Indian Journal of Science and Technology, Vol 9(36), DOI: 10.17485/ijst/2016/v9i36/102154, September 2016 ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Segmentation, Feature Extraction and Classification of Astrocytoma in MR Images

Sheifali Gupta*, Prabhpreet Walia, Chaitanya Singla, Shivani Dhankar, Tanvi Mishra, Ayush Khandelwal and Mohit Bhardwaj

Chitkara University, Rajpura - 140401, Punjab, India; Sheifali.gupta@chitkara.edu.in, zinniawalia101@gmail.com, chaitanya.singla@chitkara.edu.in, shivani_dhankar@hotmail.com, tanvi333.in@gmail.com, ayushkhandelwal86@gmail.com, bhardwajmohit81@yahoo.co.in

Abstract

The purpose of this research paper is to segment tumorous tissue from brain magnetic resonance images using image processing. For this the astrocytoma, a type of brain tumor is considered in which cells grow abnormally in the brain. Astrocytoma is classified as low grade or high grade using k-nn classifier and then analyzed its performance on the basis of three parameters i.e., accuracy, severity and specificity. The Magnetic Resonance Images (MRI) of astrocytoma has been taken from BRATS database so as to explore different algorithms for segmentation of brain tumour. The enhancement of MRI images is done using Contrast Limited Adaptive Histogram Equalization (CLAHE). After that basic global thresholding method is used for the automatic segmentation of tumorous tissue. In feature extraction step, the shape based features and texture based features are extracted. On the basis of features extracted from the segmented image, K-nn classifier is used to classify the images in two grades i.e., low grade or high grade. The performance of the system is evaluated by parameters accuracy, severity and specificity. The accuracy is coming out to be 93%.

Keywords: Astrocytoma, CLAHE, Global Thresholding, Grading, GLCM, k-NN Classifier, Tumor

1. Introduction

Astrocytoma (brain tumor) is a life threatening disease in which cells grow abnormally in the brain. According to the WHO (world Health Organization) tumors of central nervous system classify into four grades[1]. Astrocytomas are also divided into four grades. The grade 1 and grade 2 astrocytoma can be considered as low grade astrocytoma and grade 3 and grade 4 astrocytoma referred to high grade astrocytoma. There are various image modalities which are used to study the internal structure of the human brain such as Computerized Tomography (CT) and MRI. There are many advantages of MRI over CT scan as MRI does not contain any harmful radiation. MRI is suitable for brain studies because of its excellent contrast of soft issues, non-invasive characteristic and a high spatial resolution. Due to above all the advantages MR images are used for astrocytoma detection and grading in our proposed methodology. The manual segmentation of the tumor region from the MR image takes very long time and a lot of challenges associated with it. Therefore, automated segmentation method is required. So, the approach used for non-invasive astrocytoma segmentation and grade identification includes automatic thresholding, feature extraction and classification using k-NNclassifier. The MR image dataset is taken from the BRATS database. The performance of the above approach is evaluated on the basis of quality metrics that are accuracy, severity and specificity.

The structure of this paper is classified as follows: Section 2 details the methodology used in the proposed approach. The performance metrics is explained in Section 3. Section 4 discusses the conclusion and the future work.

2. Methodology

The proposed methodology consists of a lot of steps starting from collecting Astrocytoma MR images to classification as low grade or high grade. The important

* Author for correspondence

steps used in the methodology are shown in Figure 1. Here the preprocessed CLAHE enhanced Astrocytoma image is used for further process. As concluded in our previous paper based on performance analysis of image enhancement techniques for Astrocytoma detection in MR images that CLAHE enhancement technique outperforms other two techniques in terms of Peak signal to noise ratio and mean square error. Basic Global thresholding technique is used for the automatic segmentation of the tumorous tissue from the enhancement image. Various features are extracted from the tumor extracted image which are explained in detail in the below sections.

