International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

A Hierarchical Model to Classify Brain Cancer using GLCM

Vaishnavi. B [1], Uma. R [2]

[1]MPhil Scholar, Department of CS & IT, Nadar Saraswathi College of Arts and Science, Theni, Tamilnadu, India

[2]Head, Department of Computer Application, Nadar Saraswathi College of Arts and Science, Theni, Tamilnadu, India

Abstract: ***The Cancer is the next stage or effect of tumorous cells. This research aims to detection of brain cancer using Matlab and Data Mining. This proposed research is classifying the cancer in MR images of the cancer affected patients. The major steps involved in the system are Preprocessing, Feature Extraction and Decision Tree Classification. The Image Processing techniques such as Median filtering have been developed for detection of noise in the MR images. In this research, extraction of features in the cancer image is achieved by using Gray Level Cooccurrence Matrix (GLCM). Then J48 classifier is used to classify MRI brain images into abnormal and healthy image. The proposed algorithm is very useful to gives more information about the cancer which is very helpful to the doctor for subjective analysis.***

Keywords: Data mining, Decision Tree, Magnetic Resonance Imaging (MRI), Gray Level Cooccurrence Matrix (GLCM)

1. Introduction

Brain cancer is a disease of the brain in which cancer cells (malignant) arise in the brain tissue [1]. Statistics suggest that brain cancer occurs infrequently and is likely to develop in about 22,000 new people per year with about 13,000 deaths as estimated by the National Cancer Institute (NCI) [9].

The importance and necessity in accurate brain pathology diagnosis and treatment requires more accuracy in automatically classifying MRI images in distinguishing disease without human interference. MRI is currently the most effective imaging modality for brain cancer screening. However, 10-30% of brain cancers are missed at MRI scan. For this, this proposed system developed using Digital Image Processing and Data Mining. Data mining of medical images is used to collect effective models, relations, rules, abnormalities and patterns from large volume of data. This procedure can accelerate the diagnosis process and decision-making. Different methods of data mining have been used to detect and classify anomalies in MRI images such as GLCM , statistical methods and most of them used feature extracted using image processing techniques.

2. Methodology

2.1 Generalized View

The proposed framework consists of following main steps which are
1) Give MRI image of brain as input.
2) Convert it to gray scale image.
3) Apply median filter to enhance the quality of image.
4) Extracting features by using GLCM
5) Construct decision tree based on features

The architecture workflow depicts the steps involved and the corresponding input and output. The methodology used for MRI brain images is as shown in Fig. 2.1

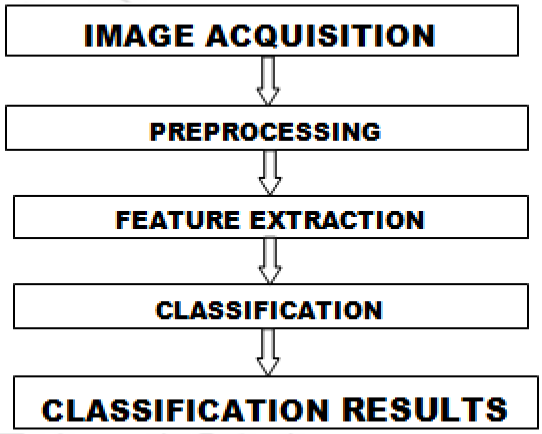

Figure 2.1: Block diagram of the system

2.2 Image Acquisition

The research consists of 60 images which belong to two types as Normal and Abnormal. There are 24 normal 36 abnormal images.

2.3. Image Preprocessing

After obtaining digital images, image preprocessing techniques can be further used for analysis of region of interest. The pre-processing is used to read the input image into the MATLAB and also to remove the noise present in the image. Image preprocessing consists mainly of following steps.
- Resize
- Gray conversion
- Median filter

The acquired MRI scanned image, stored in database is converted to gray scale image of size 256*256. It includes median filter for noise removal.

2.3.1 Median filter

In this work an efficient filter referred to as the median filter, is applied to the image. The median filter is a non linear digital filtering technique, is often used to remove noise. Median filtering is very widely used in digital image processing because, under certain conditions, it preserves edges while removing noise. It works by moving pixel by pixel through the image, replacing each value with the median value of neighboring pixels. The pattern of neighbors is called the ""window,"" which slides pixel by pixel over the entire image. The median is calculated by first sorting all the pixel values from the window into numerical order, and then replacing the pixel being considered with the middle (median) pixel value. Image processing researchers commonly assert that median filtering is better than linear filtering for removing noise in the presence of edges.

2.4 Feature Extraction

The third step of the proposed work is feature extraction. Transforming the input data into the set of features is called feature extraction. Features are used as inputs to classifiers that assign them to the class that they represent. In this work Gray Level Cooccurrence Matrix (GLCM) features are extracted.

