International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 14 (2018) pp. 11686-11692

Automatic Segmentation and Classification of Brain Tumor using Machine Learning Techniques

Himaja Byale[1], Dr Lingaraju G M[2] and Shekar Sivasubramanian[3]

[1]M.Tech, Information Science and Engineering, Ramaiah Institute of Technology, MSRIT Post Road, M S Ramaiah Nagar, MSR Nagar, Bengaluru, Karnataka 560054

[2] Information Science and Engineering, Ramaiah Institute of Technology, MSRIT Post Road, M S Ramaiah Nagar, MSR Nagar, Bengaluru, Karnataka 560054

[3]Researcher, Research in Knowledge Management, Information Retrieval and Machine Learning Carnegie Mellon University (CMU), Language Technologies Institute, Pittsburgh, USA.

Abstract

In the field of Medicine, especially in the applications that are used for diagnosis purpose, detecting a fault is a major task and needs lots of attention of the radiologist. Early defect detection is necessary to avoid the further complications. The most actively developing field in the recent technology is the MRI scanning. The size of the tumor in brain can vary differently for different patients along with the minute details of the tumor, It is a complicated and tedious job for the radiologists to diagnose and classify the tumor from a large number of images. Sometimes, cerebral fluid also seems to be seen as a mass of tissue in the MRI image. The project aims to have an automated system which plays a important role in assessing whether a lump (mass of tissue) in the brain could be benign (clump thickness) or malignant (marginal adhesion) by classification. The proposed model uses machine learning algorithms in order to improve the accuracy of classification. The system is carried out in four steps that includes pre-processing for noise removal using adaptive median filter, segmentation using Gaussian Mixture Model (GMM) for finding the region of interest, feature extraction using Grey Level Co-occurrence Matrix GLCM for extracting the features of different types of tumors and classification using Neural Networks (NN) to determine and classify the tumor as benign or malignant.

The experimental results of the proposed model shows that 93.33% accuracy, 96.6% specificity, 93.33% sensitivity and precision with 94.44%. From these results the proposed model works better when compared with the classical machine learning algorithms like Adaboost (Adaptive Boosting) which classifies the image into different classes (Normal, Benign, Malignant) with 89.90% accuracy.

Keywords: Adaptive Median Filter, K-means, GMM Segmentation, GLCM, Neural Networks,

INTRODUCTION

Brain tumor is known to cause major backdrop for the rapid increase in the mortality among the children, adult and especially in the old aged people. As human body is made of millions of cells, also from biology we understand that the cells multiply, grow and divide to form new cells and tissues. Due to some external factors the cells might grow uncontrollably resulting in the tumor formation. Tumors are of two types, Benign tumors are the cancerous cells and are less harmful as they do not spread across the other cells. Whereas malignant tumors are those mass of cells which are cancerous, harmful and are more likely to spread across other cells and tissues. Tumor is basically a gathering of cells forming a tissue which does not the have control as of healthy cells and has unconditional growth rate. In past decade, according to the statistics in most of the developed countries shows that there are over 300 people who die because of developing brain tumor, the count is perhaps increasing year by year. Research about the international statistics say that in the U.S the increasing growth rate of tumor probability in a particular amount of time has 11% – 12% people suffering from the cancer every year. Thus all these scenarios lead us in developing the detection model for the brain tumor. It is very important to diagnose the presence of tumor in brain as early as possible. MRI technique is known for the clarity of images it can scan. In MRI the appearance of the tumor is very precise and high, for further treatment and medications the help of the physicians is also needed. Most of the time the medical diagnosis is done using the MRI scanning as it produces better results. Thus MRI is gaining a significance attention and has lot of scope in the future. With the advances of computational intelligence and machine learning techniques, detection of tumors in the early stage is possible. The image (fig 1) is taken from the hospital which shows the MRI image of the brain.

