Human Brain Tumor Classification Using Genetic Algorithm Approach

E. S Benelin John[#1] and Mr. Ravikumar[*2]

[#]*UG scholar, Department of ECE, Vels Institute of Science, Technology and Advanced studies, Chennai, India.*
[*] *Assistant Professor, Department of ECE, Vels Institute of Science, Technology and Advanced studies, Chennai, India.*

Abstract— **Classification of brain tumor is a vital task in the healthcare applications. Brain tumor is a life threatening disease which develops improper cell growth of intracranial cavity in brain. Intelligent classification of the brain tumor has great potentiality in clinical medicine for detecting the abnormalities at an earlier stage. Magnetic Resonance Imaging (MRI) is widely used for capturing the abnormalities in brain. In this paper, we propose an enhanced brain MRI classifier targeting two main objectives, the first is to achieve maximum classification accuracy and secondly to minimize the number of features for classification. Gray Level Co-occurrence Matrix (GLCM) and statistical features are used for feature extraction. Two different machine learning algorithms are enhanced with a feature selection pre-processing step. Feature selection is performed using Genetic Algorithm (GA) while classifier used is Random Forest (RF) Classifier. Experimental results have shown the efficiency of the proposed algorithm.**

Index Terms— **Classification, Brain tumor, MRI, GLCM, Genetic Algorithm and Random Forest.**

I. INTRODUCTION

A brain tumor occupies space within the skull and can interfere with normal brain activity. It can increase pressure in the brain, shift the brain or push it against the skull and damage nerves and healthy brain tissue. Brain tumors are abnormal masses in or on the brain. When most normal cells grow old or get damaged, they die, and new cells take their place. Sometimes, this process goes wrong. New cells form when the body doesn't need them, and old or damaged cells don't die as they should. The formation of extra cells creates a mass of tissue called tumor. Tumor growth may appear as a result of failure of the normal pattern of cell death [1].Brain tumors may have a variety of symptoms ranging from headache to stroke. Different parts of the brain control different functions, so symptoms vary depending on the tumor's location [2]. The function of clustering is to identify the image areas that can have maximum chances of tumor. In this paper, fuzzy C means clustering is used. The genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection and crossover [3].

Most of the tumor is two types namely benign and malignant. Malignant tumor is referred to as cancer. Abnormal growth of cell inside brain is called brain tumor. There are two general groups of brain tumor. Primary brain tumor starts in brain and tends to stay there. Secondary brain tumor starts somewhere else in the body but travels to brain. Secondary tumors are more common than primary tumors. The reason for brain tumor is unknown till now. It is considered that probable reasons of brain tumor can be a number of conditions like neurofibromatosis [4], exposure to chemical vinyl chloride, Epstein-Barr virus and ionizing radiation. The use of mobile phones is also considered as one of the risk factors but there is still no clear evidence. Meningioma (usually benign), Oligodendrogliomas and astrocytoma such as Glioblastomas are primary tumor commonly found in adults and Medulloblastoma in children. Diagnosis is usually done by medical Examination along with MRI. Biopsy is then conducted for confirmation.

Tumors are divided into different grades of severity depending on the report obtained from diagnosis. In grade 1, the cells look normal and grow slowly. The cells look slightly abnormal and grow slowly in grade 2. Cell starts to grow actively and look abnormal. They start to grow actively in nearby brain tissue and tend to reoccur. This happens only when it is grade 3 tumor. In grade 4 tumor are most abnormal and fast spreading. There are various literatures available on the topic of brain tumor detection and extraction on MRI images of brain. In recently, a two methods for segmentation, i.e. ACO hybrid with Fuzzy and HSOM hybrid with fuzzy is used to detect brain tumor. Though the detection is done, still the noise is remaining in the image. And another method named as Ridler's method, is used to low pass filtering morphological operation and thresholding and lastly region growing methods to extract the brain tumor. But the size of the tumor is not accurate as that were present in original pictures. And also another method a multi-modality framework and AdaBoost classifier is used to detect the tumor [5]. Even if the tumor is detected, it still has noise in it and the accuracy of the detected tumor is not good. It is not like that present in the ground image.

