BJMHR
British Journal of Medical and Health Research
Journal home page: www.bjmhr.com

FCM and KNN Based Automatic Brain Tumor Detection

R. Ragupathy[1], B. Abirami[1*]
1.Department of computer science and Engineering, Annamalai University, Chidambaram

ABSTRACT

A brain tumor is formed when abnormal cells get accumulated within the brain. These cells multiply in an uncontrolled manner and damage the brain tissues. Magnetic Resonance Image scans are commonly used to diagnose brain tumors. However, segmenting and detecting the brain tumor manually is a tedious task for the radiologists. Hence, there is a need for automatic systems which yield accurate results. A fully automatic method is introduced to detect brain tumors. It consists of five stages Image Acquisition, Preprocessing, Segmentation, using Fuzzy C-means technique; Harris Corner Detection based feature extraction and classification using K-NN. Performance metrics such as accuracy, precision, sensitivity and specificity are used to evaluate the performance.

Keywords: Brain Tumor, Magnetic Resonance Imaging, Fuzzy C Means, Harris Corner Detector, K Nearest Neighbor.

*Corresponding Author Email: cse_ragu@yahoo.com
Received 20 April 2017, Accepted 01 May 2017

Please cite this article as: Ragupathy R *et al*., FCM and KNN Based Automatic Brain Tumor Detection. British Journal of Medical and Health Research 2017.

INTRODUCTION

A brain tumor is tissue mass that is formed by abnormal cells. Tumors do not die like normal cells. Brain tumors can be both noncancerous and malignant. Cancerous brain tumors grow faster than noncancerous ones, and they attack surrounding tissue. Noncancerous tumors are easier to remove than malignant ones and are not typically considered to be life-threatening. Cancerous brain tumors put pressure on the under, around and inside the skull and cause inflammation in the brain. Brain tumor symptoms vary, but headaches are a common symptom. Other brain tumor symptoms include seizures, balance problems, personality changes, vision and speech issues, and difficulty concentrating. The system deals with pre-processing and post-processing. Pre-processing stage consists of contrast adjustment, histogram, and re-sampling[1]. In post-processing it consists of segmentation, feature extraction and classification. The system deals with segmentation using FCM clustering technique, Harris Corner Detection using feature extraction, classification using K-NN.

FCM is used to segment the infected region of brain in MRI images. The goal of clustering is to reduce the amount of data by categorizing or grouping similar data items together[2]. Such grouping is pervasive in the way human's process information, and one of the motivations for using clustering algorithms is to provide automated tools to help in constructing categories or taxonomies. The methods may also be used to minimize the effects of human factors in the process[3]. There are different types of clustering techniques such as hard clustering, soft clustering, spectral clustering, dirichlet clustering, streaming k-means clustering. In hard clustering a given data point is in n-dimensional space only belongs to one cluster. This is also known as exclusive clustering. The k-means clustering mechanism is an example of hard clustering. In soft clustering a given data point can belong to more than one cluster in soft clustering. This is also known as overlapping clustering. The fuzzy k-means algorithm[4] is a good example of soft clustering. The spectral clustering algorithm is helpful in hard, nonconvex clustering problems. It clusters points using the eigenvectors of matrices derived from data. Dirichlet clustering fits a model over a dataset and tunes parameters to adjust the model's parameters to correctly fit the data. This approach is suitable to address the hierarchical-clustering problem. Streaming k-means is to read data points sequentially, storing very few data points in memory. Then, after the first step, a better representative set of weighted data points is produced for further processing. The final K number of clusters is produced in the ball k-means step. During the second step, potential outliers are eliminated.

