed and wavelet entropy feature is computed for all the blocks by sliding 3x3 filter. In the case of larger images 5x5 window size is selected. After wavelet entropy feature computation the image is converted into vector. Finally the tumour segmented image is obtained by performing k-means clustering the clustered image shown in figure-3. For ground truth figure-1 is taken and mask is created after selecting the region. Region selection and mask image shown in figure-4 and 5. The figure from 6-14 are other sample images which shows input skull stripped and tumour segmented image.
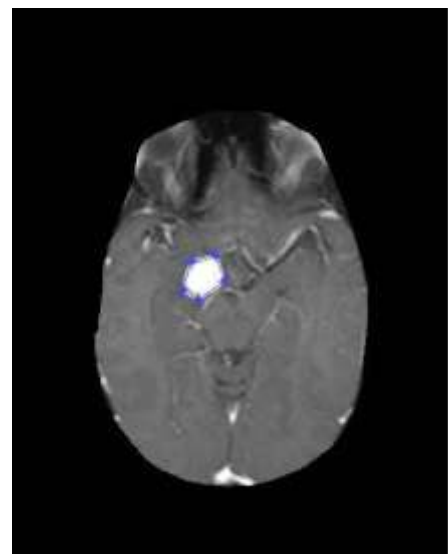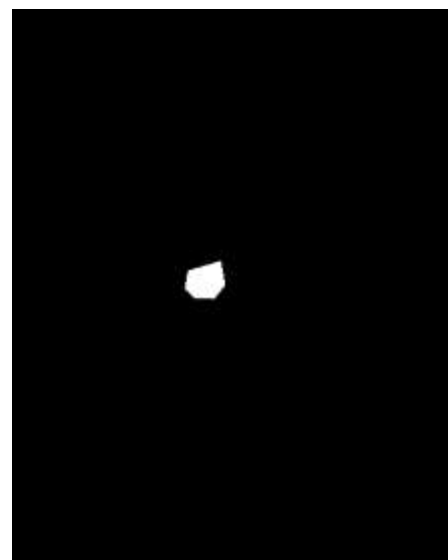
4. RESULTS
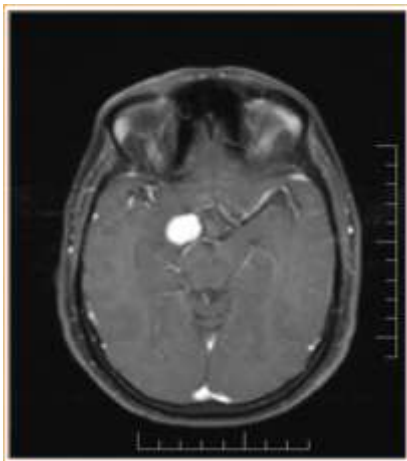
Fig-1: MRI input image
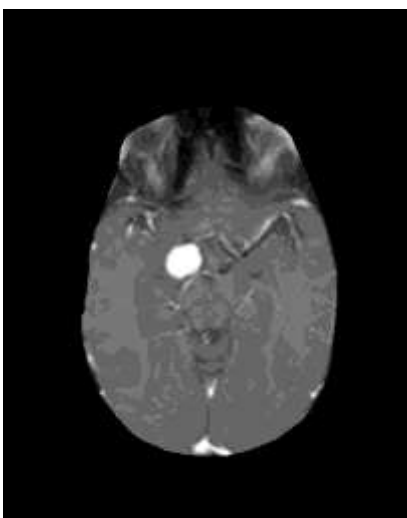
Fig-2: skull stripped after morphological operation
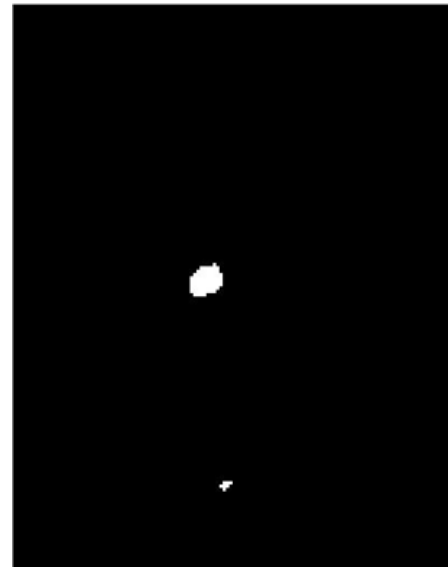
Fig-3: Tumour segmented image of k-means clustering

