International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Benign and Malignant Tumor Classification using Machine Learning Technique

Ch. Usha Kumari, N. Madhusudhana Rao, M.N.V.S.S Kumar, Kishore Pinninti, M Pala Prasad Reddy

Abstract: ***Brain tumor is the growth of large mass of abnormal cells in human brain. Brain tumors are of different types benign and malignant. Benign tumor is noncancerous and malignant tumor is cancerous. This research paper classifies benign and malignant using SVM classifier. The image processing techniques are used for image enhancement and restoration. Total 60 images are taken out of which 30 images are benign i.e., noncancerous images and 30 images are malignant i.e., tumor images. Image Filtration is performed on the input MRI image by using median filter. Segmentation is done using thresholding technique and Gray Level Concurrence matrix is used for the feature extraction. Features such as entropy, energy, homogeneity, correlation, contrast, IDM, RMS, standard deviation and mean are extracted for tumor region. After feature extraction all these features are given to SVM classifier to classify benign and malignant tumors. The SVM classifier has given 97.7% accuracy.***

Keywords: Brain Tumor, Benign Tumor, Malignant Tumor, Feature Extraction, Classification, SVM classifier.

I. INTRODUCTION

Tumor refers to an unwanted growth of disease cells inside the human body. Brain tumor is an uncontrollable and abnormal growth of disease cells inside the fibrous region of the brain. A brain tumor is of two types Benign or Malignant. Benign tumor has uniformity structure and less active cells whereas malignant tumor has non uniformity structure and more active cells[1]. It is the most dangerous disease and improper detection leads to death of the person.

Generally MRI (Magnetic Resonance Imaging) is used for tumor detection because of its high resolution images. But due to noise and disturbances in the image it is difficult to detect and classify at the early stage accurately [2]. In turn it leads to the death of the people creating serve loss to the lives of the people.

By using effective digital image processing techniques like image acquisition, gray scale conversion, preprocessing, features extraction and classification on the MRI image of the patient tumor can be detected at the early stage itself and can be classified using a SVM classifier[3].

Digital Image Processing has great scope in the field of Medicine. It uses several computer algorithms to perform many digital image operations on the input image to overcome from distortions, noise and many more which creates degradation in the input image signal [4]. In the field of medicine Digital Image Processing has many applications. Digital image processing plays an important role in detection of Brain Tumor from input MRI image. Generally Brain Tumor refers to abnormal growth of disease cells in the fibrous region of the brain and there are several medical diagnosis techniques for the detection of tumor, but MRI is considered as one of the most accurate and reliable methods of detection as it provides high resolution images of internal organs of the human body [5]. Though the images obtained from MRI has good resolution due to external disturbances the images suffer from distortions and sometimes leads to improper detection and misclassification of Tumor. By using reliable and accurate Digital Image Processing techniques on the input MRI image misclassification of tumor and inappropriate detection of tumor can be prevented[6],[7].

These images contain only intensity information and no colour information and each pixel is a shade of gray. Each pixel can be represented using 8bits or simply one byte. Intensity levels of each pixel ranges from (0-255). MRI images are magnetic resonance images, MRI images of a patient are obtained when patient's body is scanned through MRI machine [8],[9].

Pre-processing is one of the image filtration technique for noise removal. Noise creates degradation in the image signal due to external disturbances, while transferring the image electronically. Salt and pepper noise is the most common type of noise in the images and also found in MRI images. Removing the noise in the image by preprocessing plays an important role in tumor detection and classification [10],[11].

Segmentation is a process of dividing an object into certain partitions. In Digital Image Processing Image segmentation refers to method of dividing an image into certain partitions based on some properties. Segmentation is done based on assumption that images in one particular region possess similar properties and exhibits similar characteristics [12].

Manuscript published on 30 September 2019
* Correspondence Author
Ch. Usha Kumari*, Gokaraju Rangaraju Institute of Engineering and Technology, Dept. of ECE, Hyderabad, Telangana, India
N.Madhusudhana Rao, Gokaraju Rangaraju Institute of Engineering and Technology, Dept. of ECE Hyderabad, Telangana, India
M.N.V.S.S Kumar, Aditya Institute of Technology and Management, Dept. of ECE, Tekkali, India
Kishore Pinninti, VNR Vignana Jyothi Institute of Engineering and Technology, Dept of ECE , Hyderabad, Telangana, India
M Pala Prasad Reddy, Institute of Aeronautical Engineering, Hyderabad, Telangana, India