Corner detection is an approach used within computer vision systems to extract certain kinds of features and infer the contents of an image. Corner detection is frequently used in motion detection, image registration, video tracking, image mosaicing, panorama stitching, 3D modeling and object recognition. In feature extraction stage, Harris corner points are extracted from segmented image. Some of the corner detectors are Susan corner detector, Moravec corner detector, Forstner corner detector, Fuzzy logic for corner detector. In Susan corner detector a circular mask is applied around every pixel, and the grayscale values of all the pixels within the mask are compared to that of the centre pixel. Calculate the number of pixels within the circular mask which have similar brightness to the centre pixel. In Moravec corner detector the corner strength is defined as the smallest SSD between the patch and its neighbors (horizontal, vertical and on the two diagonals). If this number is locally maximal, then a feature of interest is present. One of the main problems with this operator is that it is not isotopic, if an edge is present that is not in the direction of the neighbors, then the smallest SSD will be large and the edge will be incorrectly chosen as an interest point. In Forstner corner detector a point closest to all the tangent lines of the corner in a given window and is a least-square solution. This algorithm relies on the fact that for an ideal corner, tangent lines cross at a single point. Fuzzy logic corner detector for data from natural images is always imprecise and noisy due to inherent uncertainties that may arise from the imaging process such as defocusing, wide variations of illuminations. Detection of corners becomes difficult under such imperfect situations. So fuzzy systems are well known for efficiently handling of impreciseness and incompleteness due to imperfection of data.

Classification is used to assign corresponding levels with respect to groups with homogeneous characteristics, with the aim of discriminating multiple objects from each other within the image. The level is called class. Classification will be executed on the base of spectral or spectrally. Different types of classifier are decision tree classifier, ruled based classifier, SVM classifier, Naïve bayes classifier, Neural networks classifier. In this paper we used K-NN classifier. Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. The term rule-based classification can be used to refer to any classification scheme that make use of IF-THEN rules for class prediction. SVM performs classification by finding the hyper plane that maximizes the margin between the two classes. The vectors (cases) that define the hyper plane are the support vectors. Naïve Bayes is a simple technique for constructing classifiers models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. A neural network consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer.

PROPOSED SYSTEM

In proposed system the segmentation algorithm is simple which is followed by labeling. Fuzzy c-means clustering technique which uses the fuzzy logic to establish a degree of belonging to each pixel is used. This technique is also known as soft clustering and gives accurate results. Finally, the tumor is extracted and mean feature extracted for classification. Classification is done by K-Nearest Neighbor (KNN). Classification can be useful to assign weight to the contributions of the neighbors, so that the nearest neighbors contribute more to the average than the more distant ones.

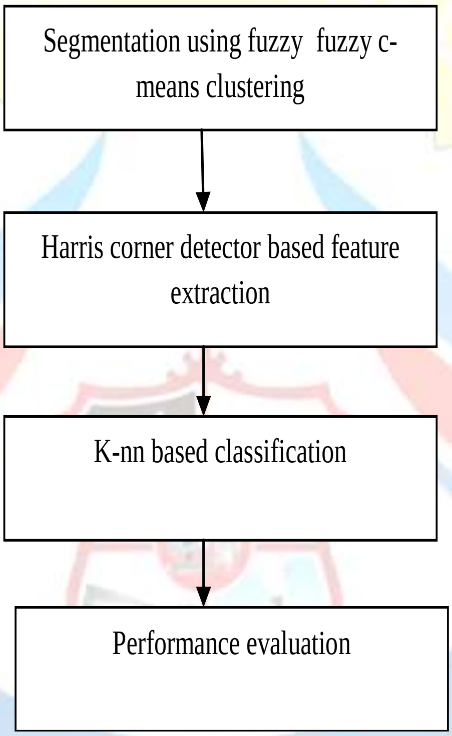

Figure 1: Block Diagram

SEGMENTATION USING FUZZY C-MEANS CLUSTERING ALGORITHM

Figure 1 shows the block diagram for proposed methodology. FCM is used for segment the infected region of brain in MRI images[5]. The goal of a clustering analysis is to divide a given set of data or objects into a cluster, which represents subsets or a group. The partition should have two properties. They are homogeneity and heterogeneity. Homogeneity inside clusters; the data, which belongs to one cluster, should be as similar as possible. Heterogeneity between the clusters; the data, which belongs to different clusters, should be as different as possible.