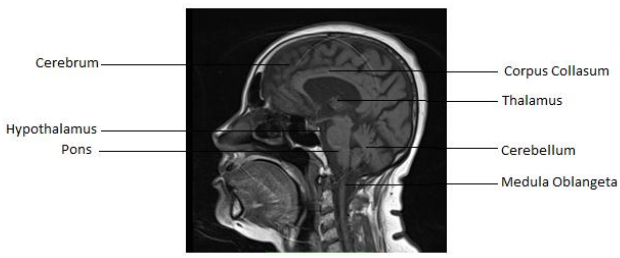

Figure 1. Brain Image obtained from MRI scanning

LITERATURE REVIEW

Astina Minz [1] proposed a model in which the image processing and machine learning techniques were used for detection of the tumor, Median filters are used for cleaning the images, by using the thresholding technique segmentation of the brain is done, GLCM technique is used for extracting the features, and finally adaboost technique is used for the classification of the tumor. Their results show 89.90% accuracy in classifying the tumor.

The artificial neural networks in classifying the tumors into normal or astrocytoma type of tumor in the MRI images of various patients gains a lot of attention. The paper [2] employs GLCM as the feature extraction and trained the neural networks. Their results show that few more diverse have to be considered for having a efficient classification of the tumor into pilocytic (grade1), low grade (grade2), anaplastic (grade3) and glioblastoma multiforme (grade4)

The proposed system is based on the GLCM feature extraction and applied feed forward neural networks on the MRI and CT images, their results show that the proposed system is 97% accurate. In this project [3] the pre processing is done to make the data noise free, no segmentation is carried out and hence the entire pre processed image has been used for the feature extraction.

Artificial intelligence techniques like, neural networks and fuzzy logic neural networks which has great impact in this work [5]. The feature extraction is done using PCA and PNN. The conclusions say that the the performace of PNN is quite efficient and prominent for the diagnosing the tumor. The accuracy of the system is between 73-80% depending upon the training data.

The model that predicts the tumor into malignant or benign based upon the features that is fed while training the algorithm, the techniques used for the prediction in the proposed model are the supervised learning techniques which comprises of Linear Regression and Logistic Regression [6]. These algorithms predict which type a tumor belongs to based upon the features observation. The two features considered for the classification are clump thickness and marginal adhesion. The proposed system uses the sigmoid and cost function to minimize the sum of errors. Gradient decendent function is used to find the global minimal of the data and then classify the tumor.

The classification of the tumor form the MRI images using machine learning concepts gains a lot of attention. The techniques used in the proposed model [7] are the discrete cosine transformation which is used for dimensionality reduction of the images and cleaning the image to have noise free and probabilistic neural networks for classifying the tumor of its respective type. The network is trained for 20 data samples of the brain tumor and has 100% accuracy in the classification and has significantly less computational time.

The brain tumor detection model using the MRI images. This system revolves around the multi-model framework for detecting the presence of tumor in the brain automatically. In this system different MRI modalities are used training and testing the system [8]. For each MRI image features like the intensity, shape, deformation, symmetry and texture. These features are fed to the Adaboost classifier. The simulated results show that the system show the accuracy of 90.11%. The author explains that no classification model would provide more accurate classification unless the images are processed using some techniques for the noise removal and segment the ROI out of the image to have more accurate results. In this [9] system median filter, max filter, min filters are used to maintain the clarity of the image. For the segmentation two techniques are used, they are thresholding and watershed for extracting the region of interest.

The classification model that classifies the tumor into different types using machine learning techniques. The proposed model [10] uses different computational strategies like feedback pulse coupled neural network for segmenting the tumor from the brain, discrete wavelength transform for feature extraction and PCA for reducing the dimensions and for classifying the image into type1 or type2 tumor feed forward and back propagation neural networks are used. The results shows that the system is able to classify the tumors with a accuracy of 99% which is significantly good.

