Brain tumour stage classification using texture based features and Random forest classifier

Zainab Kareem Shakir, [2]Prof.E.Sreenivasa reddy
[1]*PhD Scholar, Dept.of Computer Science and Engineering, A.N.U.College of Engineering and Technology, A.N.U.University, Nagarjuna Nagar, Guntur District, A.P, India*
[2]*Dept.of Computer Science and Engineering, A.N.U.College of Engineering and Technology, A.N.U.University, Nagarjuna Nagar, Guntur District, A.P, India*

Abstract- Brain diseases is one of the major cause of cancer related death among children and adults in the world. Brain diseases like brain tumor is characterized as a gathering of abnormal cells that becomes inside the brain and around the brain. There are various imaging techniques which are used for brain tumor detection. Among all imaging technique, MRI (Magnetic Resonance Imaging) is widely used for the brain tumor detection. MRI is safe, fast and non-invasive imaging technique. This paper presents a novel brain tumour stage detection algorithm. The tumour part is divided into blocks and Shape and texture based features such as Grey level co-occurrence matrix (GLCM) are extracted. Principal Component Analysis (PCA) is used for the dimensionality reduction of the features. This improves the classification accuracy of the algorithm. The proposed algorithm outperforms the existing techniques in terms of the classification accuracy.

Keywords: texture, MRI, Grey level co-occurrence matrix, Principal Component Analysis, Random forest.

I. INTRODUCTION

Brain diseases like tumors are the one of the major mainsprings for the increase in fatality among the children, male and female. The one can define the brain tumor as the growth of the abnormal cells in the human brain or around the human brain. Based on the survey it has been observed that there are many brain tumors in which some of brain tumor are cancerous or malignant and some of the brain tumor are noncancerous or benign. The NBTF (National Brain Tumor Foundation) of United States has been observed that the brain tumor is the reason for one-fourth of all cancer deaths in children [1]. Early recognition of the brain diseases is the imperative and the inspiration for further studies. The brain images generated by the MRI (Magnetic resonance imaging) is more accurate for the examination of the brain diseases if any are present and for the further analysis of tumor area the physician also needs the help of computer and image processing techniques. On the other side, the quick development of an automatic system are taken place last few decades. The example of such system is CAD (Computer-aided diagnostic) system. The main motive or idea of the CAD system is to facilitate the radiologists in the analysis of the medical images with the help of dedicated computers. Basically, the CAD systems are used to enhance the diagnostic accuracy of the radiologists. The CAD system helps to reduce the workload, chance of miss classification due to fatigue [1]. But the final decision is made by the radiologists. Subsequently, radiologists expect that CAD system can enhance their analytic capacities in light of synergistic impacts between the radiologist and the computer with medical image investigation and machine learning methods [2]. Along these, the CAD system ought to have the capacities like doctors and radiologist as far as in terms of learning and identification of the brain diseases. Hence for the improvement of the CAD system pattern recognition techniques like machine learning play the important roles.

In recent years, for the feature extraction and classification of the brain MR images various technique have been suggested by different researchers. Extracting essential feature from brain MR image is very important for further analysis and classification.

Chaplot et al. [3] have introduced a scheme for feature extraction and classification. To validate the introduced system they are taken a standard dataset of 52 brain MRI images. For feature extraction, they consider coefficient of level-2 approximation subband of 2D DWT. Daubechies-4 (DAUB4) filter is used as decomposition filter. After getting the features they employed self organizing map (SOM) and support vector machine (SVM) as classifier and they achieved higher classification rate for SVM with radial basis function (RBF) classifier.

Maitra and Chatterjee [4] have proposed a scheme for feature extraction and classification. For the feature extraction they have used slantlet transform (ST) and for the classification they used back-propagation neural network (BPNN) and archived ideal result. In [5] they introduced a scheme, they used ST for feature extraction and fuzzy c-means for classification and from the experimental result they observed that the proposed scheme outperformed.

Selvaraj et al. [6] suggested a system for brain MR image classification. For classification they have used many classifier i.e. SVM classifier, Neural classifier, statistical classifier. El-Dahshan et al. [7] suggested a technique. The suggested technique comprises three stages i.e. feature extraction, feature reduction and classification. For feature extraction the approximation subband of DWT is considered. Principal component analysis (PCA) is used for feature reduction and for the classification feed forward back-propagation neural network (FP-ANN) and k-nearest neighbor (k-NN) used as classifier.

Zhang et al. [8] have proposed a scheme for classification. They have taken 160 images (20 normal,140 abnormal) to validate the scheme. For feature extraction level-3 approximation component using Haar wavelet is used. After feature extraction, PCA is used for feature reduction and for the classification forward neural network is used.