The algorithm of fuzzy c-means is a classification algorithm based on fuzzy optimization of quadratic criterion of classification where each class is represented by its center of gravity. The algorithm requires knowing the number of classes in advance and generates classes through an iterative process minimizing an objective function. Thus, it provides a fuzzy partition of the image by giving each pixel a degree of belonging to a given region[6].

The fuzzy c-means clustering algorithm is as follows:

Let x = $\{x_1, x_2, x_3 \ldots . x_n\}$ be the set of data points and v = $\{v_1, v_2, v_3 \ldots . v_c\}$ be the set of centers.

1. Randomly select 'c' cluster centers.

2. Calculate the fuzzy membership 'μ_{ij}' using

$$\mu_{ij} = 1 / \sum_{k=1}^{c} \left(d_{ij} / d_{ik} \right)^{(2/m-1)} \qquad (1)$$

3. Compute the fuzzy centers 'v_j' using

$$v_j = (\sum_{i=1}^{n} (\mu_{ij})^m x_i) / (\sum_{i=1}^{n} (u_{ij})^m) \forall_j = 1,2,\ldots c \qquad (2)$$

4. Repeat step 2 and step 3 until the minimum 'J' value is achieved

Where,

'n' is the number of data points

'v_j' represents the j^{th} cluster center

'm' is the fuzziness index m £ [1, ∞]

'c' represents the number of cluster center

'μ_{ij}' represents the membership of i^{th} data to j^{th} cluster center

'd_{ij}' represents the Euclidean distance between i^{th} data and j^{th} cluster center.

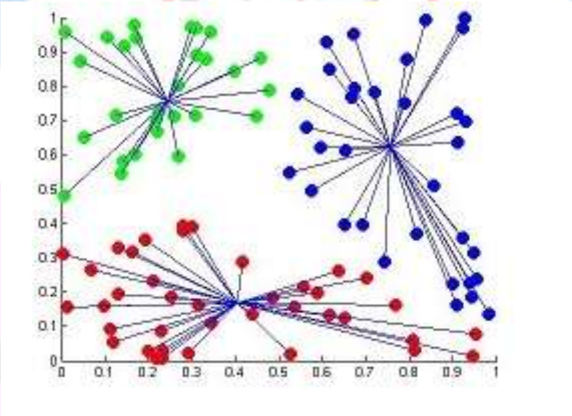

Figure 2: Fuzzy c-means

Figure 3: Clustered image

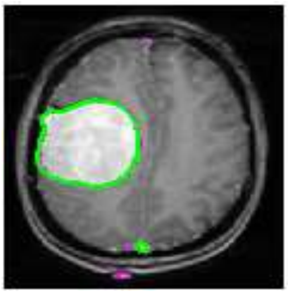

Figure 4: Final Segmented image

FCM works by assigning membership to each data point corresponding to each cluster centre on the basis of distance between the cluster and the data point figure 3 shows the clustered image. More the data is near to the cluster centre more is its membership towards the particular center and figure 4 shows the final segmented image.

HARRIS CORNER DETECTOR BASED FEATURE EXTRACTION

In feature extraction stage, Harris corner points are extracted from segmented image. These feature having high dimensionality so reduce these features, we are calculating the mean feature[7].

The Harris Corner Detector is a mathematical operator that finds features in an image. It is simple to compute, and is fast enough to work on computers. It is popular because it has rotation, scale and illumination variation independent. A corner can be defined as the intersection of two edges[8]. A corner can also be defined as points for which there are two dominant and different edge directions in local neighborhood of the point and figure 5 shows the Harris corner detection.