PROPOSED METHODOLOGY

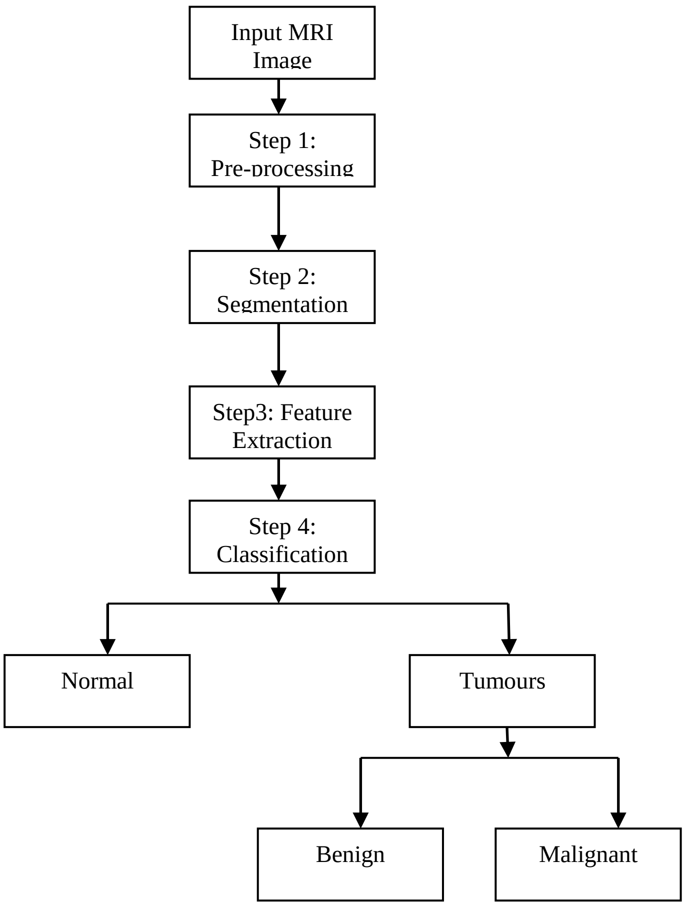

Figure 2

Input Brain MRI image

Image Acquisition is the first and foremost step of the process as shown in the fig 2, in which the MR images of the brain having tumor and normal brain images are taken. The images used in this project are acquired form MS Ramaiah Memorial Hospital Bangalore. Over 60 samples of the MR images have been collected out of which 30 images are used of training and 30 images are used of testing purpose. The below figures(3) shows the MRI images in dicom format.

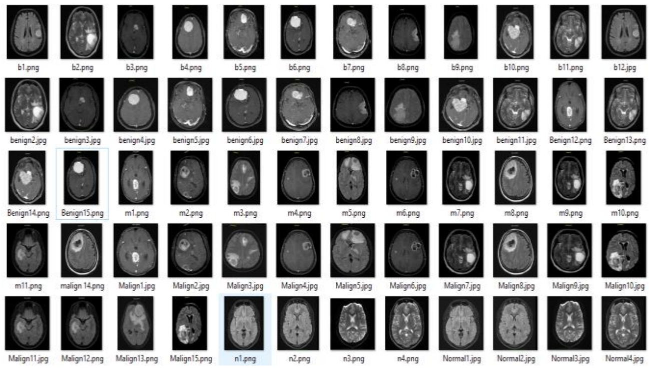

Figure 3

Step1: Pre-Processing

It is always necessary to have improved image quality to achieve better results in the preceding steps. Pre-processing of the image plays an important role in delivering the improved features of the image. In this proposed system to improvise the clarity of the pictures of MRI, efficient adaptive median filter technique is used. This is a very efficient way of filtering the impulse data from the images. On the other hand, in the conventional median filter method, all the pixels of the image are filtered equally i.e. both noisy and noiseless pixels of the image. Thus the pre processed image will result in having faded corners, eliminated edges and blur image. Hence many variations have been introduced [11].

One such variation is the adaptive median filter, in this method the mechanism that is used is the filter which is bigger in size is applied to the area of pixels having higher noisy resolution and the filter in a smaller size is applied to the pixels having less noise in the input image. The research [15] used this method by considering the values of min, max and med intensity using the window as the measure for size. Adaptive median filter consists of two stages i.e. in the first stage the filter gets adaptive to the image by differentiating noisy and noise-free pixels class. In the next stage only noisy pixels are considered and taken into the consideration by applying adaptive median filter to it. In the further process no action is performed on the noise less pixels and are copied the same to the processed image.

