ISSN No: 2348-4845
International Journal & Magazine of Engineering,
Technology, Management and Research
A Peer Reviewed Open Access International Journal

Methodology for Identifying Brain Tumors by Image Mining Techniques



T.Ravichandra Babu
Associate Professor & HOD,
Department of ECE,
Krishnamurthy Institute of
Technology and Engineering.

S.Mounika, M.Tech
Assistant Professor,
Department of ECE,
Krishnamurthy Institute of
Technology and Engineering.

Jangala Sandeep
PG Scholar-SSP,
Department of ECE,
Krishnamurthy Institute of
Technology and Engineering.

Abstract:
This paper is focused on the comparison of three different intensity based feature extraction method for the abnormal patterns in brain tumors. Physician's interpretation of brain tumors may lead to misclassification sometime. Hence an automated system is needed to solve our problem. The following major categories of brain tumor images are taken into our consideration. They are Metastatic bronchogenic carcinoma, Astrocytoma, Meningioma, sarcoma. The performance factor was evaluated against BRATS (Brain Tumor Segmentation) dataset. For the purpose of calculating and extracting various intensity related features MATLAB tool is used. The experimental results suggest that among the intensity based feature extraction methods GLCM (Gray Level Co-Occurrence) method is showing better results than the other methods.

INTRODUCTION:

An automated system is very much needed to classify and detect the tumors in medical images. This task requires higher amount of accuracy since it directly deals with the life of the human being. The technique which we apply is also plays a vital importance to acquire a better result. Various methods are there to collect the medical images easily [3][4]. The idea behind this paper is to obtain different intensity based features for brain tumors[l]and to produce the output of the study as a base for the brain tumor classification purposes. Several intensity features are extracted from the MRI medical images [4]. The textural features are coming into our consideration. These intensity based features are extracted through the MATLAB environment. For image processing task texture features are very important aspect. Textures features [5] have been used in brain tumor classification. Medical images of non-tumor and tumor type can be found and classified quickly by the physician through analyzing intensity based features of medical images. Alteration and variation in the surface of an image is defined as a texture. Also we can define the distribution of gray levels in a neighborhood as texture. An extracted texture feature provides information about the extracted image.

Bio medical imaging includes neural network, support vector machine and fuzzy [10] classifier [6]. There are four types of brain tumor image types taken for our analysis purpose. Astrocytoma is a type of tumor that contains recurrence of masses. Meningioma can be identified by speech hesitation and low memory recall. Metastatic type of tumor contains large mass of edema. Sarcoma is a brain tumor type that carries vision problems associated with it. The intensity histogram and intensity features are coming under first order texture calculation. The extracted texture features from the medical images can be used as a parameter to enhance the classification result. For capturing the pattern in an image the texture is used.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Intensity features like entropy, contrast and uniformity etc. Automatic classification done on abnormal patterns is based on the extracted texture features from the MRI images.

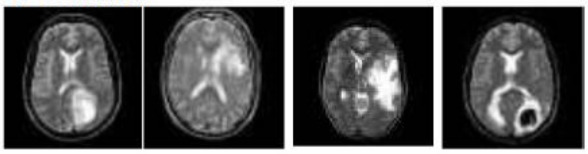
II. RELATED WORK:

In the Literature different techniques are discussed to identify and classify the presence of a tumor. Large amount of research has been done on the textural analysis on MRI images. Textural features for brain tumor classification using GLCM obtained accuracy more than 97% by NitishZulpel and VrushsenPawar2 . An Approach to Medical image classification using Neuro Fuzzy Logic and ANFIS Classifier shows the accuracy greater than 90% by AnantsBhardwaj and Kapil Kumar Siddhu.An improved Image Mining technique for brain tumor classification using efficient classifier by P.Rajendran and M.Madeswaran. Textural features for image classification by R.M.Haralick, K.Shanmugam and I.Distein. Feature Extraction for image mining by P.G.Foschi, D.Kolippakkam.

III. GLCM METHODOLOGY:

The GLCM methodology brain tumor classification task follows the steps.

A. Data Set



This method was evaluated against the dataset BRATS. The BRATS dataset can be obtained through the online by some basic authentication process. The dataset provides the dataset in different dimensions for the experimental usage. Every taken MRI image is of size 256 x 256.