The rest of the paper is organized as follows: Section II presents the related work; Section III presents the proposed work; Section IV presents the experimental analysis and finally concludes in Section V.

II. RELATED WORK

This section presents the prior works suggested on brain tumor classification. In [6] focused on survey of well-known brain tumor detection algorithms that have been proposed so far to detect the location of the tumor. In [7], a clustering based approach using a Self Organizing Map (SOM) algorithm for medical image segmentation. They presented a new unsupervised MR image segmentation method based on fuzzy C-Means clustering algorithm for the Segmentation. In [8] showed that Watershed Segmentation can successfully segment a tumor provided the parameters are set properly in MATLAB environment. Their paper explores a method to identify tumor in brain disorder diagnosis in MR images.

In [9], the images are segmented a set of MR images using standard image segmentation techniques to isolate a brain tumor from the other regions of the brain. Subsequently the resultant images from the different segmentation techniques were compared with each other and analyzed by professional radiologists to find the segmentation technique which is the most accurate. Experimental results showed that the Otsu's thresholding method is the most suitable image segmentation method to segment a brain tumor from a Magnetic Resonance Image. In [10], the author inferred that the brain tumor classification and segmentation is best done using SVM with dominant run length feature extraction method than SVM with wavelet based texture feature extraction method and SVM with SGLDM method. In their work, they attempted to improve the computing efficiency as it selects the most suitable feature extraction method that can use for classification and segmentation of brain tumor in CT images efficiently and accurately.

The author in [11] studied a Grey level Co-occurrence Matrix (GLCM) for texture feature extraction, ANFIS (Adaptive Network Fuzzy inference System) plus Genetic Algorithm for feature selection and FCM (Fuzzy C-Means) for segmentation of Astrocytoma (Brain Tumor) with all four Grades. The comparative study between FCM, FCM plus K-mean, Genetic Algorithm, ANFIS and proposed technique showed improved Accuracy, Sensitivity and Specificity. The author in [12] suggested a new method combining genetic algorithm and K-Means algorithm for clustering medical images. In this combined technique, variable string length genetic algorithm (VGA) is used for the determination of the optimal cluster centres.

In [13], the authors explored Extreme Learning Machine Classifier algorithm for implementing genetic algorithm. This method is modeled for automatic brain tissue and pathological tumor classification and segmentation of 3D MRI tumor images. Then a genetic algorithm for automating the segmentation of the prostate on two-dimensional slices of pelvic computed tomography (CT) images. In this approach they represented segmenting curve using a level set function, which is evolved using a genetic algorithm (GA).

In [14], the authors suggested an algorithm which is a combination of SVM and fuzzy c-means, a hybrid technique for prediction of brain tumor. Here, the image is enhanced using contrast improvement, and mid-range stretch. Double thresholding and morphological operations are used for skull striping. Fuzzy c-means (FCM) clustering is used for the image segmentation. Grey level run length matrix (GLRLM) is used for extraction of feature. Then, Linear, Quadratic and Polynomial SVM technique is applied to classify the brain MRI images. Real data set of 120 patients MRI brain images have been used to detect 'tumor' and 'non-tumor' MRI images. The SVM classifier is trained using 96 brain MRI images, after that the remaining 24 brain MRI images was used for testing the trained SVM. SVM classifier with Linear, Quadratic and Polynomial kernel function give 91.66%, 83.33% and 87.50% accuracy respectively and 100% specificity.