In the approach, adaptive median filer has been used - the results of pre-processing using median and adaptive median filter are as shown in the fig (5,6). As seen in the image (fig 5), adaptive median filter is able to preserve the fine lines, edges while the median filter (fig 4) has faded the lines in the image and causing blurriness

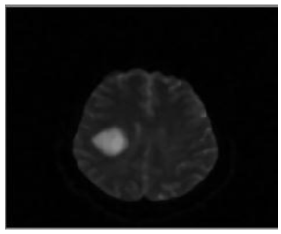

Figure 5

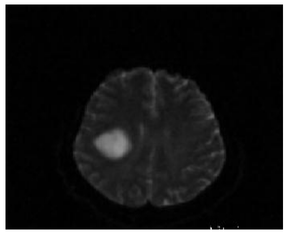

Figure 6

Processes using median filter processes using adaptive median

Step 2: Segmentation of ROI

Segmentation plays an important role in segmenting the image into different parts and only the useful part can then be considered. This process has to be carried out in order to reduce the amount of work that is to be done in the further steps i.e. feature extraction becomes easier when there is only tumor region whose features need to be extracted over the entire image.

K-means is a iterative unsupervised clustering technique. Every cluster is differentiated by its centroid which is placed randomly, by k-means it finds a local minimum of the cost function and converges. To find the distance between the centroid and the data points the Euclidean distance is the measure used. K-means is hard clustering technique which means the data point can either belong to any one of the cluster. Here assigning the data point to the respective cluster is based on the only the distance between the data point and centroid which is chosen randomly.

The proposed system uses Gaussian mixture model (GMM) segmentation technique [12] for segmenting the tumor region out of the brain from the image. GMM is the advanced technique over K-means algorithm. In GMM it is a soft clustering technique it means that the data point can participate in both the clusters depending upon the probability which is calculated by considering mean along with the co-variance. GMM further uses Expectation Maximization (EM) technique for clustering the brain into different regions and thus helps in reducing the data set. The results of K-means and GMM techniques are as in the figure (7,8). The fig (8) shows that only the most brighter pixels are clustered and the tumor part is extracted out, while the fig (7) shows that some other parts apart from the tumor is also segmented. Since only the tumor is the region of interest (ROI), GMM does better work

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 14 (2018) pp. 11686-11692

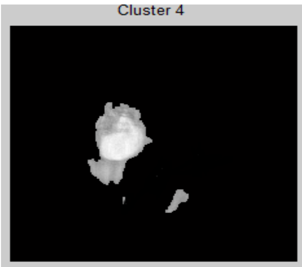

Figure 7

K-means Segmentation

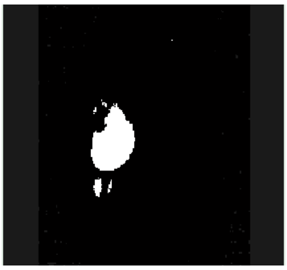

Figure 8

GMM Segmentation

Step 3: Feature Extraction

Feature Extraction is necessary as it reduces the representation of large data set. It is way of extracting the features from the image in the form of vector. The features extracted are those that differentiate the characteristics among the different class of tumor it belongs to. Transforming the data(image) into the useful features that are to be considered for the classification is called feature extraction. The proposed model uses Grey Level Co-Occurrence Matrix (GLCM) technique for the feature extraction [2]. This technique is applied on the segmented tumor region of the MR image thus extracting the textural properties of the tumor. These features are extracted separately for different types of tumor (benign and malignant) and normal brain. Finding the textural properties also plays important role in extracting the surface structural arrangements of the tumor.

GLCM is the statistical method of extracting the features [3]. To find the GLCM, matlab provides the function of determining the GLCM matrix i.e. by using graycomatrix function. Once the GLCM matrix is obtained using the above function, several statistics can be derived from the matrix.