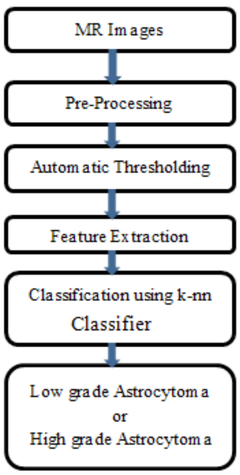


Figure 1. Methodology.

2.1 Automatic Thresholding

Thresholding is a technique used for segmentation of image into different regions based on a threshold value[2]. This threshold value can be compared with intensity value based on which image is classified into different regions. Basic global thresholding technique is used to automatically calculate the threshold value on the basis of which tumor region is extracted from input image[3]. Basic global thresholding is a technique which uses the histogram of the input image for thresholding. The steps to implement this technique or to calculate the threshold value are as follows:-

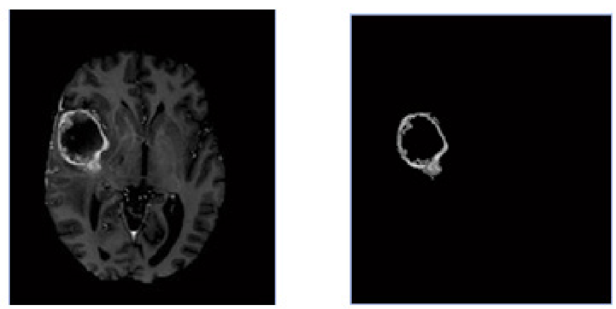
- The first step is set an estimated value of T which mainly equals to the average of the grey values in the image. The value of T for the input image (Figure 2(a)) is 0.18913.
- The next step is to partition the pixels into two groups on the value of T i.e., pixels having grey value greater than T are in one group and those having value less than that of T are in another group.
- Calculate the average of grey level values in group 1 and say it may equals to u1. Similarly calculate the average of grey level values in group 2 and say it may equals to u2. The value of u1 and u2 in our taken image are 0.0598 and 0.464 respectively.
- The next step is to calculate the new value of T which is equals to the average of u1 and u2.

$$T = \frac{u1 + u2}{2}$$

The value of T computes out to be 0.262.

- Repeat the above three steps until the difference between the values of T is less than the predefined value.

The original image taken from BRATS database is shown in Figure 2(a) and its tumor segmented image is shown in Figure 2(b).



(a) (b)

Figure 2. (a) Original image. (b) Tumor segmented image.

2.2 Feature Extraction

Different texture based and shape based features are extracted from the tumor segmented image which helps in further classification of that image[4,5]. Texture features are extracted using Gray level co-occurrence matrix. Shape features are extracted using region props.

2.2.1 Texture based Features

A total of 22 texture based features are extracted using GLCM from the image. Some of these features are given below and can be expressed using the equations.

- Energy: It measures the degree of pixel pair repetitions. The uniformity of image is also measured using this feature.

$$E = \sqrt{\sum_{i,j} p^2(i,j)}$$

- Entropy: It measures the disorder in an image. In homogeneous scenes there is low entropy whereas for homogeneous scenes the entropy is high.

$$En = -\sum_{i,j} p(i,j) \times \log(p(i,j))$$

- Dissimilarity: Dissimilarity is a measure that defines the variation of gray level pairs in an image. It is slightly different from the contrast feature as the weights in the contrast measure increases exponentially whereas the weights in the dissimilarity measure increases linearly.

$$Dissimilarity = \sum_{i,j} |i-j|\, p(i,j)$$

Contrast: It can be explained as the difference between the maximum and minimum pixel intensity in an image.

$$Contrast = \sum_{i,j} |i-j|^2\, p(i,j)$$

- Correlation: The correlation feature measures the linear dependency of grey levels on those of neighboring pixels. It has a value between -1 and 1. -1 implies maximally uncorrelated and 1 means maximally correlated.

$$Correlation = \sum_{i,j} p(i,j)\left[\frac{(i-\mu_x)(j-\mu_y)}{\sqrt{(\sigma_x^2)(\sigma_y^2)}}\right]$$

$$\mu_x = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} i.\,p(i,j) \qquad \mu_y = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} j.\,p(i,j)$$

$$\sigma_x = \sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-\mu)^2\, p(i,j)} \qquad \sigma_y = \sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(j-\mu)^2\, p(i,j)}$$

μx, μy and σx, σy are the mean and standard deviations. N is the number of grey levels.