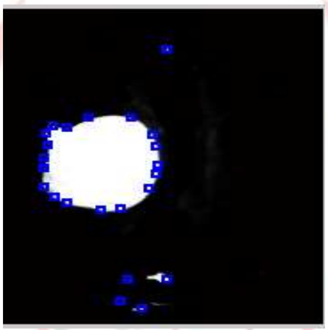

Figure 5: Harris Corner Detection

K-NN BASED CLASSIFICATION

The classification process is done over the segmented images. The main novelty here is the adoption of K-Nearest Neighbor (KNN). K-NN classifier is applied over the segmented images and the classification is done[9].

K-Nearest Neighbor is a non-parametric method used for classification. The input consists of the k closest training examples in the feature space. K-NN is a type of instance based learning.

The K-NN classification is among the simplest of all machines learning algorithm[10]. In K-NN classification an object is classified by a majority vote of its neighbors with the object being assigned to the class most common among its k- nearest neighbor and k is appositive integer. If k=1, then the object is simply assigned to the class of that single nearest neighbor.

The algorithm of K-NN classification is as follows:

Step 1: The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples.

Step 2: In the classification phase, K is a user defined constant, and an unlabelled vector is classified by assigning the label which is most frequent among the K- training samples nearest to that query point.

Step 3: A commonly used distance metric for continuous variables is Euclidean distance.

Step 4: K-NN has also been employed with correlation co-efficient such as Pearson and spearman.

Step 5: The class accuracy of KNN can be improved significantly if the distance metric is learned with specialized algorithm such as Neighborhood component analysis.

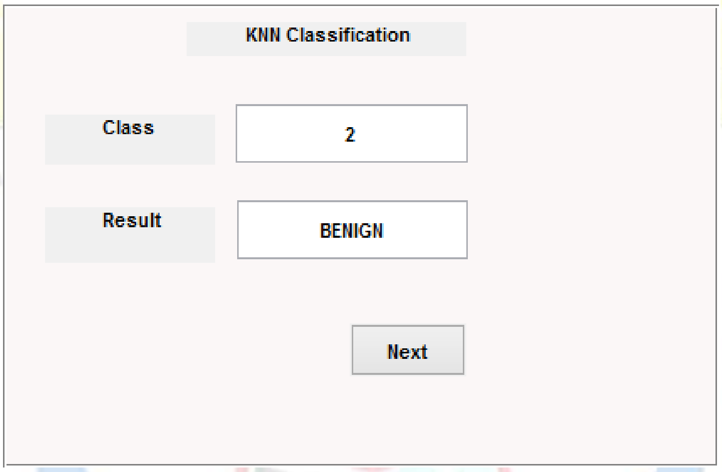

Figure 6: K-NN Classification

The classification process is done over the segmented image. The K-NN classifier is applied over the segmented images is shown in figure 6.

RESULTS AND DISCUSSION

The Harvard whole brain atlas a rarity, a comprehensive and well designed site that fulfills its objectives. The site contains a huge compilation of modern cross sectional imaging, including CT, MRI and SPECT in health and disease. There are sections like normal anatomy and test the top 100 brain structures, or review the imaging appearances of a number of more common neurological conditions. There are no advertisements and no excessive animation or unnecessary frills.

The dataset consists of magnetic resonance images, which were downloaded from the website of Hardvard Medical School(URL: http://med.harvard.edu/AANLIB/). Table 1 shows the normal images are 9 in numbers. The benign images are 17 in numbers and malignant are in 14 in numbers.