Illustration of GLCM:

GLCM matrix [14] is obtained using the graycomatrix function. The function further determines how much frequently the pixel with value i, appears to be in the same spatial area as of the pixel with value j. To calculate the matrix of GLCM, the function auto identifies the pixel that has to be processed along with the pixels that is horizontally and is adjacent right to it. Matlab also provides the feasibility for the user to define the pixels in spatial relationship. Every pixel in the resultant graycomatrix has the value i,j , which is nothing but the number that represent how often the pixels with the same intensity value occurred in the input matrix. In this way finding out the graycomatrix helps in revealing the textural properties of the image.

The function is applied on the input image which converts into matrix (initial matrix). From the fig (9) the resultant matrix of graycomatrix function, the value in [1,1] is 1 since there is only one such instance in the input matrix where two pixels that are horizontal has 1 as their values respectively. The default processing happens in a horizontal way, but it can also be customized to the directions needed. The next value in resultant matrix i.e. [1,2] contains the value 2 because in the input matrix there happens to be two such instance where the values (1,2) occurs in horizontal manner. Similarly, the third value i.e. [1,3] the value is 0 since the there is no [1,3] instance of values in the input matrix in horizontal manner. This process is continued until all the pixels of the input matrix undergo this function and forming the graycomatrix at the end of the process.

Process Used to Create the GLCM

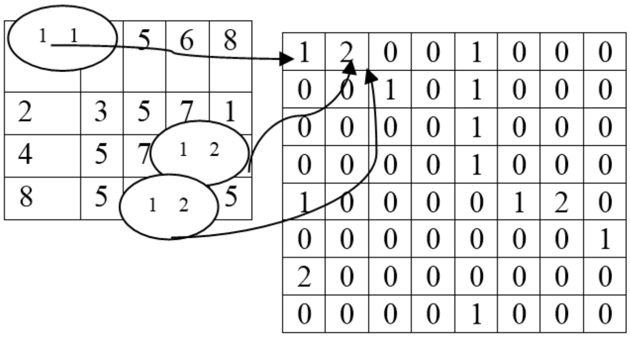

Figure 9. Process used to create GLCM matrix

There are totally 21 statistical features that are derived in this proposed system. Few of them are shown in the below table 1

Table 1. Features extracted for a given input MR image

Statistics	Description
Contrast	Determines the difference in the intensity value between the adjacent pixels of the resultant graycomatrix
Correlation	It is one of the statistic measure used to determine how mutual and strong a relationship is between the pixels of the matrix
Energy	It identifies the sum of squared elements if the graycomatrix which is known to be uniformity.
Homogeneity	Determines the quality, how much close the pixels of interest and adjacent are relatable to each other in the graycomatrix.

Step 4: Classification

In the classification process where the brain MRI image has to be identified as normal, benign or malignant, it analyses the statistical properties of the input image and systematically group the data into various categories. In the proposed system Neural network is used for training the data (trained data) and testing data (test data) thus it is 2 phase model. In the first phase features of the image are isolated and a distinct description of each classification category is created. In the second phase, test data is taken and validated against the system. Hence verifying if the images are correctly classified.

The concept of Machine Learning Techniques in the proposed system revolves around the concept of differentiating the MRI scanned images of brain into either normal, benign or malignant. The main goal of the machine learning is to auto learn from the training and finally make a clever decision with high accuracy. Neural Networks are used for classification purpose. Neural networks are the supervised learning algorithms which uses features of the images that are extracted in the feature extraction process using GLCM. These features acts as the input to the neural network. Neural networks [5] are used to learn and based upon the learning they classify the image into tumors or non-tumours category by detecting the presence of tumor in the brain MR image. In the proposed model the input layer of the network consists of 19 neurons representing 19 statistical features which are extracted. The output layer consists of 3 neurons indicating Normal, Benign, Malignant. The hidden layer changes according to the rules that give the best recognition rate for each group of features.