- Homogeneity: It calculates the local homogeneity of an image. It is also known as inverse difference moment. The value of the weight is the inverse of the value of weight of contrast feature.

$$Homogeneity = \sum \frac{1}{1-(i-j)^2} p(i,j)$$

- Cluster Shade: It calculates the skewness of the matrix. The image is said to be asymmetric if the value of the cluster shade is high.

$$Cluster\ Shade = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(i+j-\mu_x-\mu_y\right)^3 p(i,j)$$

- Cluster Prominence: It also measures the symmetry of the image. If the value of the cluster Prominence is high the image is less symmetric.

$$Cluster\ Prominence = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(i+j-\mu_x-\mu_y\right)^4 p(i,j)$$

- Maximum Probability: This feature maintains the record of the center pixel of the window the largest pij value found within the window.
- Sum Average:

$$Sum\ Average = \sum_{i=0}^{2(N-1)} i \times p_{x+y}(i)$$

$$p_{x+y}(k) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j),\ k = i+j = \{0,1,2 \ldots\ldots 2(N-1)\}$$

- Sum Variance:

$$Sum\ Variance = \sum_{i=0}^{2(N-1)} (i - Sum\ Average)^2 p_{x+y}(i)$$

- Sum Entropy:

$$Sum\ Entropy = -\sum_{i=0}^{2(N-1)} p_{x+y}(i) \log p_{x+y}(i)$$

- Difference Variance:

$$Difference\ Variance = \sum_{i=0}^{N-1} (i-t)^2 p_{x-y}(i)$$

$$p_{x-y}(k) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j),\ k = |i-j| = \{0,1,2 \ldots\ldots (N-1)\}$$

$$t = \sum_{i=0}^{N-1} i.\,p_{x-y}(i)$$

- Difference Entropy:

$$Difference\ Entropy = -\sum_{i=0}^{N-1} p_{x-y}(i) \log p_{x-y}(i)$$

- Information Measure of Correlation Feature 1:

$$M = \frac{(HXY - HXY1)}{\max(HX, HY)}$$

- Information Measure of Correlation Feature 2:

$$M2 = \sqrt[2]{\left(1 - e^{[-2(HXY2-HXY)]}\right)}$$

$$p_x(i) = \sum_{j=0}^{N-1} p(i,j) \quad p_y(i) = \sum_{i=0}^{N-1} p(i,j)$$

$$HX = -\sum_{i=0}^{N-1} p_x(i)\log p_x(i) \quad HY = -\sum_{i=0}^{N-1} p_y(i)\log p_y(i)$$

$$HXY = -\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)\log p(i,j)$$

$$HXY1 = -\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)\log p_x(i)p_y(j)$$

$$HXY2 = -\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p_x(i)p_y(j)\log p_x(i)p_y(j)$$

HX and HY are entropies. The texture based features extracted for original image is shown in Table 1.

2.2.2 Shape based Features

A total of 16 shape based features are extracted from the image using region props and some of them are explained below.

- Area: The total number of pixels in the specified region is called area of that region.
- Eccentricity: It is the ratio of distance between the foci of the ellipse and its major axis length.
- Euler Number: It specifies the number of objects in the region minus the number of holes in those objects.
- Extent: It specifies the ratio of pixels in the region to the pixels in total bounding box.
- Convex Area: The total number of pixels present in the convex Image.
- Orientation: It specifies the angle between the x-axis and the major axis of the ellipse.
- Solidity: It specifies the proportion of the pixels in the convex hull that are also in the region. It is calculated as the ratio of the area and the convex area.

The shape based features extracted for original image is shown in Table 2.