B. Pre Processing

To improve the quality of the image we need to improve the quality of the system. The quality of the image can be enhanced by noise removing and anti-blurring from the image. In this work Gaussian Filter [7] is followed to preprocessing task.

C Feature Extraction Methods:

The visual contents of the images can be captured by feature extraction technique. The purpose of this extraction process is to present the raw image in its normalized form to make decision making process equivalent to the task of pattern Classification. Texture features of MRI [l3] images can be taken through techniques like intensity histogram, co-occurrence matrix and intensity based image features. For better classification rate we need to extract enhanced features from the MRI images. The extracted set of features allows a classifier to distinguish between normal and abnormal pattern. The textural appearance can show the abnormality in the MRI images.

C.a) Intensity Histogram Features

In an image the histogram [8] shows the intensity value per each pixel. In this system we took image in which each pixel can take up to any intensity values from 0 to 256. The features extracted from the images are coming under first order statistics. The features like smoothness, third moment, entropy and uniformity. The histogram is plotted based on these extracted values. They discriminate the two classes of brain tumor images as benign or malign. Based on the intensity value of each pixel the histogram graph was constructed and plotted. G is the number of intensity level.

$$\mu_n = \sum_{i=0}^{G-1} (z_i - m)^n \, p(Z_i) \text{ ------1. } (n^{th} \text{ moment of mean})$$

where z_i be a random variable indicating intensity and let p (z_i), i=0, 1, 2,

$$m = \sum_{i=0}^{G-1} z_i \; p(z_i) \text{ ------2.}$$

$$\sigma = \sqrt{\mu_2}(z) = \sqrt{\sigma^2} \text{ ------3.}$$

C. b) GLCM Features:

Through GLCM features we can extract the textural features of a medical image. GLCM method always focuses on the pixel intensity level of the neighboring

ISSN No: 2348-4845
International Journal & Magazine of Engineering,
Technology, Management and Research
A Peer Reviewed Open Access International Journal

pixel [6]. GLCM always accounts for the specific position of pixel relative to other pixel. This is a simple tabulation that shows how often the different combinations of pixel brightness values occur in medical image. By taking d values as d=l, 2, 3, 4 for the direction of data given for 0 °, 45 °, 90 °, 180 ° matrices of GLCM was constructed [9].P(i,j) represents the probability that 2 pixels with the specified separation have gray levels i and j.

$$\mu = \sum_{i,j=0}^{G-1} i\,p(i,j) \text{ ------4}$$

$$p_i(i) = \sum_{j=0}^{G-1} p(i,j) p_j(j) = \sum_{i=0}^{G-1} p(i,j) \text{--5}$$

$$\mu_i = \sum_{i=0}^{G-1} i\ \ p_i(i) \text{ -----6}$$

$$\mu_j = \sum_{j=0}^{G-1} j\ p_j(j) \text{ -----7}$$

$$\sigma_i^2 = \sum_{i=0}^{G-1} (i - \mu_i)^2\ \ p_i(i) \text{ -----8}$$

$$\sigma_j^2 = \sum_{i=0}^{G-1} (j - \mu_j)^2\ \ p_j(j) \text{ -----9}$$

C.c) Intensity based features

For the purpose of pattern matching and patternrecognition the simplest available element is a pixel. They follow the first order statisticselements. The various intensity based features of amedical image like MRI is collected as mean,median, mode and standard deviation.

IV. Classification

The proposed method uses the MATLAB tool. It captures the collection of several intensity based features of the medical image. Medical images are fed into the MATLAB tool initially. After that by using the tool the intensity based features are collected from the tool itself. Then the output of the intensity features is also compared with the .arff (Attribute relation file format) file[lO] format of the taken image file in WEKA. In the proposed system for 4 different classes I took 4 slices of brain tumor images. MATLAB is comparatively provides better classification results with the GLCM features.

TEXTURE FEATURE EXTRACTION:

This chapter deals with various feature extraction technique based on spatial, transform, edge and boundary, color, shape and texture features. A brief introduction to these texture features is given first before describing the gray level co-occurrence matrix based feature extraction technique.