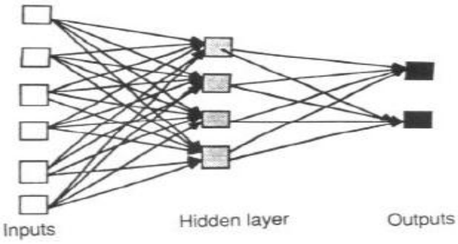

Figure 10. Example of neural network with single hidden layer

A typical neural network with a single hidden layer is as shown in the figure (10). Generally in neural networks the output of their respective layers are given as input to the next layers using the activation functions. These activation function are used to scale output of the neural network into correct ranges with better accuracy. In the proposed model there are 3 activation functions that are used:

1. Sigmoid Function
$$f(x) = \frac{1}{1 + e^{-x}}$$
2. Hyperbolic Tangent
$$f(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$
3. Linear Function
$$f(x) = x$$

EXPERIMENTAL RESULTS AND CONCLUSION

Table 2. Comparison of Adaboost and Neural Network

ML Algorithm	Total Samples	Accuracy	Sensitivity	Specificity
Neural Network	60	93.3%	93.33%	96.6%
Adaboost	50	89.90%	88.23%	62.5%

In the proposed model, an automatic recognition and classification of the tumor is developed in order to classify the input MR image into normal, benign and malignant. The classification is carried out using the features that are extracted using GLCM algorithm. Further the image is classified using Neural Network. The entire proposed model is implemented using different algorithms in each process. Adaptive Median filter is used for pre-processing the image which yielded better results when compared to median filter, the fig (5,6) shows the experimental results of comparison. GMM technique is used for segmentation purpose, GMM is an advanced method over K-means clustering. The comparison of both the techniques are as shown in the figure(7,8). For the feature extraction GLCM technique is used. Finally, for the classification neural networks are used for classifying the image into appropriate class.

The results as shown in the table 2, the proposed model shows that 93.33% accuracy, 96.6% specificity, 93.33% sensitivity and precision with 94.44%. From these results the proposed model works better when compared with the classical machine learning algorithms like adaboost which classifies the image into different classes (Normal, Benign, Malignant) with 89.90% accuracy (Table 2).