2.3 Classification

The k-Nearest Neighbor (kNN) is used for classification

Table 1. Various texture features along with their values w.r.t the input image

Sl. No.	Feature Name	Value	Sl. No.	Feature Name	Value
1.	Autocorrelation	1.211	12.	Maximum Probability	0.9841
2.	Contrast	0.034	13.	Sum of Square: Variance	1.1968
3.	Correlation: matlab	0.878	14.	Sum Average	2.0841
4.	Correlation	0.878	15.	Sum Variance	4.4117
5.	Cluster Prominence	32.13	16.	Sum Entropy	0.1160
6.	Cluster Shade	3.92	17.	Difference Variance	0.0348
7.	Dissimilarity	0.017	18.	Difference Entropy	0.0694
8.	Energy	0.968	19.	Information measure of correlation1	-0.6087
9.	Entropy	0.126	20.	Information measure of correlation2	0.3241
10.	Homogeneity: matlab	0.993	21.	Inverse difference normalized	0.9982
11.	Homogeneity	0.992	22.	Inverse difference moment normalized	0.9995

Table 2. Various shape based features along with their values w.r.t the input image

Sl. No.	Feature Name	Value	Sl. No.	Feature Name	Value
1.	Area	737	9.	Convex Area	1839
2.	Centroid	114.5726	10.	Filled Area	1703
3.	Bounding Box	84.50	11.	Euler Number	-7
4.	Major Axis Length	70.592	12.	Extrema	142.500
5.	Minor Axis Length	46.2697	13.	EquivDiameter	30.6329
6.	Eccentricity	0.7552	14.	Solidity	0.4008
7.	Orientation	-67.8020	15.	Extent	0.2955
8.	Convex Hull	142.500	16.	Perimeter	180.4092

of astrocytoma as low grade astrocytoma or high grade astrocytoma. The classifier used was a k-Nearest Neighbors (kNN) classifier that assigns labels to pixels based on the most frequent label among the k closest training points under a distance metric applied to the features (referred to as 'lazy' learning, since no explicit model is learned). The kNN algorithm is a simple and effective method for multi-class classification that is able to model non-linear distributions[6–9]. All the features extracted passed to the kNN classifier. The kNN classifier uses default Euclidean distance metric to calculate the distance between the features extracted from the training images and the features extracted from the testing images. KNN classifier calculates the distance between features (extracted from test image) with all features (extracted from each training image) and then out of these n training images k are chosen based on the least distance difference then on the majority voting basis a class is chosen and assigned to test image. By default the value of k is 1. Figure 3 shows the bock diagram for KNN classifier. The distance between the features (extracted from test image) against the features extracted from every image in training set (T_1 … T_n) and then arrange the features in the ascending order after that on the basis of majority vote class is chosen as in block diagram for k = 3 majority class is chosen.

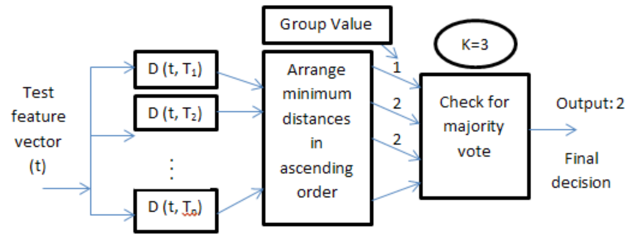


Figure 3. Block diagram for KNN classifier.

The input image taken from the database is classified using K-NN classifier and the tumor segmented from the image is classified as low grade astrocytoma tumor as shown in Figure 4 and as high grade tumor as shown in Figure 5.

3. Performance Metrics

The performance of the above methodology can be analyzed on the basis of performance metrics that are accuracy, specificity and severity. The four basic parameters that are used to calculate the above metrics are False Negatives (FN), True Positives (TP), False Positives (FP) and True Negatives (TN).