Table 1 Number of brain images collected from database

S.no	Types of images	Number of images
1.	Normal brain images	9
2.	Benign brain images	17
3.	Malignant brain images	14

PERFORMANCE EVALUATION

- True Positives (TP) – Brain tumor images are correctly recognized. This means the people who had brain tumor are correctly identified.
- False Positives (FP) – Non-brain tumor images are incorrectly recognized. It indicates the people did not had brain tumor are incorrectly identified as they had brain tumor. Healthy people incorrectly identified as sick.
- True Negatives (TN) – Non-brain tumor images are correctly recognized as they did not had brain tumor. Healthy people correctly identified as healthy.
- False Negatives (FN) – Brain tumor images are incorrectly recognized. It represent the people had brain tumor are incorrectly identified as they did not had brain tumor. Sick people incorrectly identified as healthy.

$$\text{Accuracy} = \frac{True\ Positives + True\ Negatives}{TP+TN+FP+FN} \quad (3)$$

$$\text{Precision} = \frac{True\ Positives}{True\ positives + False\ Positives +} \quad (4)$$

$$\text{Sensitivity} = \frac{True\ Positives}{True\ Positives + False\ Negatives} * 100\% \quad (5)$$

$$\text{Specificity} = \frac{True\ Negatives}{True\ Negatives + False\ Positives} * 100\% \quad (6)$$

A confusion matrix is used to describe the performance of a classification model or classifier on a set of data for which the true values are known. A confusion matrix also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning (in unsupervised learning it is usually called matching matrix).

Table 2: Confusion Matrix

	N	B	M
N	9	0	0
B	0	17	0
M	0	2	12

Table 3: Performance Evaluation

Type	Accuracy	Precision	Sensitivity	Specificity
Normal	0.95	1	1	1
Benign	0.95	0.94	1	0.91
Malignant	0.95	0.85	0.85	1
Over all %	**95%**	**93%**	**95%**	**97%**

The above table 3 shows the overall performance evaluation for accuracy, precision, sensitivity and specificity. In performance evaluation the value of accuracy is 96 percentages, precision is 96.49 percentages, sensitivity is 96.23 percentages and specificity is 97 percentages.