Saritha et al. [9] suggested a scheme, in which they have used entropy of wavelet approximation component at level-8 computed along with SWP for feature extraction. For the classification they used Probabilistic neural network (PNN) and their results indicate that they achieve high success rate.

El-Dahshan et al.[1] suggested a hybrid technique, in which feed forward pulse-coupled neural network is applied for the segmentation of the brain images. For feature extraction they consider approximation component of DWT. For feature reduction they used PCA and for the classification they used back propagation neural network.

Yang et al. [10] suggested a wavelet-energy based approach for brain MR image classification. For feature extraction they have used 2D DWT. For brain image classification SVM classifier was employed and BBO method was utilized to optimize the weights of the SVM. They noticed that their scheme was superior then KSVM, PSO-KSVM and BPNN.

Nayak et al. [11] have proposed hybrid technique for brain MR image classification. For feature extraction through brain MR images they utilizes the approximation coefficient of level-3 of discrete wavelet transform (DWT). To reduce the large set of extracted features from brain MR images they have employed kernel principal component analysis (KPCA). After getting the reduced set of features they have employed least square support vector machine (LS-SVM) as a classifier with different kernel function and they have reported that proposed scheme outperform with high accuracy.

II. PROPOSED FRAMEWORK

The proposed framework consists of the following steps:

1. Extract the tumour from the image.

Statistic	Description
Contrast	Measures the local variations in the gray-level co-occurrence matrix.
Correlation	Measures the joint probability occurrence of the specified pixel pairs.
Energy	Provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment.
Homogeneity	Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

2. Divide the tumour into blocks and extract the features of each block
3. Dimensionality reduction using PCA
4. Random forest classification

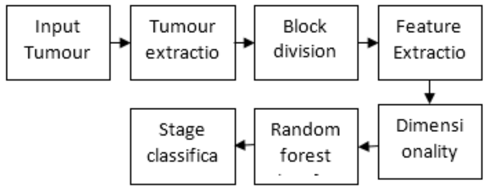

Figure 1: Block diagram of the proposed algorithm

A) Tumour Extraction

The input image is multiplied with the ground truth to extract the tumour part of the image.

B) Block division

The tumour is divided into blocks of uniform size.

C) Feature Extraction

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. (The texture filter functions, described in Texture Analysis cannot provide information about shape, that is, the spatial relationships of pixels in an image.)

After you create the GLCMs, using graycomatrix, you can derive several statistics from them using graycoprops. These statistics provide information about the texture of an image.

A) Dimensionality reduction using PCA

In statistics, machine learning, and information theory, dimensionality reduction or dimension reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables. It can be divided into feature selection and feature extraction.

The main linear technique for dimensionality reduction, principal component analysis, performs a linear mapping of the data to a lower-dimensional space in such a way that the variance of the data in the low-dimensional representation is maximized. In practice, the covariance (and sometimes the correlation) matrix of the data is constructed and the eigenvectors on this matrix are computed. The eigenvectors that correspond to the largest eigenvalues (the principal components) can now be used to reconstruct a large fraction of the variance of the original data. Moreover, the first few eigenvectors can often be interpreted in terms of the large-scale physical 1087ehaviour of the system. The original space (with dimension of the number of points) has been reduced (with data loss, but hopefully retaining the most important variance) to the space spanned by a few eigenvectors.

The following table lists the statistics.

Table 1: GLCM features

The shape based properties are discussed in table 2.

Table 2: Shape based region features

Property Name	Description
'Area'	Actual number of pixels in the region, returned as a scalar. (This value might differ slightly from the value returned by bwarea, which weights different patterns of pixels differently.) To find the equivalent to the area of a 3-D volume, use the 'Volume'property of regionprops3.
'BoundingBox'	Smallest rectangle containing the region, returned as a 1-by-Q*2vector, where Q is the number of image dimensions. For example, in the vector [ul_corner width], ul_corner specifies the upper-left corner of the bounding box in the form [x y z ...]. width specifies the width of the bounding box along each dimension in the form [x_width y_width ...].regionprops uses ndims to get the dimensions of label matrix or binary image, ndims(L), and numel to get the dimensions of connected components,numel(CC.ImageSize).
'Centroid'	Center of mass of the region, returned as a 1-by-*Q* vector. The first element of Centroid is the horizontal coordinate (or *x*-coordinate) of the center of mass. The second element is the vertical coordinate (or *y*-coordinate). All other elements of Centroid are in order of dimension. This figure illustrates the centroid and bounding box for a discontiguous region. The region consists of the white pixels; the green box is the bounding box, and the red dot is the centroid.
'ConvexArea'	Number of pixels in 'ConvexImage', returned as a scalar.
'ConvexHull'	Smallest convex polygon that can contain the region, returned as a *p*-by-2 matrix. Each row of the matrix contains the *x*- and *y*-coordinates of one vertex of the polygon.

