

Feature Extraction Using Image Mining Techniques to Identify Brain Tumors

B.Sathees Kumar¹

Asst. Professor in Dept. of Computer Science
Bishop Heber College, Tiruchirappalli,
Tamilnadu, India.
satbhc@gmail.com

Dr.R.Anbu Selvi²

Asst. Professor in Dept. of Computer Science
Bishop Heber College(Autonomous)
Tiruchirappalli, India
r.anbuselvi@gmail.com

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-Occurance) method is showing better results than the other methods. WEKA tool classification algorithm J48 also shows close correlation with GLCM Features.

Index Terms— MRI, Astrocytoma, Metastatic, Meningioma, Sarcoma, Glcm, Mat lab,

I. 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[1]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. 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.

The structure of the paper is as follows. In section II related works are discussed. The Section III deals with the Glcm methodology. The Section IV deals with classification comparisons. The Section V deals with performance evaluation. The Section VI deals with results and discussions. The section VII deals with the conclusion and the section VIII Feature Enhancement aspect is discussed.

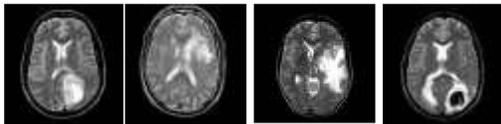
II. RELATED WORK

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 Nitish Zulpe1 and Vrushsen Pawar2. 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 [13] 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.

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

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 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 $d=1, 2, 3, 4$ for the direction of data given for $0^\circ, 45^\circ, 90^\circ, 180^\circ$ 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 .

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

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

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

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

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

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

C.c) Intensity based features

For the purpose of pattern matching and pattern recognition the simplest available element is a pixel. They follow the first order statistics elements. The various intensity based features of a medical 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[10] 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.

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)	$e = \sum_{i=0}^{G-1} p(z_i) \log_2 p(z_i)$

TABLE 2: GLCM FEATURES

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.

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

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

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).

TABLE4: FEATURE VALUES EXTRACTED FROM 4th SLICE

Feature Type	Feature	Astrocytoma (abnormal)	Meningioma (abnormal)	Metastatic (abnormal)	Sarcoma (abnormal)
<i>GLCM</i>	Contrast	1.744	1.655	2.0221	0.9987
	Energy	0.999	0.988	0.5678	1.235
	Cluster Shade	1.2545	0.9544	0.9765	2.3434
	Sum of Square Variance	1.0999	2.0982	3.8565	2.6543
<i>Intensity Histogram</i>	Smoothness	-1.2995	-1.2876	-2.2341	-4.3243
	Third Moment	3.5754	2.3445	1.9876	2.3667
	Uniformity	1.2563	0.9896	0.7654	0.2234
	Entropy	0.8903	1.2223	0.6777	1.4534
<i>Intensity</i>	Median	50.959	57.87	79.87	59.98
	Mode	58.788	52.33	76.75	56.75
	Variance	125.5	151.11	58.81	5.432
	Standard Deviation	0.6312	0.6221	0.6312	0.5892

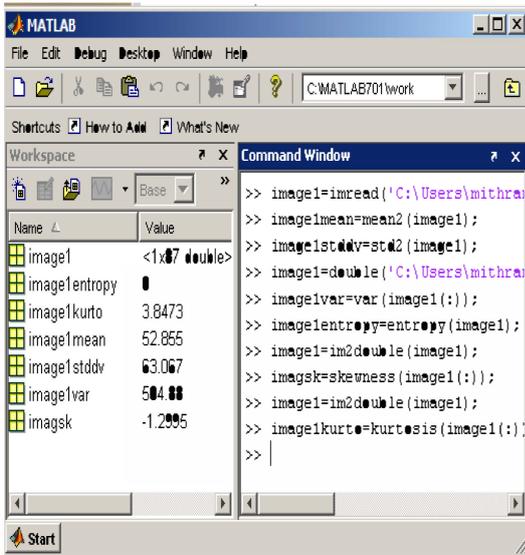


Fig.1: Feature extraction through MATLAB tool.

