Detection and Classification of HGG and LGG Brain Tumor Using Machine Learning

F. P. Polly[1], S. K. Shil[1], M. A. Hossain[1], A. Ayman[1], and Y. M. Jang[2]
[1]School of Science, Engineering & Technology, East Delta University, Chittagong, Bangladesh.
[2]Department of Electronics Engineering, Kookmin University, Seoul, Korea.
Email: arif.h@eastdelta.edu.bd, yjang@kookmin.ac.kr

***Abstract*— Gliomas are brain tumors starting in the glial cells. Gliomas can be low grade (slow growing) or high grade (fast growing). Physicians use the grade of a brain tumor based on gliomas to decide which treatment a patient needs. The condition of the tumor is of vital importance for the treatment. In this paper, we propose a computerized system to differentiate between normal brain and abnormal brain with tumor in the MRI images and also further classify the abnormal brain tumors into HGG or LGG tumors. The proposed computerized system uses k-means as the segmentation technique for clustering whilst Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA) are the main parts of the feature extraction and feature reduction mechanisms, respectively. Support vector machine (SVM) is a major part of our proposed system as it classifies the abnormal brain tumors in the LGG and HGG after the extraction and reduction of the features.**

***Index Terms*—Glioma, HGG, LGG, Brain tumor, tumor detection, MRIs, K-means, DWT, PCA, SVM.**

I. Introduction

Glial cells are the support cells in the brain that help to keep neurons in place and functioning well. Gliomas form when these glial cells mutate and grow out of control. They are classified as astrocytoma, or, oligodendroglioma, or ependymoma. At times, the tumor can be a combination of all of these too.

Medical image segmentation for detection of brain tumor from the magnetic resonance images or from other medical imaging modalities is a very important process for deciding right therapy at the right time because the earlier the detection, the faster the treatment can be started. Studies based on its detection, classification and cure are at an onset at this instant.

In this fast phased world, a computerized system for brain tumor detection and classification is a priori to save time and proceed into the next series of medications according to the achieved result. MRI images are preferred in our computerized system since it can accurately comprehend different tissues [1].

This study proposes a computerized system to differentiate between normal and abnormal MRI images and again classify the abnormal MRI's into HGG or LGG glioma tumors. The tumor is segmented by k-means. The main contribution of this study lies in the design of the computerized system for classifying the abnormal MRI images into HGG or LGG glioma tumors with the selective relevant features for the required procedure.

The rest of the paper is organized as follows. The literature review is done in Section II. The theoretical design and the phases of the scheme are described in Section III. The experimental results of the computerized system are documented in Section IV. Finally, conclusions and future work are written in Section V.

II. Related Works

Image segmentation is the most imperative part of image processing. Numeral researchers have suggested diverse methodologies and algorithms for image segmentation. P. Katti and V. R. Marathe designed a detection and classification system for Brain cancer detection distinguishing different types of brain MRI into three classes such as Benign, Malignant and Normal [2]. The features are extracted using both Discrete Cosine Transform and Discrete Wavelet Transform independently. Segmentation is performed by utilizing k-means clustering. Automated brain tumor classification is implemented using Probabilistic Neural Network (PNN). Finally, accuracy analysis for Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) are made individually.

A. Batra and G. Kaushik propose a combinational algorithm of FCM clustering and SVM classifier for classification of the tumor in conjunction with BCFCM for bias field correction and HAAR wavelet transform for feature extraction [3]. The proposed method achieves a 98.2 percent of accuracy, 97.5 percent sensitivity, 100 percent specificity, 100 percent positive predictive value and 98.7 dice similarity coefficient.

An Intelligent-Model for Automatic Brain Tumor Diagnosis is presented by M. B. M. Amien et al. which classifies MRI images as normal, edema, cancer, or not classified [4]. Noise is removed and the contrast of the image is enhanced in the preprocessing stage by using multiple steps. Texture features are extracted and principal component analysis (PCA) is used to reduce its dimensions in the latter stage. Finally, Back-Propagation Neural Network (BPNN) based on Pearson correlation coefficient is utilized to classify brain images. The proposed model achieved an accuracy of 96.8%.