Fig-4:selecting region for mask image
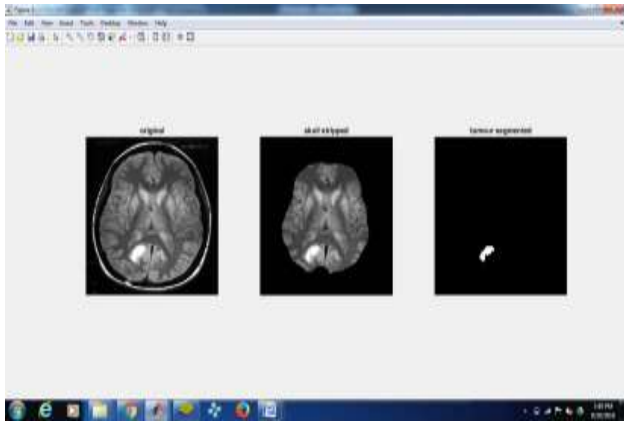

Fig-5: Masked image after cutting selected region

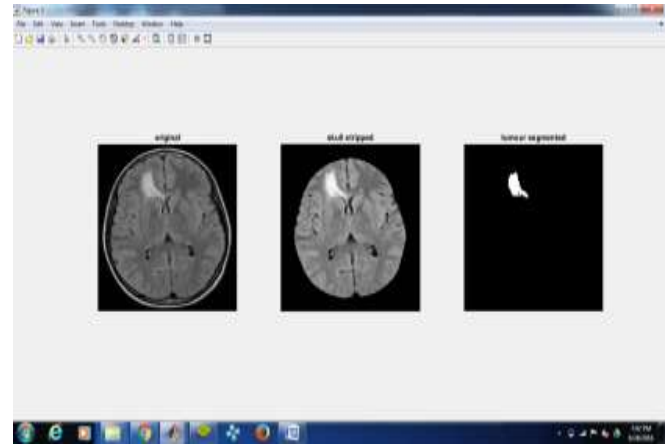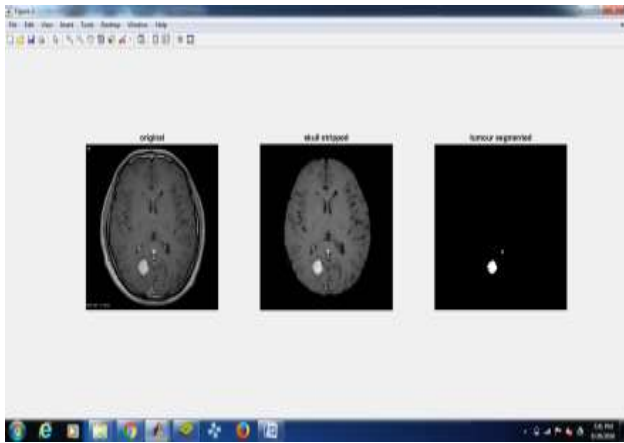**Fig-6**: Input image, Skull stripped, Segmented image
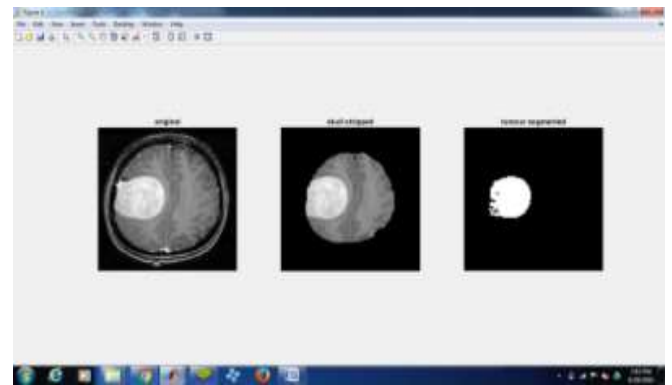

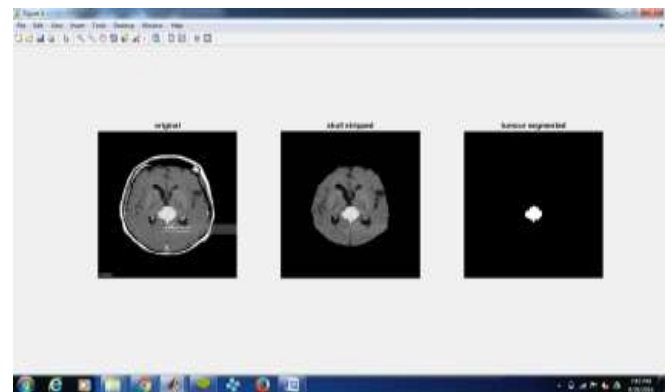**Fig-7:** Input image, Skull stripped, Segmented image

Fig-8: Input image, Skull stripped, Segmented image

Fig-9: Input image, Skull stripped, Segmented image

Fig-10: Input image, Skull stripped, Segmented image

Fig-11: Input image, Skull stripped, Segmented image
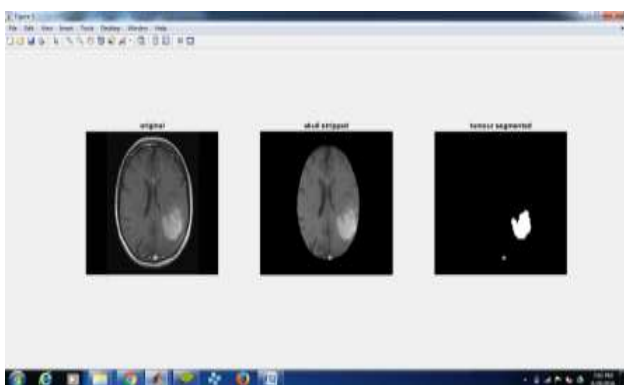