2.4.1 GLCM Features

GLCM is a statistical method of examining texture that considers the spatial relationship of pixels is the Gray Level Cooccurrence Matrix (GLCM), also known as the gray-level spatial dependence matrix. A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G, in the image.

Haralick has extracted many properties or features from GLCM. The graycomatrix function in MATLAB creates a gray-level co-occurrence matrix (GLCM) by calculating how often a pixel with the intensity (gray-level) value *i* occurs in a specific spatial relationship to a pixel with the value *j*. Grey Level Cooccurence Matrix can be defined as:

$$C_{\Delta x,\Delta y}(i,j)=\sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases}1, & \text{if } I(p,q)=i \text{ and } I(p+\Delta x,q+\Delta y)=j\\ 0, & \text{otherwise}\end{cases}$$

Where m and n are grey levels normally m=n for symmetry GLCM.

The basic GLCM algorithm is as follow:

1. Count all pairs of pixels in which the first pixel has a value i, and its matching pair displaced from the first pixel by d has a value of j.
2. This count is entered in the ith row and jth column of the matrix Pd[i,j]
3. Note that Pd[i,j] is not symmetric, since the number of pairs of pixels having gray levels[i,j]does not necessarily equal the number of pixel pairs having gray levels [j,i].
4. The elements of Pd[i, j] can be normalized by dividing each entry by the total number of pixel pairs.
5. Normalized GLCM N[i, j], defined by:

$$N[i,j]=\frac{P[i,j]}{\sum_i\sum_j P[i,j]}$$

The Following GLCM features are extracted in this research work:

- Contrast
- Correlation
- Homogeneity
- Entropy
- Energy

i) Contrast

Contrast is defined as the separation between the darkest and brightest area. It is the difference between the highest and the lowest values of a contiguous set of pixels.

$$Contrast=\sum_{i,j=0}^{n-1}P_{i,j}\,(i-j)^2$$

ii) Correlation

Correlation is a measure of gray tone linear-dependencies in the image; in particular, the direction under investigation is the same as vector displacement.

$$Correlation=\sum_{i,j=0}^{n-1}P_{ij}\,\frac{(i-\mu)(j-\mu)}{\sigma^2}$$

iii) Homogeneity

Homogeneity gives information about how little change there is in an image. Homogeneity is defined as the quality or state of being homogeneous.

$$Homogenity=\sum_{i,j=0}^{n-1}\frac{P_{ij}}{1+(i-j)^2}$$

iv) Entropy

Entropy is a measure of the uncertainty in a random variable. Higher entropy values are extracted from homogeneous scenes, and lower ones are from inhomogeneous scenes.

$$Entropy=\sum_{i,j=0}^{N-1}-ln(P_{ij})P_{ij}$$

v) Energy

Energy parameter is also called as Uniformity**.** Energy is a feature that measures the smoothness of the image.

$$Energy=\sum_{i,j=0}^{N-1}(P_{ij})^2$$

2.5 Classification by Decision Tree Induction

Classification is the process where a given test sample is assigned a class on the basis of knowledge gained by the classifier during training. Classification technique is applied to the brain cancer database. Extracted texture features using GLCM of brain cancer images are considered as values, and each image feature representation is a record**.** The data records have been created in Excel data sheet and saved in the format of CSV and then converted into the accepted WEKA format ARFF. Hence we propose a model for better classification. In this research J48 classifier is used for the classification of brain MRI image into healthy brain or cancer brain.

2.5.1 Data Description

The MRI image data description of the proposed method is shown in table.

Table 2.1: Dataset Description

S. No	*Type of Image*	*Number of images*
1.	Normal Images	24
2.	Abnormal Images	36

For each region, we calculated the following five set of features. The attributes descriptions are shown in Table 2.2 below.

Table 2.2: Attribute Description

No	*Attribute*	*Category*
1	Contrast	Numeric
2	Correlation	Numeric
3	Homogeneity	Numeric
4	Entropy	Numeric
5	Energy	Numeric
6	Brain Image	Class

J48 classifier is a simple C4.5 decision tree for classification. It creates a binary tree. The decision tree approach is most useful in classification problem. With this technique, a tree is constructed to model the classification process. Once the tree is built, it is applied to each tuple in the database and results in classification for that tuple.

```
Algorithm J48:
INPUT
D //Training data
OUTPUT
T //Decision tree
DTBUILD (*D)
{
        T=φ;
        T= Create root node and label with splitting
        attribute;
        T= Add arc to root node for each split predicate
        and   label;
        For each arc do
                D= Database created by applying splitting
                predicate to D;
                If stopping point reached for this path,
        then
                        T‟= create leaf node and label
                        with appropriate class;
                Else
                        T‟= DTBUILD(D);
                        T= add T‟ to arc;
}
```

3. Result and Discussion

An experiment has been conducted on a MRI brain image data set based on the proposed flow diagram as shown in Fig 2.1. Fig 3.1 represents the abnormal MRI as an input image. Fig 3.2 represents the gray scale image. Fig 3.3 shows the median filtered of MRI scan brain image.