C. L. Devasena and M. Hemalatha presented an efficient Hybrid Abnormal Detection Algorithm (HADA) to detect the abnormalities in any part of the human body by using MRIs [5]. The proposed technique consisted of five stages: Noise

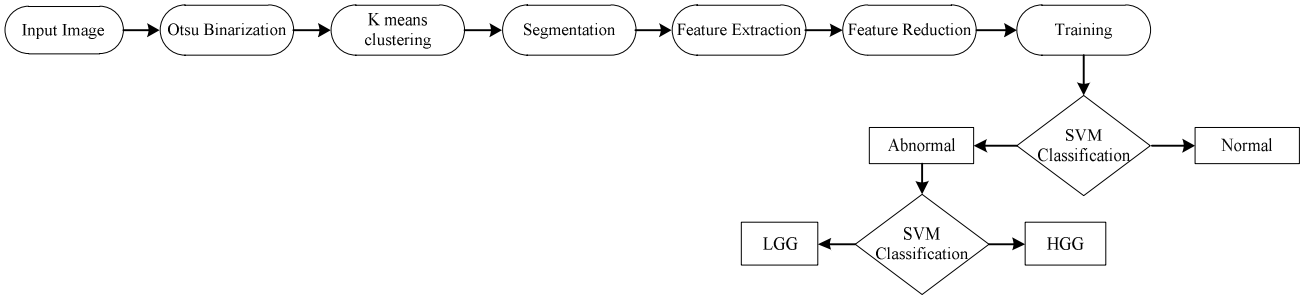


Fig. 1. Flowchart of the Proposed System.

reduction, smoothing, feature extraction, feature reduction and classification. The proposed algorithm was implemented and achieved a classification accuracy of 98.8%.

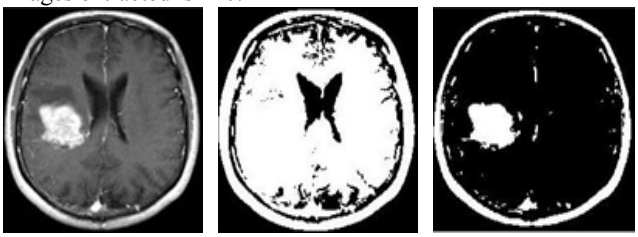
E. Dandil et al. designed a CAD system to detect tumor using T1 and T2 weighted MR images [6]. The designed system segments brain tumor region of MR image using spatial-Fuzzy C-means (FCM). FIF, SF, and GLCM techniques were used for feature extraction and PCA method was used for feature selection. Subsequently, support vector machine (SVM) is used to classify the tumors into benign and malign tumors. According to test results, the proposed CAD system recognizes brain tumors with 91.49% accuracy, 90.79% sensitivity and 94.74% specificity.

III. Proposed System

The proposed system is designed for the precise detection and classification of normal and abnormal brain MRI's and then the classification of the abnormal MRI's into HGG or LGG glioma tumor. The block diagram in Fig. 1 gives a proper idea of the methodology used in the proposed system. Brain MRI is read by the system, then otsu binarization is applied to convert the image into a binary image. After that k-means clustering is applied for segmentation. Later, DWT and PCA are applied. Finally, SVM is used for classification. In stage 1, the images are classified into normal or abnormal MRI's. In stage 2, the abnormal MRI images are classified into HGG or LGG glioma tumor MRI's as observed in Fig. 2 and Fig. 3.

A. Data Description

The computerized system is tested using a large data. They are previewed in 3D Slicer 4.6.2 Software. T2-weighted MRI images have been utilized in our system. The total number of images extracted is 440.



(a) (b) (c)

Fig. 2. Output Image of HGG glioma tumor : (a) Brain MRI Image, (b) Otsu Thresholded Image, (c) Segmented Tumor.

Among them, there are 100 normal images and 340 abnormal images. In the first stage, 100 normal images and 180 abnormal tumorous images are used. In the first stage, 130 images are utilized (70 normal and 60 abnormal images) for the first training stage and the remaining 150 images are used for the first testing stage purpose (30 normal and 120 abnormal images). In the second stage, 80 HGG and 80 LGG images are used. In the second stage, we have used 60 images (30 HGG glioma tumor images and 30 LGG glioma tumor images) for the second training stage and the remaining 100 images (50 HGG and 50 LGG) are used for the second testing stage purpose. The images have been collected from Brats17 Training Data [10] and Brats13 Training Data [9] in .nii form and from "Midas: healthy human brain database" [8] in .mha form.