Feature Extraction plays a prominent role in the field of digital image processing and one of the most important techniques for classification of different kinds of statistics. The core objective of feature extraction is to obtain the

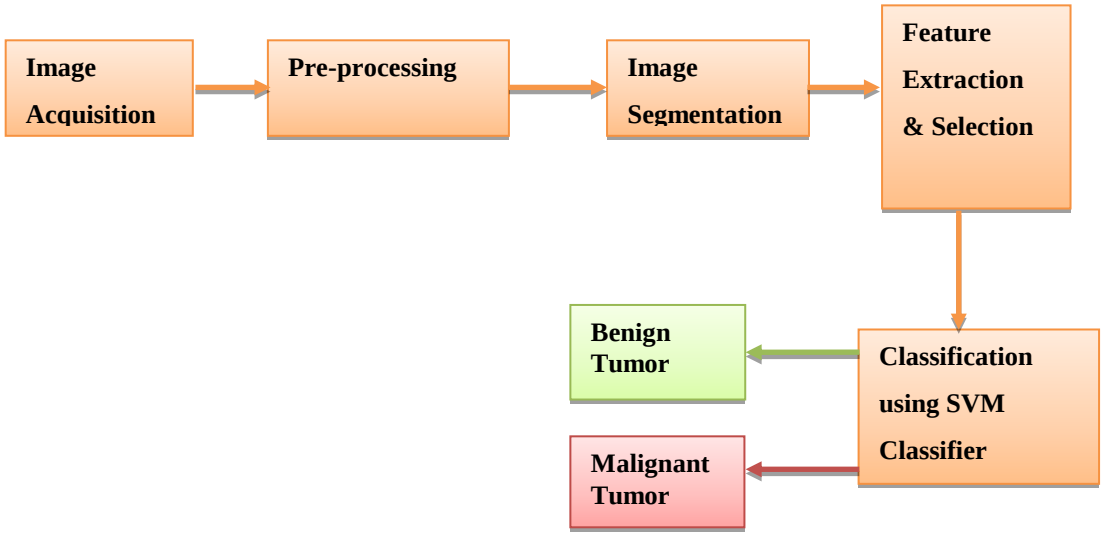

Figure1: Proposed Methodology for Benign and Malignant Tumor Classification

significant information from the original data and to represent in a lower dimensional space. The method of transforming the input data in to a set of features is called feature extraction. The features such as Contrast, Correlation, Energy, Homogeneity, Inverse Different Movement (IDM), Mean, Standard Deviation, RMSE (Root Mean Square Error) and Entropy are extracted. Contrast is used to measure the intensity of the pixels of the digital image. Correlation describes the spatial relationship among the pixels of the image. Homogeneity calculates the proximity of allocation of elements in the glcm to glcm diagonal. Energy measures the uniformity and similarity of the image and output is the sum of the square elements of GLCM. Mean is deliberated as the ratio of sum of all the pixels of an image to total number of pixels of an image. Standard Deviation determines the in homogeneity of the image and higher value indicates enhanced intensity levels and high contrast of edges of an image. Entropy feature is deliberated to exemplify the randomness of the textural image. Inverse Difference Moment (IDM) feature determines the textural property of the image by taking into consideration of the alignment of the image in terms of angles. Root Mean Square Error (RMSE) is the difference between the predicated value and actual value of a system. It is simply a measure of accuracy of a system.

Classification is a supervised learning approach in machine learning. The core objective of a classifier is to categorize to which of category the new observation belongs to based on the training set containing the observations whose category membership is known. Initially a classifier is trained using a train data and after training we load the classifier with test data on which computations should be performed.

II. METHODOLOGY

A. Image Acquisitions

It involves acquiring the digital image in JPEG format. Here a MRI (Magnetic Resonance Imaging) of a patient is acquired from an appropriate database and further operations are performed for tumor detection. After gray scale conversion entire RGB is eliminated and pixels contain only gray level information. This makes further operations easy. The function rgb2gray converts the RGB image to gray-scale image.

B. Preprocessing

Pre-processing is one of the image filtration techniques for noise removal, sharpening the contrast, highlighting the contours and detecting the edges. Noise creates degradation in the image signal. Median Filer is used for removal of noise from the MRI image of the patient. Median filter is a non-linear filer and replaces the intensity value of each pixel with median values of its surrounding pixels. It is an effective method for noise reduction. Medfilt is a function used for performing the preprocessing of input image using median filter.