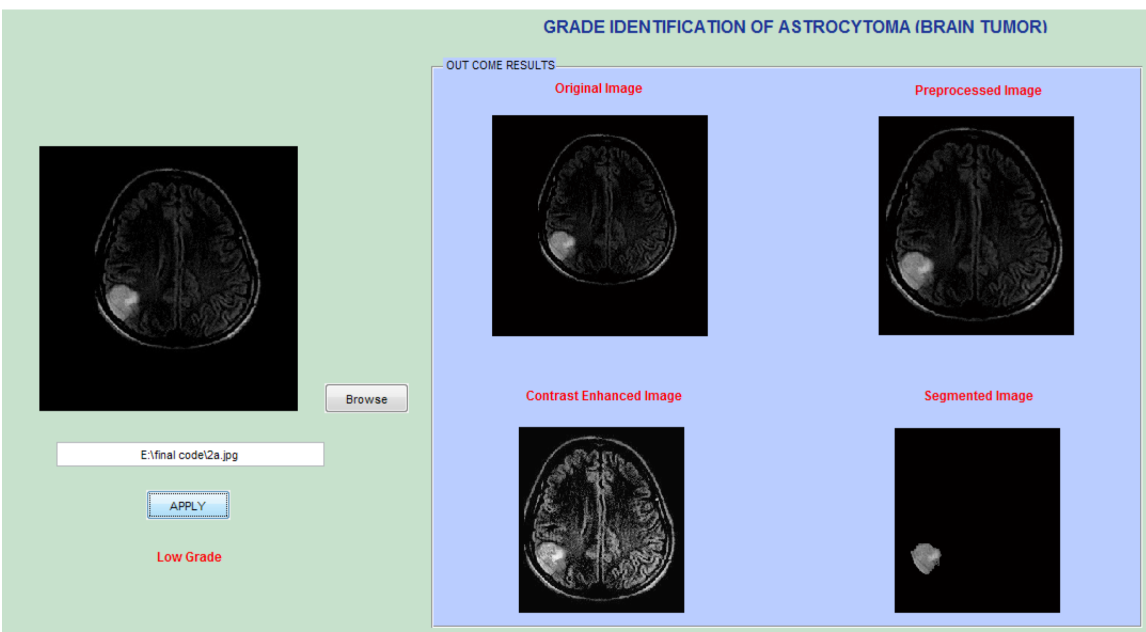


Figure 4. GUI showing the classification of input image (low grade).

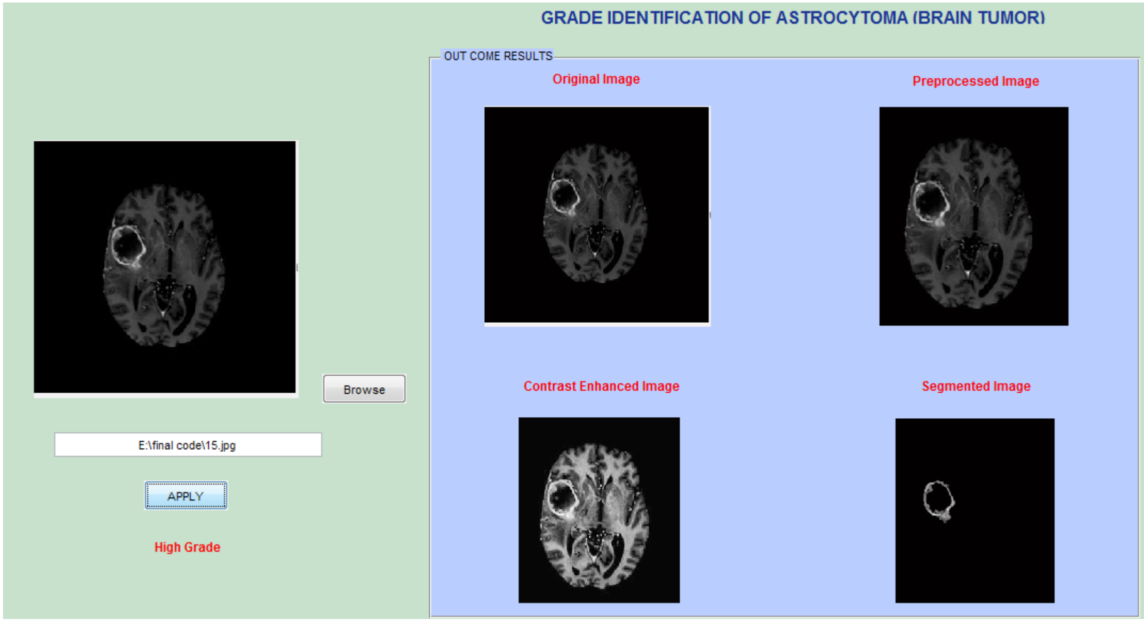


Figure 5. GUI showing the classification of input image (high grade).