In [15], the authors studied an effective automatic classification method for brain MRI is projected using the Adaboost machine learning algorithm. The proposed system consists of three parts such as Preprocessing, Feature extraction and Classification. Preprocessing has removed noise in the raw data, it transform RGB image into grayscale, median filter and thresholding segmentation is applied. For feature extraction by using GLCM technique 22 features were extracted from an MRI. For classification boosting technique used (Adaboost). It gives 89.90% accuracy and result in normal brain or in Malignant or Benign type of tumor. The authors in [17] proposed a novel technique which includes Normalization of Histogram and K-means Segmentation. First, input image is pre-processed in order to remove the unwanted signals or noise from it. To de-noise filters such as Median filter, Adaptive filter, Averaging filter, Un-sharp masking filter and Gaussian filter is used in the MRI images. The histogram of the pre-processed image is normalized and classification of MRI is done. Finally, the image is segmented using K-means algorithm [18] in order to take out the tumor from the MRI. Efficient classification of the MRIs is done using NB Classifier and SVM so as to provide accurate prediction and classification. The authors in [19] studied a model in which Image is processed through: Preprocessing, Segmentation, Feature extraction Classification stages. In preprocessing, Morphology technique using double thresholding is applied to remove the skull out of the MRI brain images. The present work presents the comparison study of two techniques used for tumor detection of MRI images. One is based on the Level set method that uses the non parametric deformable models with active contour to segment the brain tumor from the MRI brain images. The other one is the K-means segmentation algorithm.

III. PROPOSED WORK

This section presents the proposed method for detecting the brain tumor classification. The proposed model composes of Gray Level Correlation Matrix (GLCM), Statistical features and Genetic Algorithms.

A. Gray level Correlation Matrix (GLCM)

A co-occurrence matrix or co-occurrence distribution is a matrix that is defined over an image to be the distribution of co-occurring pixel values (grayscale values, or colors). 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, i.e., the spatial relationships of pixels in an image.) After you create the GLCMs, you can derive several statistics from them using the graycoprops function. These statistics provide information about the texture of an image.

B. Statistical features

A statistics use is motivated by the fact that the distribution of samples in the data is characterized by their asymmetry, concentration and dispersion around the mean. Interestingly, the differences are correctly captured using the statistics of the IMFs. Using the following quantities these statistics are obtained for an IMF.

  i) Mean: The arithmetic mean or average of a set of values is the ratio of the sum of these values to the number of elements in the set. It is given as follows:

$$\mu_t = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad (3.1)$$

  ii) Standard Deviation:

The Standard Deviation is a measure of how spread out numbers is. It is given as follows:

$$\sigma_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \mu_t)^2} \qquad (3.2)$$

C. Genetic algorithm

In computer science and operations research, a genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms (EA). Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on bio-inspired operators such as mutation, crossover and selection. A typical genetic algorithm requires:

  • A genetic representation of the solution domain,
  • A fitness function to evaluate the solution domain.

During each successive generation, a portion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where fitter solutions (as measured by a fitness function) are typically more likely to be selected. Certain selection methods rate the fitness of each solution and preferentially select the best solutions. Other methods rate only a random sample of the population, as the former process may be very time-consuming. The fitness function is defined over the genetic representation and measures the quality of the represented solution. The fitness function is always problem dependent. For instance, in the knapsack problem one wants to maximize the total value of objects that can be put in a knapsack of some fixed capacity. A representation of a solution might be an array of bits, where each bit represents a different object, and the value of the bit (0 or 1) represents whether or not the object is in the knapsack. Not every such representation is valid, as the size of objects may exceed the capacity of the knapsack. The fitness of the solution is the sum of values of all objects in the knapsack if the representation is valid or 0 otherwise. In some problems, it is hard or even impossible to define the fitness expression; in these cases, a simulation may be used to determine the fitness function value of a phenotype (e.g. computational fluid dynamics is used to determine the air resistance of a vehicle whose shape is encoded as the phenotype), or even interactive genetic algorithms are used.