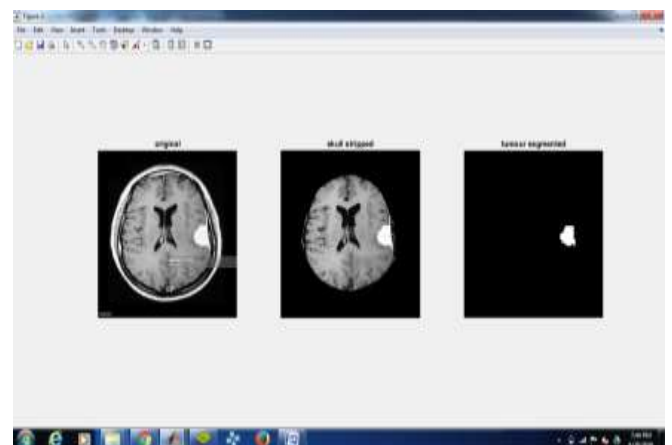**Fig-12**: Input image, Skull stripped, Segmented image

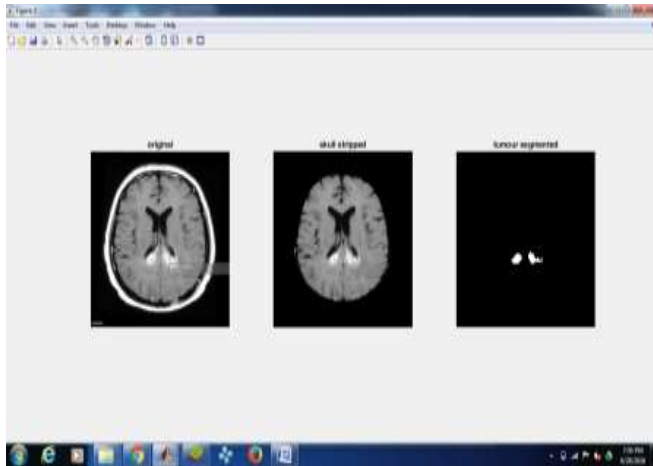**Fig-13**: Input image, Skull stripped, Segmented image

Fig-14: Input image, Skull stripped, Segmented image

5. CONCLUSION

Brain tumour segmentation is performed on different sample images. Skull stripping is done by using morphological operations and thresholding image. Then this image is filtered or eroded if any noise is present and feature is selected i.e. wavelet entropy type norm with power 5 and feature extraction is performed by sliding 3x3 or 5x5 filter on skull stripped image and feature vector is constructed and this is dataset given as input to the k-means clustering by taking the no of clusters as two the result is classified into two regions tumour and non tumour. For the purpose of validation ground truth is obtained by creating a mask from the input image using region of interest polygon tool then generated mask is change to type double and confusion matrix is done by taking, asked image and clustered image. This confusion matrix generate comparison table through which accuracy is calculated.

REFERENCES

[1]. Jay patel and kaushal dosh A Study of Segmentation Methods for Segementation of Tumour in Brain MRI "Advance in Electronic and Electric Engineering." ISSN 2231-1297, Volume 4, Number 3 (2014), pp. 279-284

[2]. Ayse Demirhan, Mustafa Toru, and Inan Guler "Segmentation of Tumor and Edema Along With Healthy Tissues of Brain Using Wavelets and Neural Networks" IEEE journal of biomedical and health informatics, vol.19,no. 4, july 2015

[3]. Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu "an efficient k-means alogorithm: analysis and implementation" IEEE transactions on pattern analysis and machine intelligence vol. 24, no. 7, july 2002

[4]. Coifman, R.R. M.V. Wickerhauser (1992), "Entropy-based Algorithms for best basis selection," IEEE Trans. on Inf. Theory, vol. 38, 2, pp. 713–718.

BIOGRAPHY

Mohammed Hameedullah Sharief was born in Guntur in the year 1993. He received B.Tech degree in electronics and communication engineering from the Kallam Haranadha Reddy institute of technology, Guntur, india, in 2014 and present pursuing M.Tech in communication engineering and signal processing from the VR.siddhartha engineering college, Kanuru, India, respectively.

Mr.S.Yallamandaiah was born in guntur in the year 1986. He received B.Tech degree in electronics and communication engineering from Sri Sarathi Institute of technology, Nuzvid, India, in 2008 and completed M.Tech degree in communication engineering and signal processing from the VR.Siddhartha engineering college, kanuru, India and currently working as Assistant Professor in .Siddhartha engineering college respectively