Texture Features

Guiying Li (2012) defined texture is a repeated pattern of information or arrangement of the structure with regular intervals. In a general sense, texture refers to surface characteristics and appearance of an object given by the size, shape, density, arrangement, proportion of its elementary parts. A basic stage to collect such features through texture analysis process is called as texture feature extraction. Due to the signification of texture information, texture feature extraction is a key function in various image processing applications like remote sensing, medical imaging and content based image retrieval. There are four major application domains related to texture analysis namely texture classification, segmentation, synthesis and shape from texture.

Texture classification produces a classified output of the input image where each texture region is identified with the texture class it belongs. Texture segmentation makes a partition of an image into a set of disjoint regions based on texture properties, so that each region is homogeneous with respect to certain texture characteristics. Texture synthesis is a common technique to create large textures from usually small texture samples, for the use of texture mapping in surface or scene rendering applications. The shape from texture reconstructs three dimensional surface geometry from texture information. For all these techniques, texture extraction is an inevitable stage. A typical process of texture analysis is shown in Figure 4.7.

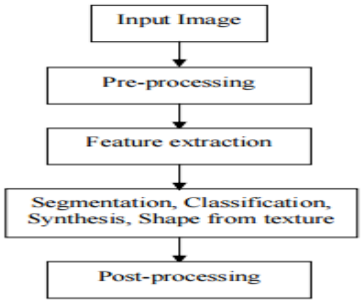


Figure: Feature extraction Process block diagram

Second Order Gray Level Co-occurrence Matrix Features

Some previous research works compared texture analysis methods; Dulyakarn et al. (2000) compared each texture image from GLCM and Fourier spectra, in the classification. Maillard (2003) performed comparison works between GLCM, semi-variogram, and Fourier spectra at the same purpose. Bharati et al. (2004) studied comparison work of GLCM, wavelet texture analysis, and multivariate statistical analysis based on PCA (Principle Component Analysis). In those works, GLCM is suggested as the effective texture analysis schemes. Monika Sharma et al (2012) discussed GLCM is applicable for different texture feature analysis. The GLCM is a well-established statistical device for extracting second order texture information from images. A GLCM is a matrix where the number of rows and columns is equal to the number of distinct gray levels or pixel values in the image of that surface.

GLCM is a matrix that describes the frequency of one gray level appearing in a specified spatial linear relationship with another gray level within the area of investigation. Given an image, each with an intensity, the GLCM is a tabulation of how often different combinations of gray levels co-occur in an image or image section. Texture feature calculations use the contents of the GLCM to give a measure of the variation in intensity at the pixel of interest. Typically, the co-occurrence matrix is computed based on two parameters, which are the relative distance between the pixel pair d measured in pixel number and their relative orientation. Normally, is quantized in four directions (e.g., 0º, 45 º, 90 º and 135 º), even though various other combinations could be possible. GLCM has fourteen features but between them most useful features are: angular second moment (ASM), contrast, correlation, and inverse difference moment, sum entropy and information measures of correlation. These features are thoroughly promising.

GRAY LEVEL CO-OCCURRENCE MATRIX

In 1973, Haralick introduced the co-occurrence matrix and texture features which are the most popular second order statistical features today. Haralick proposed two steps for texture feature extraction. First step is computing the co-occurrence matrix and the second step is calculating texture feature based on the co-occurrence matrix. This technique is useful in wide range of image analysis applications from biomedical to remote sensing techniques. 4.5.1 Working of GLCM Basic of GLCM texture considers the relation between two neighboring pixels in one offset, as the second order texture. The gray value relationships in a target are transformed into the co-occurrence matrix space by a given kernel mask such as 3 3 , 5 5 , 7 7 and so forth. In the transformation from the image space into the co-occurrence matrix space, the neighboring pixels in one or some of the eight defined directions can be used; normally, four direction such as 0°, 45°, 90°, and 135° is initially regarded, and its reverse direction (negative direction) can be also counted into account. It contains information about the positions of the pixels having similar gray level values.