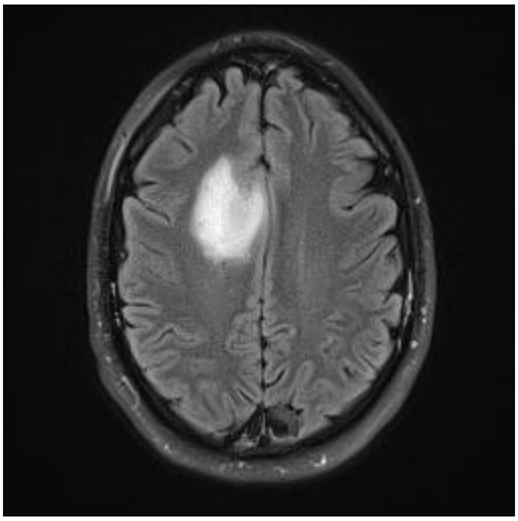

Figure 3.1: Cancer Affected Brain Image

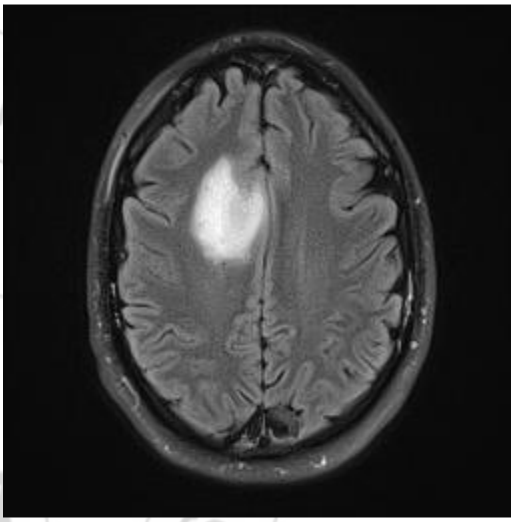

Figure 3.2: Grayscale Image

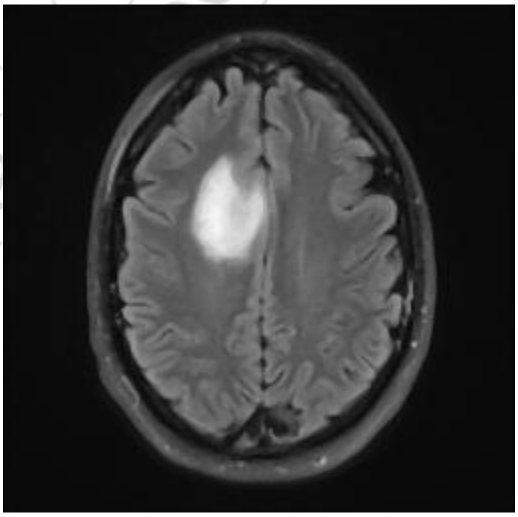

Figure 3.3: Median Filtered Brain Image

Input values for the decision tree are features extracted using GLCM such as contrast, correlation, homogeneity, energy, entropy. Then texture feature set listed in “Table 3.1” is extracted using GLCM statistical method.

Table 3.1: GLCM Features for Normal &Abnormal MRI

No	*GLCM Features*	*Normal*	*Abnormal*
1	Contrast	0.1570	0.2465
2	Correlation	0.9271	0.9737
3	Homogeneity	0.9356	0.9108
4	Energy	0.3330	0.3161
5	Entropy	1.5918	1.9494

From this database the decision tree has been constructed using the J48 algorithm. Decision tree J48 implements Quinlan‟s C4.5 algorithm for generating pruned tree. The constructed decision tree was shown in Fig 3.2. In this proposed method J48 Classifier was used to classify the MRI brain images into two categories namely normal and abnormal.

J48 pruned tree

CON <= 0.1439: A (24.0)
CON > 0.1439
| COR <= 0.9473: N (24.0)
| COR > 0.9473: A (12.0)

Number of Leaves: 3
Size of the tree : 5
Time taken to build model: 0.03 seconds

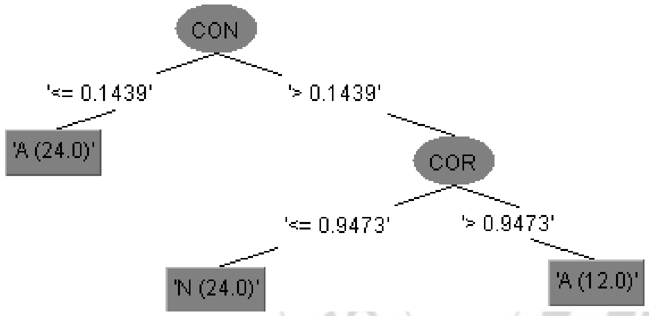

Figure 3.4: J48 Decision Tree for Brain Cancer

3.1 Performance Evaluation

The J48 algorithm gives the following correctness results for the given dataset.