C. Segmentation

In this research image segmentation is done using Thresholding technique. In thresholding method the image will be converted into binary image (black and white) with vales 0 and 1. Binary image contains one bit per pixel. In this method pixels with value greater than threshold will be converted in to white and pixels with values less than threshold will be converted into black. Finally the entire image will be converted in to black and white. In the proposed system the tumor region is separated from all other regions of digital image, white region represents the tumor affected area of the brain.

D. Feature Extraction and Selection

For feature extraction GLCM (Gray level co-occurrence matrix) is used. The function graycomatrix () creates a GLCM and the function graycoprops () calculates four important features they include Contrast, Correlation, Energy and Homogeneity [25].

GLCM is method of calculating the texture of an image by considering the spatial relationships among the pixels of an image. Every element (i,j) in the resultant glcm represents sum of number of times the pixel with value i occurred in the spatial relationship to a pixel with value j in the input image. In the proposed system total nine features are deliberated for extracted tumor region they include Contrast, Correlation, Energy, Homogeneity, Inverse Different Movement (IDM), Mean, Standard Deviation, RMSE (Root Mean Square Error) and Entropy.
of the brain.

E. Classification using a SVM (Support Vector Machine) Classifier

In machine learning SVM is a supervised learning model which uses supervised algorithms for performing classification and regression on the specified input model. The core objective of a SVM is to find a hyper plane separating different kinds of data and the hyper plane with high marginal distance is selected to minimize the error rate of the system. Hyper planes are the conclusion limitations for classifying the dissimilar kinds of data. The points near to the hyper plane are called support vectors which controls the location and direction of the hyper plane. The process of transforming non separable data to separable data is called kernel trick.

In the proposed system the SVM classifier is used to classify whether the tumor is malignant or benign, initially the classifier is trained using train data taken from reliable and appropriate database and after training the classifier is loaded with new data and classification is performed on 46 images taken from reliable and appropriate data base.

III. Results and Discussions

Benign tumor image processing is shown in Fig: 2. Initially tumor image in JPEG format is converted into gray scale by eliminating RGB shown in Fig2.a. The noise from gray image is removed using median filter shown in Fig2.b. The segmentation is done using thresholding technique shown in Fig2.c. Fig:3 shows Malignant Tumor image processing. Fig3.a is gray scale conversion, Fig:3.b is noise removal from gray image and Fig:3.c is segmented output.

The extracted features for Benign and Malignant tumor image are Contrast, Correlation, Homogeneity, Energy, Mean, Standard deviation, Entropy, RMS and IDM shown in Table:1