D. Random Forest Classifier

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. Decision trees are a popular method for various machine learning tasks. Tree learning "come[s] closest to meeting the requirements for serving as an off-the-shelf procedure for data mining", say Hastie et al., because it is invariant under scaling and various other transformations of feature values, is robust to inclusion of irrelevant features, and produces inspect able models. However, they are seldom accurate.
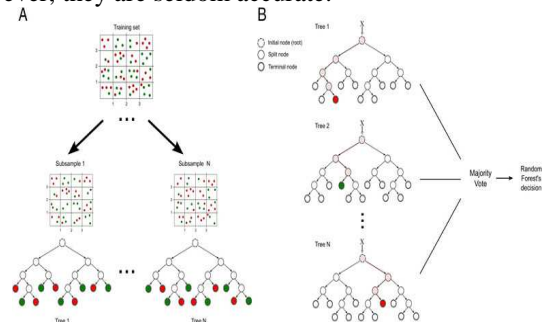

Fig.1. Random Forest classifier- Working

In particular, trees that are grown very deep tend to learn highly irregular patterns: they over fit their training sets, i.e. has low bias, but very high variance. Random forests are a way of averaging multiple deep decision trees, trained on different parts of the same training set, with the goal of reducing the variance. This comes at the expense of a small increase in the bias and some loss of interpretability, but generally greatly boosts the performance of the final model.
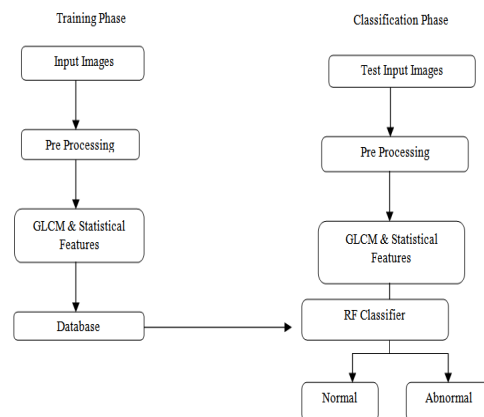

Fig.2 Proposed architecture

IV. EXPERIMENTAL RESULTS

This section presents the experimental analysis of our proposed algorithms.
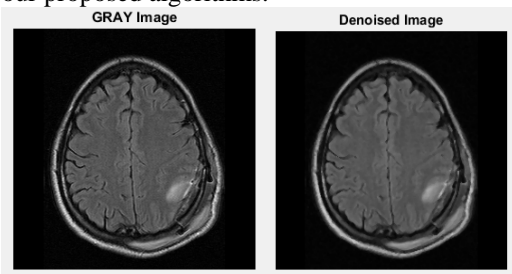

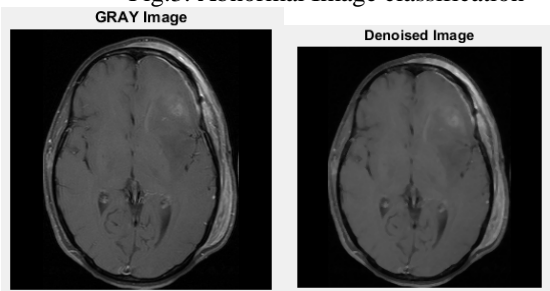Fig.3. Abnormal Image classification
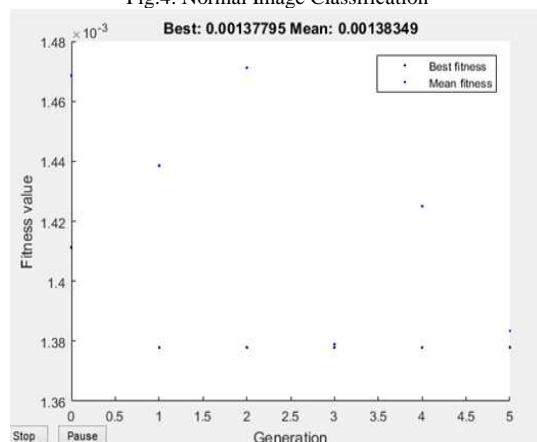