True positives (TP): It implies that the classifier correctly identifies the infected brain with astrocytoma tumor.

True Negatives (TN): It implies that the classifier correctly identifies the MRI images which do not have tumor.

False positives (FP): It implies that the classifier incorrectly identifies the MRI images. The brain images which do not infected from astrocytoma classifier predicts those as infected from astrocytoma.

False Negative (FN): It implies that the classifier incorrectly identifies the MRI images. The brain images which are infected from astrocytoma classifier predict those as that are not infected from astrocytoma.

Total Positives (P): The sum of false negatives and true positives is referred as total positives.

Total Negatives (N): The sum of false positives and true negatives is referred as total negatives.

Accuracy: The accuracy of a classifier on the particular set of images is the proportion of the total number of images that are precisely identified by the classifier.

$$\text{Accuracy} = (TP+TN) / (P+N)$$

Sensitivity: The ratio of the true positives over the total positives is referred as sensitivity. It calculates the fraction of total number of positives that are accurately classified.

$$\text{Sensitivity} = TP/P$$

Specificity: The ratio of the true negatives over the total negatives is referred as specificity. It calculates the fraction of total number of negatives that are accurately identified.

$$\text{Specificity} = TN/N$$

4. Conclusion and Future Work

The result shows that 93% accuracy of the proposed approach is achieved. The astrocytoma images are taken from the BRATS database. In future in place of k-nn classifier nature inspired or bio inspired algorithm such as genetic algorithm is used for classification.

5. References

1. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW, Kleihues P. The 2007 WHO classification of tumours of the central nervous system. Acta neuropathologica. 2007 Aug; 114(2):97–109.
2. Natarajan P, Krishnan N, Kenkre NS, Nancy S, Singh BP.

Tumor detection using threshold operation in MRI brain images. IEEE International Conference on Computational Intelligence and Computing Research (ICCIC); 2012 Dec. p. 1–4.

3. Ehab FB, Esraa GM, Nadder H. An algorithm for detecting astrocytomas in MRI images. IEEE International Conference on Computer Engineering and Systems (ICCES); 2010 Dec. p. 368–73.
4. Kumar BS, Selvi RA. Feature extraction using image mining techniques to identify brain tumors. 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS); 2015 Mar. p. 1–6.
5. Patel O, Maravi YP, Sharma S. A comparative study of histogram equalization based image enhancement techniques for brightness preservation and contrast enhancement. arXiv preprint; 2013 Nov.
6. Gupta V, Sagale KS. Implementation of classification system for brain cancer using backpropagation network and MRI. 2012 Nirma University International Conference on Engineering (NUiCONE); 2012 Dec 6. p. 1–4.
7. Vaidyanathan M, Clarke LP, Velthuizen RP, Phuphanich S, Bensaid AM, Hall LO, Bezdek JC, Greenberg H, Trotti A, Silbiger M. Comparison of supervised MRI segmentation methods for tumor volume determination during therapy. Magnetic Resonance Imaging. 1995 Dec; 13(5):719–28.
8. Deekshatulu BL, Chandra P. Classification of heart disease using k-nearest neighbor and genetic algorithm. Procedia Technology. 2013 Dec; 10:85–94.
9. Zhang G, Lu Z, Ji G, Sun P, Yang J, Zhang Y. Automated classification of brain MR images by wavelet-energy and k-nearest neighbors algorithm. 2015 Seventh International Symposium on Parallel Architectures, Algorithms and Programming (PAAP); 2015 Dec. p. 87–91.