Each element (i, j) in GLCM specifies the number of times that the pixel with value i occurred horizontally adjacent to a pixel with value j . In Figure 4.8, computation has been made in the manner where, element (1, 1) in the GLCM contains the value 1 because there is only one instance in the 103 image where two, horizontally adjacent pixels have the values 1 and 1. Element (1, 2) in the GLCM contains the value 2 because there are two instances in the image where two, horizontally adjacent pixels have the values 1 and 2.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

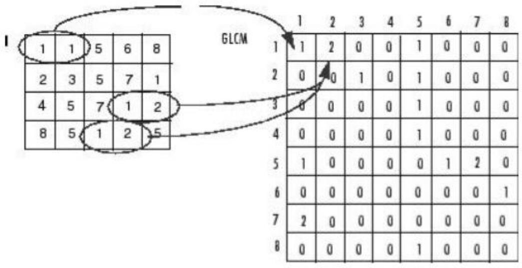


Figure 4.10 Creation of GLCM from image matrix

4.6 APPLICATION OF TEXTURE:

Texture analysis methods have been utilized in a variety of application domains such as automated inspection, medical image processing, document processing, remote sensing and content-based image retrieval.

4.6.1 Remote Sensing

Texture analysis has been extensively used to classify remotely sensed images. Land use classification where homogeneous regions with different types of terrains (such as wheat, bodies of water, urban regions, etc.) need to be identified is an important application.

4.6.2 Medical Image Analysis

Image analysis techniques have played an important role in several medical applications. In general, the applications involve the automatic extraction of features from the image which is then used for a variety of classification tasks, such as distinguishing normal tissue from abnormal tissue. Depending upon the particular classification task, the extracted features capture morphological properties, colour properties, or certain textural properties of the image.

TABLE 1: INTENSITY HISTOGRAM FEATURES

Features	Expression
Smoothness (Smoothness level of intensity in a histogram)	$R = 1 - \frac{1}{1+\sigma^2}$
Third Moment (Skewness of a Histogram)	$\mu = \sum_{i=0}^{G-1} (z_i - m)^3 p(z_i)$
Uniformity (Uniformity of intensity in a histogram)	$U = \sum_{i=0}^{G-1} p^2 z(i)$
Entropy (Measure of randomness)	$e = \sum_{i=0}^{G-1} p(z_i) \log_2 p(z_i)$

Features	Expression
Contrast (Intensity contrast between a pixel and its neighbor)	$\sum_{i,j=0}^{G-1} (i-j)^2 p(i,j)$
Cluster Shade (It is a measure of skewness)	$\sum_{i,j=0}^{G-1} (i + j - \sigma_i - \sigma_j)^3 \ P(i,j)$
Energy (Uniformity of angular second momentum)	$\sum_{i,j=0}^{G-1} p(i,j)^2$
Sum of square variance Puts high weight that differ from the average value of p(i,j)	$\sum_{i,j=0}^{G-1} p(i,j)(1-\mu)^2$

TABLE 3: INTENSITY FEATURES.

Feature	Description
Median	Mean dataset are arranged in ascending order and then middle value is taken as median.
Mode	The mode of mean dataset is the value that occurs most often in mean dataset
Variance	The variability of values in the mean dataset $\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (mean(i) - M)$, wh $M = \frac{1}{n} \sum_{i=1}^{n} mean(i)$
Standard deviation	It is the square root of the variance $SD = \sqrt{\sigma^2}$

V. MEASURES FOR PERFORMANCE EVALUATION:

Different types of techniques are commonly used to evaluate the performance of the proposed method. These measures include the classification analysis accuracy (AC) Mathews Correlation Coefficient (MCC) are calculated from confusion matrix. The confusion matrix describes actual and predicted classes of the proposed methods.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

$$AC = \frac{(TP+TN)}{(TP+FP+TN+FN)}$$

$$MCC = \frac{(TPXTN-FPXFN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

MCC is used to measure the quality of binary classification. The MCC can be calculated from the confusion matrix using the formula. It returns a value from - 1 (inverse prediction) to + 1 (perfect prediction).

RESULTS AND DISCUSSIONS:

From the proposed experiment we observe theperformance of J48 Algorithm is showing close relation with the GLCM feature extraction method. GLCM feature extraction techniques show it isclosely correlating with the leading J48 algorithm in terms of the accuracy. The J48 algorithm features are compared with all the intensity related features. From the discussed features GLCM the textural extraction technique is the one that is very much closer to the J48 tool output of WEKA This Experiment is done in a different dimension to show the textural feature extraction mechanism GLCM is the best performing one to till date apart from other techniques. To show the close correlation we proposed this method. This method uses the .arff file format for tool manipulation.