B) Random forest classifier

Random Forest is a supervised learning algorithm. Like you can already see from it's name, it creates a forest and makes it somehow random. The "forest" it builds, is an ensemble of Decision Trees, most of the time trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result.

One big advantage of random forest is, that it can be used for both classification and regression problems, which form the majority of current machine learning systems. I will talk about random forest in classification, since classification is sometimes considered the building block of machine learning. Below you can see how a random forest would look like with two trees:

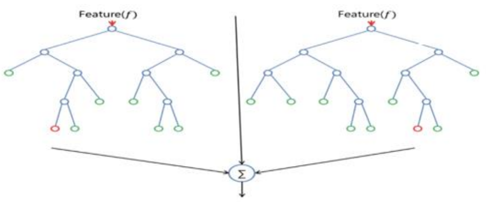

Figure 2: Random forest representation

Random Forest has nearly the same hyper-parameters as a decision tree or a bagging classifier. Fortunately, you don't have to combine a decision tree with a bagging classifier and can just easily use the classifier-class of Random Forest. Like I already said, with Random Forest, you can also deal with Regression tasks by using the Random Forest regressor.

Random Forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.

Therefore, in Random Forest, only a random subset of the features is taken into consideration by the algorithm for splitting a node. You can even make trees more random, by additionally using random thresholds for each feature rather than searching for the best possible thresholds (like a normal decision tree does).

III. EXPERIMENTAL RESULTS

The following images are the results obtained by the proposed framework.