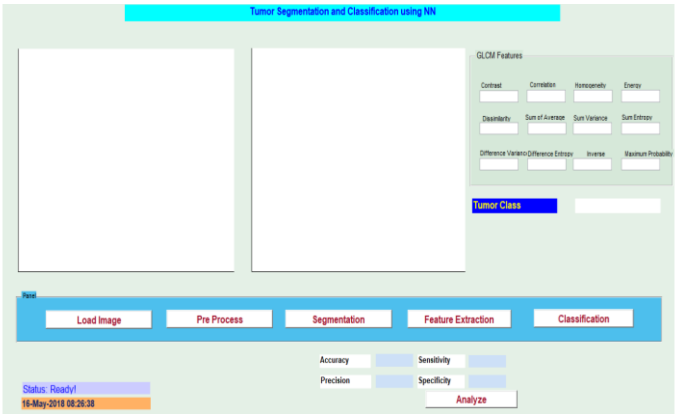

Figure 11. The GUI layout of the proposed system

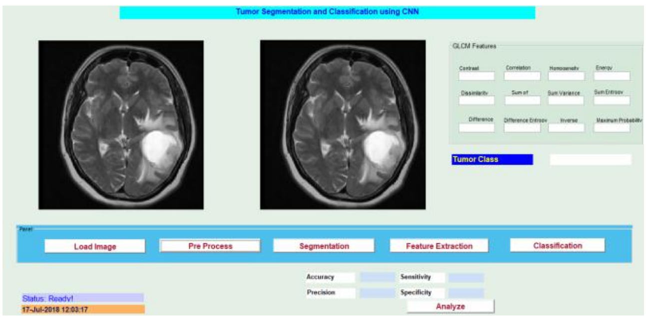

Figure 12. Brain MRI is loaded and Preprocessed

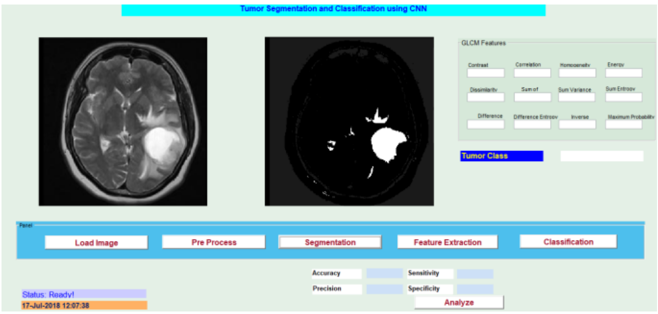

Figure 13. Tumor segmentation using GMM

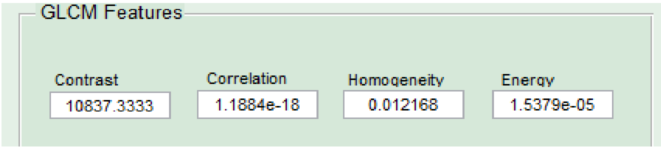

Figure 14. Statistical features extraction using GLCM

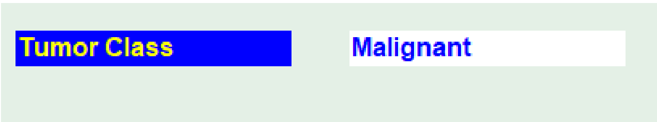

Figure 15. Decision of the neural networks.

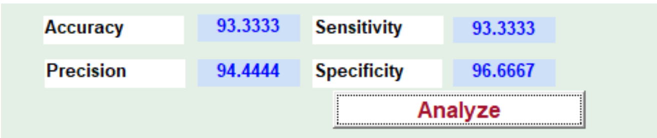

Figure 16. Analysis

FUTURE SCOPE

Future scope of this project is to train this algorithm on large dataset and also patient level accuracy on different dataset can be performed. For reducing the execution time graphical processing unit (GPU) can be incorporated in NN processing. The approach used in project will provide baseline for further researches in brain cancer diagnosis using deep learning algorithms. This system shall be extended in identifying different classes (Glioma, Meningioma) within Benign and Malignant type of cancer. The proposed method can also be explored in other medical imaging diagnosis such as Lung cancer, Breast cancer and colon Detection. Different variations of Neural networks can also be applied for the diagnosis of the brain tumor. The accurate detection of pre-cancer growth using automated tools will help the patient to get appropriate treatment well within time, as most of cancer is curable only if it is detected in early stages. Once the system achieves 100% accuracy in classifying the image, then this application can be deployed for the betterment of the hospitals.

ACKNOWLEDGEMENT

The successful completion of the project would be incomplete without the mention of the people who have constantly guided me and supported the completion of the task. First of all, special thanks to my guide **Dr Lingaraju G M,** Professor, Dept. of ISE and **Shekar Shivasubramaniyan** who guided me well throughout and focused on betterment of the project, their supervision, support and vital contribution from an early stage of this research and providing me extraordinary experiences throughout the work. With due support from **Dr. Vijaya Kumar B P** Professor & Head of Dept. of ISE I thank him for his valuable encouragement and suggestions throughout the work. I would like to thank whole heartedly **Dr. N V R Naidu**, Principal, for the kind support and permission to use the facilities available in the institute. I am highly grateful to the faculties, library staff and laboratory staff of ISE department for their extended long-term support and coordination. This project would never have been completed without the unending support form M S Ramaiah Memorial Hospital Bangalore. I would like to extend my gratitude towards **Dr Narresh Shetty** the president of M S Ramaiah Memorial Hospital Bangalore, **Dr Ashok** HOD of Radiology Department, **Dr Akshay Sethi** and **Dr Dennis Fernando** for providing the valuable dataset for the completion of the project successfully.