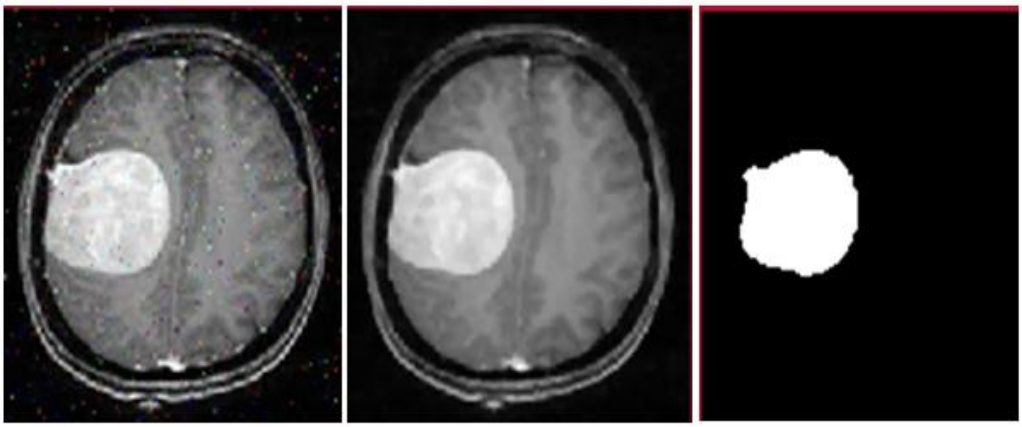

Figure2 Benign Tumor: a) Gray Image b) Filtered Output c)Segmented Output

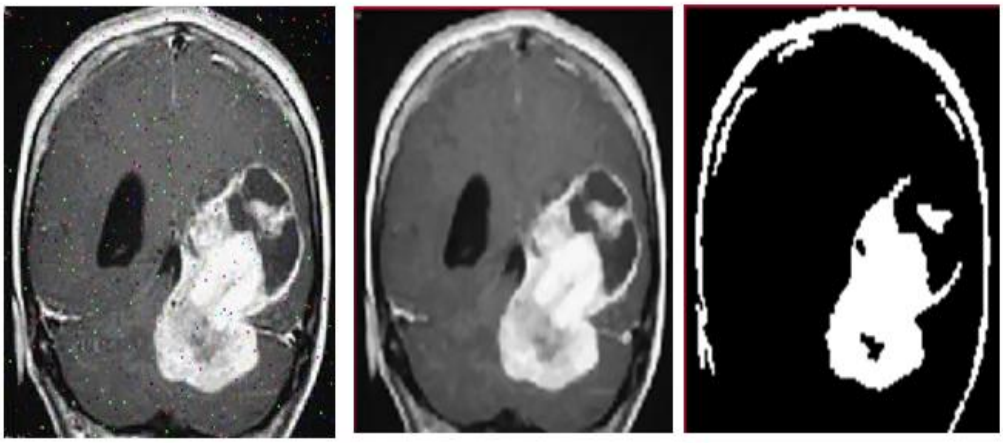

Figure3 Malignant Tumor: a) Gray Image b) Filtered Output c) Segmented Output

Table I: Extracted Features for Benign and Malignant Tumor

S.No	Extracted Feature	Benign Tumor	Malignant Tumor
1	Contrast	0.422692	0.25667
2	Correlation	0.166904	0.13248
3	Homogeneity	0.961941	0.93272
4	Energy	0.873108	0.75872
5	Mean	0.005251	0.00329
6	Standard	0.089661	0.089775
7	Entropy	1.87991	3.38585
8	RMS	0.089802	0.089802
9	IDM	3.30767	1.40956

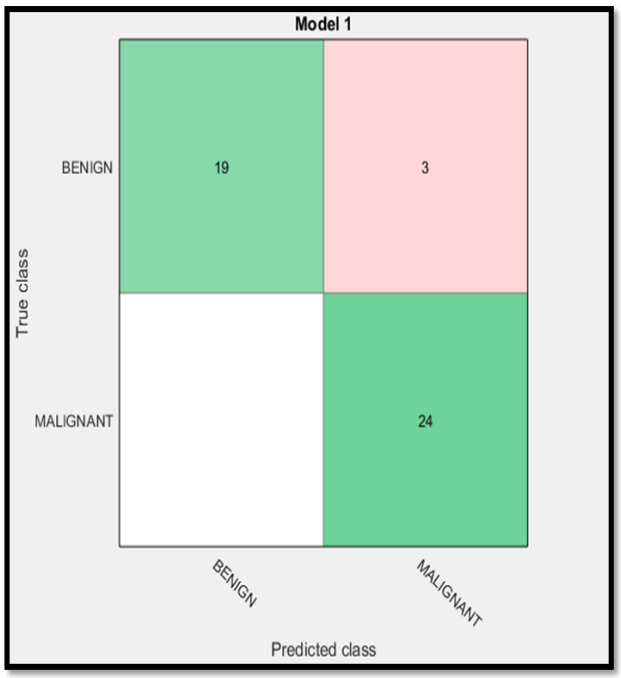

Figure 4: Confusion matrix for Benign Tumor and Malignant Tumor Classification

In this research total 46 images are used for classification of Benign and Malignant tumors, 22 are Benign images and 24 are Malignant images. SVM classifier got 97.8% accuracy with three misclassification in Benign images as shown in confusion matrix plot Fig: 4. False negative rate is zero for Malignant images, as all malignant images have been correctly classified.

IV. CONCLUSION

This research article presents the methodology of detecting and classifying the Brain Tumor at the very preliminary phase itself from MRI Image of the Brain. Digital Image Processing Techniques are used for detection of tumor. The detection and classification of brain tumor is done in five phases. In phase one the RBG image is converted into gray-scale image. In second phase image filtration techniques are applied for noise removal, sharpening the contrast, highlighting the contours and detecting the edges. In the third phase segmentation is done using thresholding technique. In fourth phase features are extracted using Gray level co-occurrence matrix (GLCM). Contrast, Correlation, Homogeneity, Energy, Mean, Standard deviation, Entropy, RMS and IDM are extracted for Benign and Malignant tumor images. SVM classifier achieved 97.8% accuracy with three misclassification in Benign images. Malignant images are correctly classified with zero false negative rate.