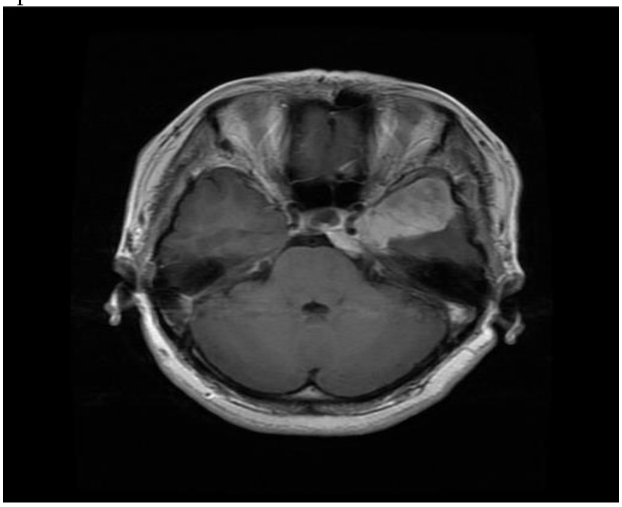

Figure 3: Input tumour image

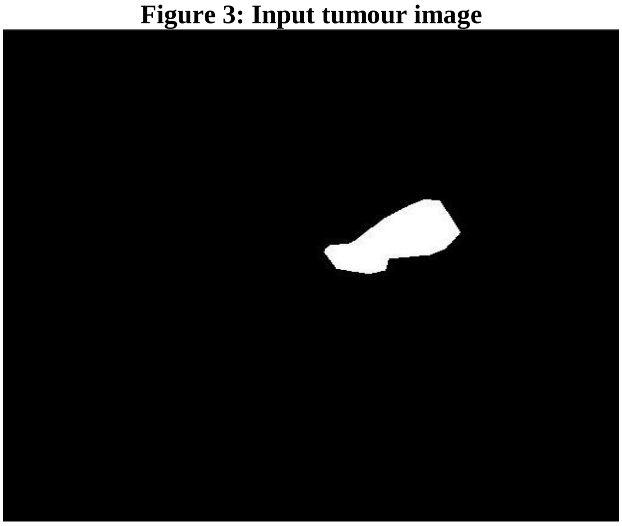

Figure 4: Mask of the tumour

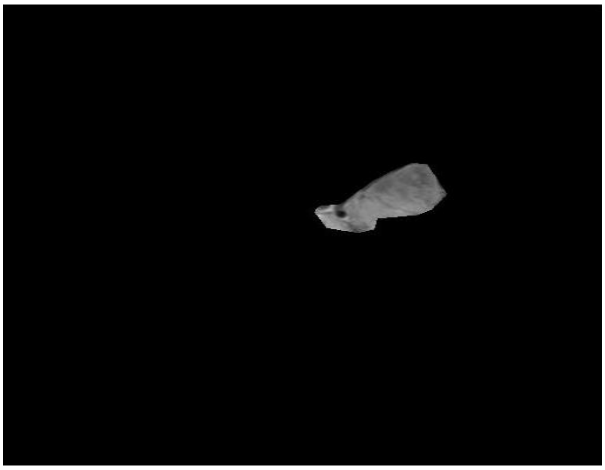

Figure 5: Segmented tumour image

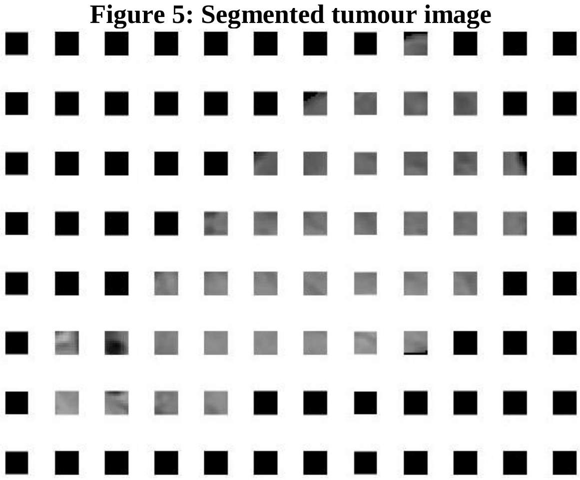

Figure 6: Block divided tumour

The algorithm is run on 3046 tumour images with images belonging to three tumour categories namely

- Meningioma
- Glioma
- Pituitary

The accuracy of the proposed algorithm is 86%.

IV. CONCLUSION

This paper presents a classification scheme for differentiating adult brain tumors using MRI images. Shape characteristics, texture features on image intensities are extracted from the tumour region. The scheme is fully automated and the help of an expert is not required. Random forest based classification of texture patterns is a very promising approach to developing an objective and quantitative evaluation of brain tumors.

V. 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] H. Arimura, C. Tokunaga, Y. Yamashita, and J. Kuwazuru, "Magnetic resonance image analysis for brain cad systems with machine learning," Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis: Medical Imaging Intelligence and Analysis, p. 258, 2012.

[3] S. Chaplot, L. Patnaik, and N. Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network," Biomedical Signal Processing and Control, vol. 1, no. 1, pp. 86–92, 2006.

[4] M. Maitra and A. Chatterjee, "A slantlet transform based intelligent system for magnetic resonance brain image classification," Biomedical Signal Processing and Control, vol. 1, no. 4, pp. 299–306, 2006.

[5] Madhubanti Maitra and Amitava Chatterjee, "Hybrid multiresolution slantlet transform and fuzzy c-means clustering approach for normal-pathological brain mr image segregation," Medical engineering & physics, vol. 30, no. 5, pp. 615–623, 2008.

[6] H. Selvaraj, S. T. Selvi, D. Selvathi, and L. Gewali, "Brain mri slices classification using least squares support vector machine," International Journal of Intelligent Computing in Medical Sciences & Image Processing, vol. 1, no. 1, pp. 21–33, 2007.

[7] E.-S. A. El-Dahshan, T. Hosny, and A.-B. M. Salem, "Hybrid intelligent techniques for mri brain images classification," Digital Signal Processing, vol. 20, no. 2, pp. 433–441, 2010.

[8] Y. Zhang, S. Wang, and L. Wu, "A novel method for magnetic resonance brain image classification based on adaptive chaotic pso," Progress In Electromagnetics Research, vol. 109, pp. 325–343, 2010.

[9] M. Saritha, K. P. Joseph, and A. T. Mathew, "Classification of mri brain images using combined wavelet entropy based spider web plots and probabilistic neural network," Pattern Recognition Letters, vol. 34, no. 16, pp. 2151–2156, 2013.

[10] G. Yang, Y. Zhang, J. Yang, G. Ji, Z. Dong, S. Wang, C. Feng, and Q. Wang, "Automated classification of brain images using wavelet-energy and biogeography-based optimization," Multimedia Tools and Applications, pp. 1–17, 2015.

[11] D. R. Nayak, R. Dash, and B. Majhi, "Least squares svm approach for abnormal brain detection in mri using multiresolution analysis," in Computing, Communication and Security (ICCCS), 2015 International Conference on. IEEE, 2015, pp. 1–6.