Fig.4. Normal Image Classification

Fig.5. Genetic algorithm for Image classification

V. CONCLUSION

Since brain diseases are dynamic and evolutionary in nature, their detection, treatment will also progress based on the dynamic nature of the disease. The treatment detection technique in brain tumor keeps on changing with time. Therefore, the image processing technique must also progress in a direction of finding tumor as early as possible. The screening techniques in tumor detection must be reliable, robust and must have high level of diagnostic value. For this purpose, right from the image acquisition to the detection of tumor, an effective work is defined enough to identify the real indicators of tumor nodules. In this paper, we propose an enhanced genetic algorithm for classifying the brain abnormalities. Initially, brain MRI images are collected and then denoised for further evaluation. The objective of the proposed work is to achieve better classification accuracy with minimized features. The relevant features are extracted using GLCM and the statistical models. The extracted features are then selected using genetic algorithm and then further classified using Random Forest (RF) classifier. Experimental analysis has shown the efficiency of the proposed algorithm.

REFERENCES

[1] Nayak, D. R., Dash, R., & Majhi, B. Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests. Neurocomputing, 2016, 177, pp. 188-197.

[2] Gupta, M., Rao, B. P., & Rajagopalan, V. Brain tumor detection in conventional MR images based on statistical texture and morphological features. IEEE International Conference on Information Technology (ICIT), 2016, pp. 129-133.

[3] Mohan kumar, S. Analysis of different wavelets for brain image classification using support vector machine. International journal of advances in signal and image sciences, 2016, 2(1), pp. 1-4.

[4] Singh, G., & Ansari, M. A. Efficient detection of brain tumor from MRIs using K-means segmentation and normalized histogram. IEEE 1st India International Conference on Information Processing (IICIP), 2016, pp. 1-6.

[5] Bhuvaneswari, K. S., & Geetha, P. Semantic feature based classification of Brain MRI using PCA and PNN. IEEE International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016, pp. 2700-2706.

[6] Anitha, V., & Murugavalli, S. Brain tumour classification using two-tier classifier with adaptive segmentation technique. IET computer vision, 2016, 10(1), pp. 9-17.

[7] Priya, K. M., Kavitha, S., & Bharathi, B. Brain tumor types and grades classification based on statistical feature set using support vector machine. IEEE 10th International Conference on Intelligent Systems and Control (ISCO), 2016, pp. 1-8.

[8] Moeskops, P., Viergever, M. A., Mendrik, A. M., de Vries, L. S., Benders, M. J., & Isgum, I. Automatic segmentation of MR brain images with a convolutional neural network. IEEE transactions on medical imaging, 2016, 35(5), pp. 1252-1261.

[9] Telrandhe, S. R., Pimpalkar, A., & Kendhe, A. Detection of brain tumor from MRI images by using segmentation & SVM. IEEE World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), 2016, pp. 1-6.

[10] Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R., and Davatzikos, C. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magnetic resonance in medicine, 2009, 62(6), pp. 1609-1618.

[11] Qurat-Ul-Ain, L.G., Kazmi, S.B., Jaffar, M.A., Mirza, A.M.: 'Classification and segmentation of brain tumor using texture analysis', Recent Adv. Artif. Intell. Knowl. Eng. Data Bases, 2010, 10, pp. 147–155

[12] Khalid, N.E.A., Ibrahim, S., Haniff, P.N.M.M.: 'MRI brain abnormalities segmentation using K-nearest neighbors (k-NN)', Int. J. Comput. Sci. Eng. (IJCSE), 2011, 3, (2), pp. 980–990

[13] Aslam, H.A., Ramashri, T., Ahsan, M.I.A.: 'A new approach to imagesegmentation for brain tumor detection using pillar K-means algorithm', Int. J. Adv. Res. Comput. Commun. Eng., 2013, 2, (3), pp. 1429–1436

[14] Ricci, P.E., Dungan, D.H.: 'Imaging of low and intermediate-grade gliomas', Semradonc, 2001, 11, (2), pp. 103–112

[15] Rajini, N.H., Narmatha, T., Bhavani, R.: 'Automatic classification of MR brain tumor images using decision tree'. Special Issue of Int. J. of Computer Applications on Int. Conf. on Electronics, Communication and Information Systems (ICECI 12), 2012, pp. 10–13.