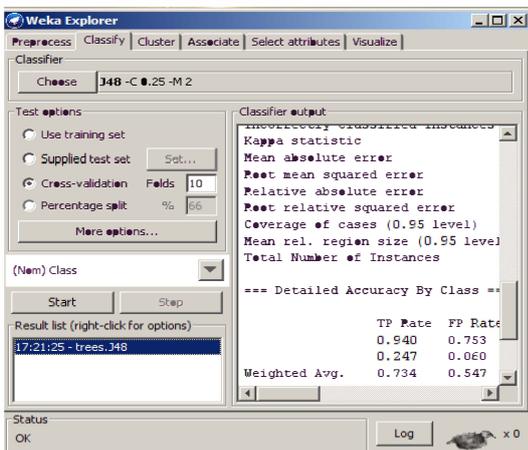


Fig.2: Feature extraction from BRATS dataset

The effectiveness of texture features are tested and compared against the J48 classifier.

The dataset used for this experiment is composed of 250 brain tumor MRI images from the BRATS database which includes 150 normal and 150 abnormal, 80% (200 out of 250) set of images are used for training and 20% (50 out of 250) used for testing. The results are compared with the outcome of the J48 output.

TABLE6: CONFUSION MATRIX FOR INTENSITY HISTOGRAM FEATURES

Actual	Predicted	
	Cancer (Positive)	Normal (Negative)
Cancer(Positive)	24 (TP)	4 (FP)
Normal(Negative)	2 (FN)	20 (TN)

TABLE 7: CONFUSION MATRIX FOR INTENSITY FEATURES

Actual	Predicted	
	Cancer (Positive)	Normal (Negative)
Cancer(Positive)	24 (TP)	3 (FP)
Normal(Negative)	1 (FN)	23 (TN)

**TABLE 8: CONFUSION MATRIX
BASED ON GLCM FEATURES**

Actual	Predicted	
	Cancer (Positive)	Normal (Negative)
Cancer(Positive)	24 (TP)	1 (FP)
Normal(Negative)	1 (FN)	24 (TN)

**TABLE 9: INTENSITY FEATURE COMPARED
WITH J48**

Measures	Intensity histogram	GLCM	J48	Intensity
AC(%)	89	95.25	97.50	93
MCC(-1 to +1)	0.84	0.96	0.98	0.92

VI. RESULTS AND DISCUSSIONS

From the proposed experiment we observe the performance of J48 Algorithm is showing close relation with the GLCM feature extraction method. GLCM feature extraction techniques shows it is closely 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.

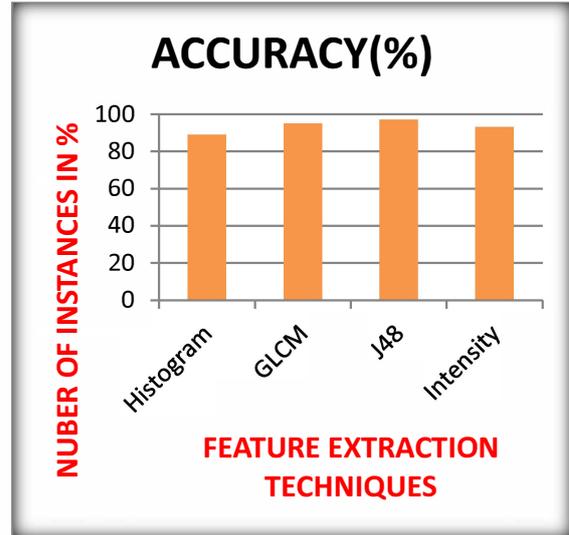


Fig3: Performance analysis on feature extraction methods.

VII. 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 J48 algorithm. The experimental results using the comparative study with BRATS dataset depict the various feature outputs. That is compared with the J48 classifier in WEKA environment with the arff file 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 the J48 algorithm of WEKA tool.

VIII. 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.

*IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and
Communication Systems ICIIECS'15*

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] Nitish Zulpe1 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.J., 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.J., 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:1, 31-02 August 2014
- [7] W. Chu, C. J. 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, J. Jaffer Basha, R. Murugesan,"A novel intelligent wavelet domain noise filtration technique: Application to medical images", *Expert System with Application*, 2010.
- [9] Abe, Shigeo / Inoue, Takuya, " Fuzzy Support Vector Machines for Pattern Classification" *Neural networks*, 2001. *Proceedings. IJCNN* ,01
- [10] M.Haralick, H.KShanmugam, and Dinstein,"Texture features for image for imageclassification," *IEEETrans. on Syst. ManCybernet.*, 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,"A new fuzzy support vector machine to evaluate 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. Mahip M.Bartere. "Survey on various techniques of Brain Tumor Detection from MRII mages" *International Journal of Computational Engineering Research*,Vol4.Issue January 2014,1.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.