B. Initialization

The proposed system is started with the input of the T2-weighted MRI images. But before beginning the actual procedure, the image must be converted into the suitable applicable form. Otsu binarization is applied to the image to convert the image into a binary image where 0 stands for black and 1 stands for white pixel values. The image is thresholded depending upon the minimum variance within the class and converted into the readable image form.

C. Segmentation

The desired region of interest is segmented from the background in the segmentation stage. Segmentation is necessary for classification because all the decisions depend upon the segmented tumor. In the proposed method, segmentation is applied using k-means clustering [7].

Euclidean distance is used to measure a point's proximity to the centroid. A new binding is done between the nearest new centroid and the same data set points. A loop is then generated. As a result of this loop, the locations with similar pixel values are separated out into two or more clusters.

D. Feature Extraction and Feature Reduction

(i) Feature Extraction is done by using discrete wavelet transform (DWT). It is used to find the wavelet coefficient from the MRI images. All the important features of the images are captured by a subset of DWT coefficients [2]. In tumorous images, feature extraction is used for collecting the feature

variables from the tumor portion. It is also used to remove Gaussian noise from the respective MRI images.

(ii) Feature Reduction is done using Principal component analysis (PCA) [4]. PCA is a mathematical formulation used for data compression (i.e. reducing data dimensions). It acknowledges standards in data (i.e. their features) to be identified and expressed by emphasizing their similarities and differences. PCA allows a large number of variables to be replaced with fewer variables (i.e. called the principal component) using a linear relationship between the variables. The number of principal components that can be identified for any dataset is equal to the number of the variables in the dataset. The principal components are of utmost benefit in the computerized system since the variables may or may not be capable of any intuitive human interpretation. In summary, the PCA formulation may be used as a digital image compression algorithm with a low level of loss. Such a reduction is advantageous for image compression, data representation, calculation reduction necessary in subsequent processing, etc. The useful feature formulae used in our computerized system are listed below.

(1) Contrast: Contrast is a measurement of the difference between maximal and minimal pixel intensities of an image.

$$C_{on} = \sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(x-y)^2 G(x,y) \tag{1}$$

(2) Correlation: Correlation measures the linear dependency of grey levels of neighboring pixels.

$$C = \frac{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(x,y)\,G(x,y) - \mu_X \mu_Y}{\sigma_X \sigma_Y} \tag{2}$$

(3) Energy: Energy is a measure of the localized change of the image. It measures the uniformity of an image.

$$\varepsilon = \sqrt{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} G^2(x,y)} \tag{3}$$

(4) Mean: The mean of an image measures the contribution of individual pixel intensity values for the entire image.

$$\mu_X = \left(\frac{1}{m\times n}\right)\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} G(x,y) \tag{4}$$

(5) Entropy: Entropy measures the loss of information or message in a transmitted signal and also measures the image information.

$$E = -\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} G(x,y)\log_2 G(x,y) \tag{5}$$

(6) Kurtosis: The shape of a random variable's probability distribution is described by the parameter called Kurtosis.

$$k = \left(\frac{1}{m\times n}\right)\frac{\sum (G(x,y)-\mu_X)^4}{\sigma_X^{\,4}} \tag{6}$$

(7) Skewness: Skewness is a measure of symmetry or the lack of symmetry.

$$\mathbb{S} = \left(\frac{1}{m\times n}\right)\frac{\sum (G(x,y)-\mu_X)^3}{\sigma_X^{\,3}} \tag{7}$$

(8) Inverse Difference Moment (IDM): Inverse Difference Moment measures homogeneity of an image.

$$I = \sum_{x=0}^{m-1}\sum_{y=0}^{n-1}\frac{1}{1+(x-y)^2} G(x,y) \tag{8}$$

(9) Coarseness: Coarseness is a measure of unevenness in the texture of an image. Rough textures have higher coarseness values, whereas; fine textures have smaller values.

$$c = \frac{1}{2^{m+n}}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} G(x,y) \tag{9}$$