REFERENCES

1. S.Murugavalli, V.Anitha,"Brain tumor classification using two-tier classifier with adaptive segmentation technique," IET Computer Vision, Vol 10 issue: 1pp. 9-17, 18 January 2016.
2. Springer Berlin and Heidelberg "Application of Wavelet Transforms and Bayes Classifier to Segmentation Ultrasound Images" IEEE Transactions on Medical ImagingVol.3523/2005, pp. [336-342, 2005].
3. S.Murugavalli, V.Anitha,"Brain tumor classification using two-tier classifier with adaptive segmentation technique," IET Computer Vision, Vol 10 issue: 1pp. 9-17, 18 January 2016.
4. H. Wynne, L.L Mong, and J. Zhang, "Image mining: trends and developments. Journal of Intelligent Information Systems," 19 (1): 2002, pp [7–23].
5. Kumari CU, Mounika G, Prasad SJ. Identifying Obstructive, Central and Mixed Apnea Syndrome Using Discrete Wavelet Transform. InInternational Conference on E-Business and Telecommunications 2019 Mar 22 (pp. 16-22). Springer, Cham..
6. Kumari CU, Prasad SJ, Mounika G. Leaf Disease Detection: Feature Extraction with K-means clustering and Classification with ANN. In2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) 2019 Mar 27 (pp. 1095-1098). IEEE..
7. Kumari CU, Vignesh NA, Panigrahy AK, Ramya L, Padma T. "Fungal Disease in Cotton Leaf Detection and Classification using Neural Networks and Support Vector Machine", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol 08 issue:10, pp-3664-3667, August 2019
8. Kumari CU, Kora P, Meenakshi K, Swaraja K. Short Term and Long Term Path Loss Estimation in Urban, SubUrban and Rural Areas. In2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) 2019 Mar 27 (pp. 114-117). IEEE.
9. M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.
10. (Basic Book/Monograph Online Sources) J. K. Author. (year, month, day). *Title* (edition) [Type of medium]. Volume(issue). Available: http://www.(URL)
11. J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: http://www.atm.com
12. (Journal Online Sources style) K. Author. (year, month). Title. *Journal* [Type of medium]. Volume(issue), paging if given. Available: http://www.(URL)

AUTHORS PROFILE

Dr Ch Usha Kumari is presently working as a professor in Department of ECE, Gokaraju Rangaraju Institute of Engineering and Technology (GRIET), Hyderabad. She completed her Ph.D from Jawaharlal Nehru Technological University, Hyderabad. She completed her M.Tech from Andhra University Visakhapatnam. She had 14years of teaching experience. She published many journals and research papers in national and international conferences. She is life associate member of IETE and fellow member of IEEE.

N.Madhusudhana Rao received B.E.,degree in the Branch of Electronics and Communication Engineering from Anna University, Chennai in 2006, M.Tech., degree in Systems and Signal Processing JNTU University, Hyderabad in 2012. Presently, he is working as Assistant Professor in the Department of Electronics and Communication Engineering, in Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad.

M.N.V.S.S Kumar is presently working as Associate Professor in the Department of Electronics and Communication Engineering at Aditya Institute of Technology and Management, Tekkali. He completed his B.Tech, M.Tech and Ph.D from Andhra University. He published 104 research papers in various journals and conferences. He completed 2 research projects funded by UGC and DLRL.

Kishore Pinninti received B.E degree in Electronics and Communications Engineering from Andhra University in 2004, M.Tech in VLSI System Design from JNTU Hyderabad in 2007 and received Ph.D from JNTU Kakinada in 2018. He is having a total teaching experience of 13 Years and currently working as a Assistant professor in VNRVJIET, Hyderabad, His research areas includes Low Power CMOS Digital and Analog IC design. He had more than 30 publications in various national and International journals/conferences. Dr. Kishore is a member of IEEE, IETE, MISTE professional bodies.

M Pala Prasad Reddy is currently working as Associate Professor in the department of Electrical and Electronics Engineering. He received B.Tech degree from Jawaharlal Nehru Technological University, Hyderabad in the year 2007. He obtained his M.Tech degree with specialization Instrumentation & Control and awarded with Ph.D in Control systems from National Institute of Technology Calicut, Kerala in 2009 and 2019 respectively. He has 10 years of teaching and research experience and published 10 articles in international journals. He is also a member of IEEE and IAENG. His areas of interest include linear and non-linear control systems, soft computing techniques, signal processing and bio-medical instrumentation.