REFERENCES

[1] Astina Minz, “MR Image classification using adaboost for brain tumor type”, 2017 IEEE 7th International Advance Computing Conference (IACC)

[2] Shweta Jain, “Brain Cancer Classification Using GLCM Based Feature Extraction in Artificial Neural Network”, International Journal of Computer Science & Engineering Technology (IJCSET)

[3] Nitish Zulpe and Vrushsen Pawar, “GLCM Textural Features for Brain Tumor Classification”, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012

[4] Daljit Singh, Kamaljeet Kaur, “Classification of Abnormalities in Brain MRI Images using GLCM, PCA and SVM”, International Journal of Engineering and Advanced Technology (IJEAT)

[5] Mohd Fauzi Othman and Mohd Ariffanan Mohd Basri, “Probabilistic Neural Network for Brain Tumor Classification”, 2011 Second International Conference on Intelligent Systems, Modelling and Simulation

[6] Ashish Shah, “ Prediction of Malignant and Benign Tumor using Machine Learning”, International Journal of Computer Applications (0975 – 8887) Volume 135 – No.5, February 2016.

[7] D.SRIDHAR, MURALI KRISHNA, “Brain Tumor Classification Using Discrete Cosine Transform and Probabilistic Neural Networl”, 2013 International

Conference on Signal Processing, Image Processing and Pattern Recognition [ICSIPRJ]

[8] Sahar Ghanavati, Junning Li, "Automatic Brain Tumor Detection in Magnetic Resonance Images", 2012 IEEE

[9] HebaMohsen, EI-Sayed Ahmed EI-Dahshan, "A Machine Learning Technique for MRI Brain Images", The 8th International Conference on INFOrmatics and Systems (INFOS2012) - 14-16 May

[10] Kalpana U. Rathod, "MATLAB Based Brain Tumour Extraction Using Artificial Neural Network", International Journal on Recent and Innovation Trends in Computing and Communication

[11] Haidi Ibrahim, "Simple Adaptive Median Filter for the Removal of Impulse Noise from Highly Corrupted Images", IEEE Transactions on Consumer Electronics, Vol. 54, No. 4, NOVEMBER 2008

[12] U. Baid, S. Talbar et al, " Comparative Study of K-means, Gaussian Mixture Model, Fuzzy C-means algorithms for Brain Tumor Segmentation", ICCASP/ICMMD-2016. Advances in Intelligent Systems Research. Vol. 137, Pp. 592-597. © 2017

[13] Prof. Chandrakant Mahobiya, "MR Image classification using adaboost for brain tumor type", 2017 IEEE 7th International Advance Computing Conference (IACC)

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

[15] Yunxiang Mao, Joseph Schober et al, "A Deep Convolutional Neural Network Trained on Representative Samples for Circulating Tumor Cell Detection", IEEE 2017

[16] Rasel Ahmmed, Md. Foisal Hossain, "Tumor Detection in Brain MRI Image Using Template based K-means and Fuzzy C-means Clustering Algorithm", 2016 International Conference on Computer Communication and Informatics (ICCCI -2016), Jan. 07 – 09, 2016, Coimbatore, INDIA

[17] Zhe Xiao, Ruohan Huang, Yi Ding, Tian Lan,, "A Deep Learning-Based Segmentation Method for Brain Tumor in MR Images", IEEE 2017

[18] Xiaoguang Chen[1], Xuan Yang et all, "Convolution Neural Network for Automatic Facial Expression Recognition ", Proceedings of the 2017 IEEE International Conference on Applied System Innovation

[19] Saddam Hussain, Syed Muhammad Anwar, Muhammad Majid, "Saddam Hussain, Syed Muhammad Anwar, Muhammad Majid", IEEE 2015

[20] Raymond H. Chan, Chung-Wa Ho, and Mila Nikolova, "Salt-and-Pepper Noise Removal by Median-Type Noise Detectors and Detail-Preserving Regularization", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 14, NO. 10, OCTOBER 2014

[21] Jiafu Jiang, Jing Shen, "An Effective Adaptive Median Filter Algorithm for Removing Salt & Pepper Noise in Images", IEEE 2016.