Fig3: Performance analysis on feature extraction methods.

CONCLUSION:

GLCM textural feature extraction is an important textural feature extraction mechanism and it is widely used for important decision making process for the physician. In this paper it has proposed the comparison of J48 algorithm already provided in WEKA with the intensity based features. The comparative method ensures the entire process of GLCM is having nearest accuracy to the J48algorithm. The experimental results using the comparative study with BRATS dataset depict the various feature outputs. That is compared with theJ48 classifier in WEKA environment with the affine format output. This comparative study between GLCM and J48 shows the close correlation exists among these two techniques compared to other techniques. Our experimental result proves that the comparative study produces better results with theJ48 algorithm of WEKA tool.

FEATURE ENHANCEMENT:

Modern world is well equipped with the physician and machineries. There are several tools to their decision making purpose as well. But Machine vision is so important in medical images. Because of the inability to expose the inbound masses that is present in the brain areas there is still problem in physician decision making process. This problem can be taken further for our consideration and resolved in future.

REFERENCES:

[1] Ricci P.E., Dungan D. H., "Imaging of low- and intermediate-grade gliomas", Seminars in Radiation Oncology, 2001, 11(2), p. 103-112.

[2] NitishZulpel and Vrushsen Pawar2 "GLCM Textural Features for Brain Tumor Classification", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012

[3] M. C. Clark, L. O. Hall, D. B. Goldgof, L.P. Clarke,R. P. Velthuizen, and M. S.Silbiger, "MRI Segmentation using Fuzzy Clustering Techniques", IEEE Engineering in Medicine and Biology, pp. 730-742, 1994.

[4] Chang R.F., Wu W.1., Moon W.K., Chou Y.H., Chen D.R., "Support vector machines for diagnosis of breast tumors on US images", Academic Radiology,2003, 10(2), p. 189-197.

[5] Chang R.F., Wu W.1., Moon W.K., Chou Y.H., Chen D.R., "Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis", Ultrasound in Medicine and Biology, 2003, 29(5), p. 679-686.

[6] M.Kasthuri and S.Britto Ramesh Kumar, "Multilingual Phonetic Based Stem Generation", Second International Conference on Emerging Research in Computing, Information Communication and Applications (ERCICA-2014) ,Volume: I, 31-02 August 2014

[7] W. Chu, C. 1. Ong, and S. S. Keerthi "An improved conjugate gradient scheme to the solution of least squares SVM", IEEE Transactions on Neural Networks 16(2): pp. 498-501, 2005.

[8] V. Thavavela, 1. JafferBasha, R. Murugesan,"A novel intelligent wavelet domain noise filteration technique: Application to medical images", Expert System with Application, 2010.

[9] Abe, Shigeo I Inoue, Takuya, " Fuzzy Support Vector Machines for Pattern Classification" Neural networks, 200I. Proceedings. IJCNN ,0 I

[10] M.Haralick, H.KShanmugam, and Dinstein, "Texture features for image for image classification," IEEE Trans. on Syst. Man Cybernet.,vol. 3, no. 6, pp. 610-621, 1973

[11] Graham Sexton, Keshav Dahal. "An adaptive ensemble classifier for mining concept drifting data streams," Elsevier, Expert Systems with Applications, 40, pp 5895-5906, 2013.

[12]Yongqiao Wang,Shouyang Wang, and K. K.Lai,"Anew fuzzy support vector machine toevaluate credit risk", IEEE Trans. on Fuzzy Systems, vol. 13, no. 6, pp. 820-831, 2005.

[13] Shu-Hsien Liao, Pei-Hui Chu, Pei-Yuan Hsiao. "Data mining techniques and applications A decade review from 2000-2011," Elsevier, Expert Systems with Applications, 39, pp 11303-11311, 2012.

[14] Mr.Deepak, C.Dhanwani, Prof. MahipM.Bartere. "Survey on various techniques of Brain Tumor Detection from MRII mages" International Journal of Computational Engineering Research, Vol4.Issue January 2014,l.Issn 2250-3005.

[15] M.BShweta Jain "Brain Cancer Classification Using GLCM Based Feature Extraction in Artificial Neural Network".International Journal of Computer Science & Engineering Technology.Vol 4,.7 July 2013, Issn 2229 -3345.