Table 3.2: Performance Results from J48 Classification Algorithm

	No. of Instances	*Percentage*
Correctly Classified Instance	60	100%
Incorrectly Classified Instances	0	0

Confusion Matrix: The confusion matrix is used to measure the performance of two class problem for the given data set. The right diagonal elements TP (true positive) and TN (true negative) correctly classify Instances as well as FP (false positive) and FN (false negative) incorrectly classify Instances. Confusion Matrix Correctly Classifiy Instance and TP+TN Incorrectly Classifiy Instance FP+FN.

Table 3.3: Confusion Matrix for Proposed Method

Actual	*Predicted*	
	Abnormal	Normal
Abnormal	36	0
Normal	0	24

Accuracy: It is defined as the ratio of correctly classified instances to total number of instances. The effectiveness of the proposed method has been estimated using accuracy measure.

Accuracy = (TP + TN) / (TP+TN+FP+FN)

where,

True Positive (TP): Number of Abnormal images correctly classified

False Positive (FP): Number of Normal images classified as Abnormal

True Negative (TN): Number of Normal images correctly classified

False Negative *(FN):* Number of Abnormal images classified as Normal.

The derived features sets are then applied for classifying the new dataset of MRI images as normal and abnormal cases automatically. The result of the prediction provides 100% accuracy in classification.

4. Conclusion

Experiments are conducted on various real-world datasets and the results concluded that the proposed algorithm yield good results when compared with the other classifiers. The goal is classifying the tissues into two classes of normal and abnormal brain MR images. The median filtering techniques have efficiently reduces the noises present in the MRI scan brain images. The brain cancer detection system is designed by using Gray Level Co-Occurrence Matrix (GLCM). The tree generated using decision tree induction method clearly shows that the time taken to classify abnormal and normal cases in just 0.03 seconds. The accuracy of 100% is found in classification of brain cancer. The developed brain cancer classification system is expected to provide valuable diagnosis techniques for the physicians.

Our future work is to extend our proposed method for color based segmentation of 3D images. In future, other feature extraction techniques like Spatial Gray Level Dependence Method (SGLDM) and Run Difference Method (RDM) can also be used to extract the features. Finally, we should do comparison between different classifiers‟ methods.

References

[1] http://neurosurgerybasics.com/brain/brain-mri-a-systematic-reading
[2] http://www.braintumourresearch.org/statistics
[3] Dipali M. Joshi, Dr.N. K. Rana, V. M. Misra, ”Classification of Brain Cancer Using Artificial Neural Network”, 2010 IEEE ,Electronic Computer
[4] Gonzalez, R.C. Richard, E.W; “Digital Image Processing,” (2004), II Indian Edition, Pearson Education, New Delhi, India.
[5] Haralick R. M., Shanmugam K., Dinstein I., “Textural Features of Image Classification”, IEEE Transactions on Systems, Man and Cybernetics. 1973. vol. 3(6). pp: 610-621.
[6] A.Padma ,R.Sukanesh, “ Automatic Classification and Segmentation of Brain Tumor in CT Images using Optimal Dominant Gray level Run length Texture Features”, (IJACSA) International Journal of Advanced

Computer Science and Applications, vol. 2, no. 10, pp: 53-59, 2011
[7] http://www.mathworks.in/
[8] Masud Karim, Rashedur M. Rahman, "Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing", Journal of Software Engineering and Applications, no.6, pp:196-206, April 2013
[9] WHO, 2014, World cancer report, World health organization. , 2014
[10] BTAS, 2015, "Brain tumour facts 2011". Brain tumour alliance Australia. Retrieved 9 april 2015.
[11] Laboratory of Neuro Imaging (LONI), web, http://www.loni.ucla.edu/.
[12] Aswini Kumar Mohanty, Saroj Kumar Lenka, "Efficient Image Mining Technique for Classification of Mammograms to Detect Breast Cancer", International Conference [ICCT-2010], vol. 2, no. 2, December 2010
[13] Jiawei Han, Micheline Kamber.: Data Mining: Concepts and Techniques. Simon Fraser University, 2001.
[14] Rafael C. Gonzalez, Richard E. Woods.: Digital Image Processing, 1992.
[15] http://weka.wikispaces.com/Use+WEKA+in+your+Java+code#Classification-Building a Classifier.
[16] Preeti Chouhan, Mukesh Tiwari, "Image Retrieval Using Data Mining and Image Processing Techniques", International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering, vol. 3, no. 12, ISSN:2321 – 2004, December 2015