(10) Directional Moment: Directional moment is a textural property calculated by considering the alignment of the image measured in terms of the angle.

$$DM = \sum_{x=0}^{m-1}\sum_{y=0}^{n-1} G(x,y)\,|x-y| \tag{10}$$

E. Classification

SVM is a binary classifier based on supervised learning. SVM classifies between two classes by creating a hyperplane in high-dimensional feature space. The data is already labelled, hence the name "supervised learning". It works by depending upon the training data to find the hyperplane that has the largest distance to the nearest training data point of any class [3]. The larger the margin the lower the generalization error of the classifier. The distance between the hyperplane and the closest data points is referred to as the margin. The optimal line that can separate the two classes is the line that has the largest margin. This is called the Maximal-Margin hyperplane. The margin is calculated as the perpendicular distance from the line to only the closest points. Only these points are appropriate in defining the line and in the construction of the classifier. These

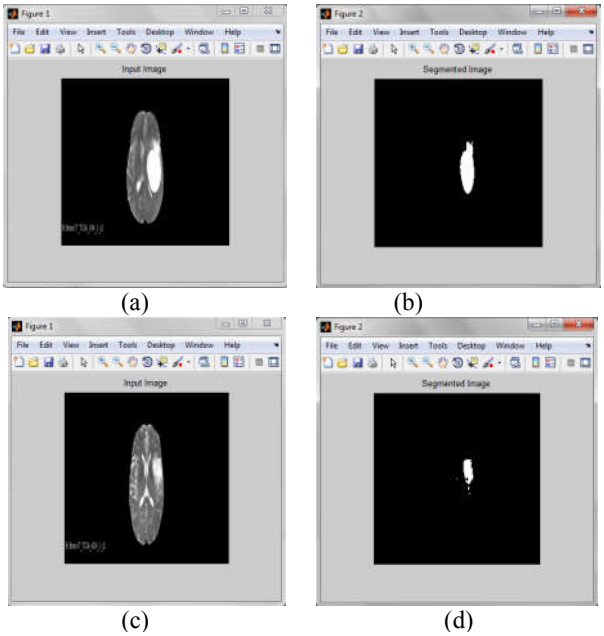


Fig. 3. (a) Input Image of an HGG tumor, (b) Segmented Image of an HGG tumor, (c) Input Image of an LGG tumor, (d) Segmented Image of an LGG tumor [9][10].

points are called the support vectors. They support or define the hyperplane.The hyperplane is learned from training data using an optimization procedure that maximizes the margin. The hyperplane can be represented for x points by (11).

$$w \cdot x - b = 0 \tag{11}$$

where w is weight vector and normal to hyperplane and b is bias or threshold. The values of w and b have to be such that the margin is maximized or the distance between the parallel hyperplanes is as far apart as possible while still separating the data points.

These hyperplanes can be described by (12) and (13).

$$w \cdot x - b = 1 \tag{12}$$

$$w \cdot x - b = -1 \tag{13}$$

The following constraint has to be fulfilled to prevent data points from falling into the margin. For each i th term,

$$w \cdot x_i - b \geq 1 \tag{14}$$

$$w \cdot x_i - b \geq -1 \tag{15}$$

$$c_i(w \cdot x_i - b) \geq 1, \text{ for all } 1 \leq i \leq n \tag{16}$$

The distance between the two classes is $\frac{2}{\| w \|}$. So to maximize the distance, $\| w \|$ has to be minimized.

IV. Experimental Results

The database consists of MRI of the brain. In our computerized system, a wide variety of images from different patients have been extracted, analyzed and classified. The results have been classified based upon the following evaluation parameters: sensitivity, specificity, and accuracy.

Sensitivity represents the proportion of actual positives that are correctly identified while specificity denotes the proportion of negatives that are correctly identified. Accuracy is the proportion of both true positives and true negatives.

$$\text{Sensitivity} = \frac{\lambda}{\lambda + \alpha} \tag{17}$$

$$\text{Specificity} = \frac{\varepsilon}{\varepsilon + \gamma} \tag{18}$$

TABLE I. Database of HGG and LGG tumor.