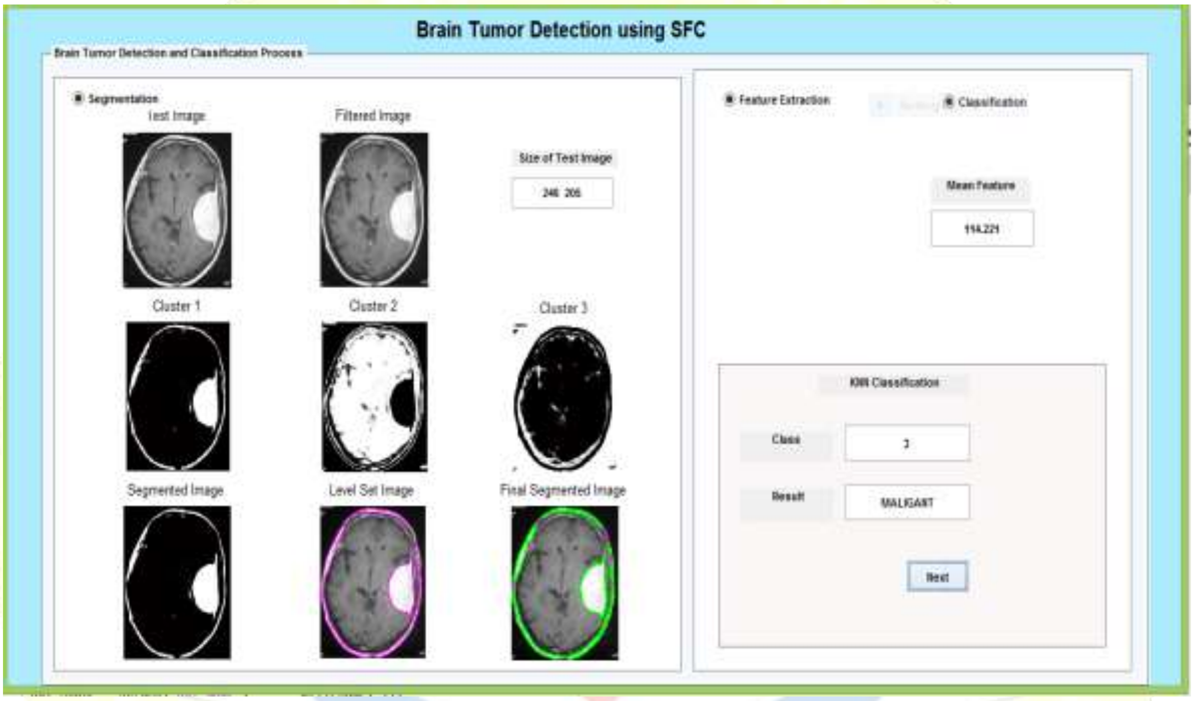

Figure 7: Experimental Results

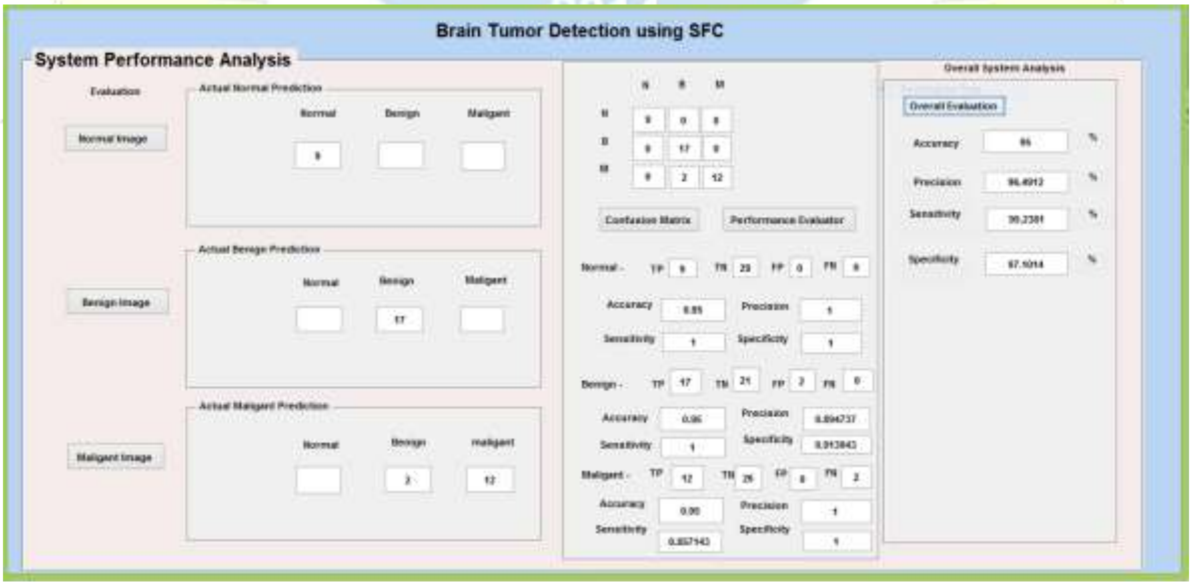

Figure 8: Performance Evaluation

Figure 7 shows the experimental results, the segmentation process is done by FCM technique, feature extraction is done by Harris corner detection to calculate the mean feature and classification is done by K-NN classifier and figure 8 shows performance evaluation.