Database	Training Dataset	Testing Dataset
HGG	30	50
LGG	30	50

TABLE II. Different parameters for accuracies of HGG and LGG Tumor.

Accuracy	Sensitivity	Specificity
99	100	98.03

TABLE III. Comparison of accuracy.

System	Accuracy
A. Batra and Dr. G. Kaushik [3]	98.2
M. B. M. Amien et al. [4]	98.6
C. L. Devasena and M. Hemalatha [5]	98.8
E. Dandil et al. [6]	91.49
Our Proposed Method	99

$$\text{Accuracy} = \frac{\lambda + \varepsilon}{\lambda + \varepsilon + \alpha + \gamma} \tag{19}$$

Here, λ, ε, γ, and α represents the true positives that detect the positive cases accurately, the true negatives that detect the negative cases accurately, the false positives that wrongly detect the negative cases, the false negatives that wrongly detect the positive cases, respectively.

After running our proposed system, the number of true positives achieved is 49, false positives are 1, and true negatives are 50 and 0 or no false negatives. Our system attains an accuracy of 99%, a sensitivity of 100% and specificity of 98.03% in tumor detection and classification successfully. The obtained results are compared with a few variable methods in Table III.

V. Conclusions and Future Works

Automatic segmentation and classification of the MRI images is the only way by which brain tumor can be detected and classified in early stages with a high percentage of accuracy. The total number of images used in our system is 440. Among them, there are 100 normal images and 340 abnormal images. In the first stage, 100 normal images and 180 abnormal tumorous images are used. In the second stage, 80 HGG and 80 LGG images are used as mentioned in Table I. As observed in Table II, our computerized system has successfully classified HGG and LGG with an accuracy, sensitivity and specificity of 99%, 100% and 98.03% respectively. The system can be made more reliable by using a larger number of data. More relevant features for accurate classification can be found out. This computerized system could be further used for the classification of other brain diseases and for other alternative medical images of different pathological condition, types, and disease status.

Acknowledgment

This research was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2017-0-00392, Development of 1 cm level Precise Localization Technique using Multiple LEDs and Multiple Time-of-Flight (ToF) Cameras for Autonomous Vehicles).

REFERENCES

[1] E.-S. A. El-Dahshan, H. M. Mohsen, K. Revett, and A.-B. M. Salem, "Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm," *Expert Systems with Applications,* vol. 41, no. 11, pp. 5526–5545, 2014.

[2] P. Katti, V. R. Marathe, "Implementation of Classification System for Brain Tumor using Probabilistic Neural Network," *International Journal of Advanced Research in Computer and Communication Engineering,* Vol. 4, no. 10, pp. 188-192, October 2015.

[3] A. Batra, Dr. G. Kaushik, "SECTUBIM: Automatic Segmentation And Classification of Tumeric Brain MRI Images using FHS (FCM, HWT and SVM)," *International Journal of Engineering Science and Computing,* Vol. 7, no.6, pp. 13190-13194, June 2017.

[4] M. B. M. Amien, A. Abd-elrehman, and W. Ibrahim, "An intelligent model for automatic brain-tumor diagnosis based-on MRI images," *International Journal of Computer Applications,* vol. 72, no. 23, pp. 21‑24, June 2013.

[5] C. L. Devasena and M. Hemalatha, "Efficient computer aided diagnosis of abnormal parts detection in magnetic resonance images using hybrid abnormality detection algorithm," *Central European Journal of Computer Science*, vol. 3, no. 3, pp. 117–128, 2013.

[6] J. Cheng, X. Chen, H. Yang, and M. Leng, "An enhanced k-means algorithm using agglomerative hierarchical clustering strategy," *International Conference on Automatic Control and Artificial Intelligence (ACAI 2012).* pp. 407‑410, 2012.

[7] E. Dandil, M. Cakiroglu, Z. Eksi, "Computer-aided diagnosis of malign and benign brain tumors on MR images," *ICT Innovations 2014*, pp. 157–166, 2015.

[8] Midas: healthy human brain database. Available online: http://www.insight journal.org/midas/collection/view/27.

[9] Brats database. Available online: http://www2.imm.dtu.dk/ projects/BRATS2013.

[10] Brats database. Available online: http://www2.imm.dtu.dk/ projects/BRATS2017.