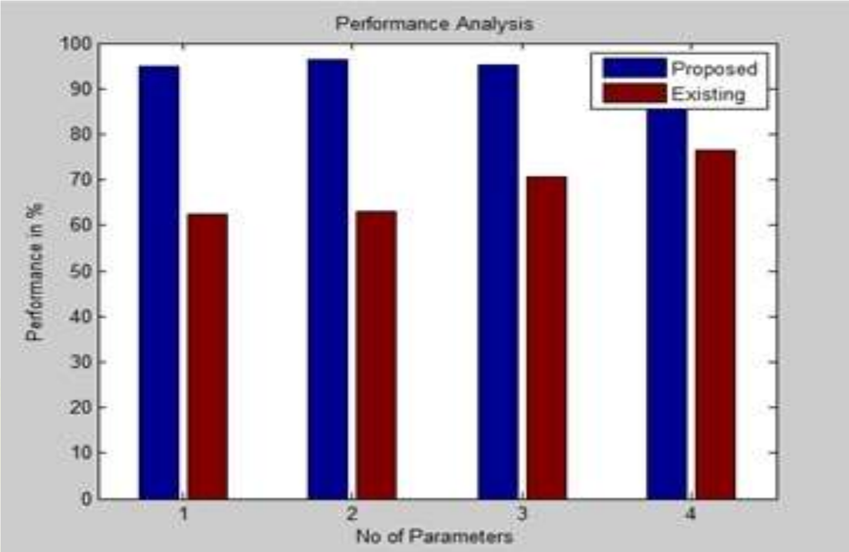

Figure 9 : Graphical representation

Figure 9 shows the graphical representation of performance evaluation for proposed and existing system. Thus Accuracy, Precision, Sensitivity, Specificity are compared in the above given chart.

CONCLUSION

An automatic system for segmentation and detection of brain tumor in magnetic resonance images is proposed. The image is acquired from the database and preprocessing operations like wiener filter are applied. The preprocessed output image is then segmented using the Fuzzy C-Means algorithm which is fast and yields good results, Harris Corner Detection based feature extraction to calculate the mean feature and classification using K-NN. Finally classification was done over segmented images.

REFERENCES

1. R Ragupathy, B Abirami, and N Aswini, "Automatic Brain Tumor Detection," International Journal For Science and Advance Research in Technology, vol. 3, no. 3, pp. 920 - 925, Mar 2017.
2. Ritu Khatri Sangeeta Sehrawat, "Brain tumor Detection Using FCM and BPNN," International Journal of Basic and Applied Biology, vol. 2, no. 1, pp. 83 - 88, Oct 2014.
3. Jianxin Wang, Fangxiang Wu, Tianming Liu, Yi Pan Jin Liu Min Li, "A Survey of MRI-Based Tumor Segmentation Methods," in Tsinghua Science and Technology, vol. 19, 2014, pp. 578 - 595.

4. Tripti Singhal Heena Hooda Om Prakash Verma, "Brain Tumor Segmentation A Performance Analysis using K-Means, Fuzzy C-means and Region Growing Algorithm," in IEEE International Conference on Advanced Communication Control and Computing Technologies, 2014, pp. 1621 - 1626.
5. S.R.Suralkar Jinal A.Shah, "Brain Tumor Detection from MRI Images using Fuzzy C-Means Segmentation," International Journal of Advanced Research in Computer and Communication Engineering, vol. 5, no. 6, pp. 178 - 183, Jun 2016.
6. Abdolreza Farzan Vida Harati Rasoul Khayati, "Fully automated tumor segmentation based on improved fuzzy connectedness algorithm in brain images," in Computers in Biology and Medicine, vol. 41, 2011, pp. 483 - 492.
7. Sandip R.Panchal Trupti P.Patel, "Corner Detection Techniques: An Introductory Survey," International Journal of Engineering Development and Research, vol. 2, no. 4, pp. 3680 - 3686, 2014.
8. Zhenxing Luo, "Survey of Corner Detection Techniques in Image Processing," International Journal of Recent Technology and Engineering, vol. 2, no. 2, pp. 184 - 185, May 2013.
9. Hugo Larochelle Mohammad Havaei Pierre-Mare Jodoin, "Efficient interactive brain tumor segmentation as within-brain KNN classification," in International Conference on Pattern Recognition, 2014, pp. 556 - 561.
10. Dr.K.Thanushkodi N.Suguna, "An Improved K-Nearest Neighbor Classification Using Genetic Algorithm," International Journal of Computer Science Issues, vol. 7, no. 4, pp. 18 - 21, Jul 2010.