International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 06, Volume 6 (June 2019) www.irjcs.com
SPECIAL ISSUE - 5th International Conference - "ACCE – 2019"

TUMOR DETECTION AND CLASSIFICATION USING RANDOM FOREST ALGORITHM IN BRAIN MRI

Sathya Narayana N | Vidya T | Sarthak U Kuchalli
Vemana Institute of Technology, Koramangala, Bangalore, India

Manuscript History
Number: IRJCS/RS/Vol.06/Issue06/JNCS10133
Received: 29, May 2019
Final Correction: 30, May 2019
Final Accepted: 02, June 2019
Published: June 2019
doi://10.26562/IRJCS.2019.JNCS10133
Editor: Dr.A.Arul L.S, Chief Editor, IRJCS, AM Publications, India

ABSTRACT–In this paper we present an approach to detect whether an MRI scan of a brain contains a tumor or not using machine learning. Once detected, it will then classify the type of tumor as either benign or malignant. In any medical field, the most important resource used by doctors is Medical Images which is a tool with high accuracy. In this work, the system correctly classifies MRI images into images with tumor and images without tumor. This has to be done with no human intervention. In order to apply several types of classifiers, there is a need to pre-process several aspects of the images such as the color, area of interest, image file extension, and contrast level.

Keywords: Magnetic Resonance Imaging (MRI), Pre-Processing, Extraction of Features and Classification.

1. INTRODUCTION

A cluster of abnormal cells growing in any region of brain is termed as brain tumor. It can appear at any region in varying sizes and also with varying pixel intensities. In our project we first classify the scanned images as images with tumor and images without tumor and then we detect the edge of the tumor region from the pre-processed image, once the tumor is detected it is then classified as either benign or malignant. Benign brain tumor, unlike malignant brain tumor, is not cancer cells with homogeneous structure and generally slower growth rate, it does not invade the nearby tissues or spread to other parts of the body the way cancer can. But benign tumor can also be serious if they appear in vital regions of the body such as blood vessels or nerves.

To identify tumor cells we can either use CT scans or MRI scans. We use MRI scans in our proposed system since, CT scan utilize radiations like X-rays to form image inside the body while MRI uses powerful magnetic fields and radiofrequency pulses to produce detailed pictures of organs and other internal body structures. MRIs provide more detailed information about the inner organs (soft tissues) such as the brain, skeletal system, reproductive system and other organ systems that is provided by a CT scan.

The effects of these tumors include blockage of ducts, reduced blood flow, tissue death (necrosis) and nerve pain ordamage. These tumors can be treated through **Surgery**, **radiotherapy-**using high-energy beams such as X-rays or proton to kill these brain tumor cells, **radio-surgery-** using multiple beams of radiation to generate highly focused form of radiation treatment to kill the tumor cells in very small area,each beam of radiation is not particularly powerful, but the point where all beams meet (at the brain tumor) receives a very large dose of radiation to kill the tumor cells, **chemotherapy-**using drugs to kill tumor cells, these drugs can be taken orally in pill form or injected into vein. The conventional method in medicine for brain MR images classification and tumor detection is human inspection.T he MR images data is by nature, a huge, complex and cognitive process. Accurate diagnosis of MR images data is not an easy task and is always time consuming.

In some extreme scenario, diagnosis with wrong result and delay in delivery of a correct diagnosis decision could occur due to the complexity and cognitive process of which it is involved. In this paper machine learning concepts are used to correctly classify MRIs into images with tumor and images without tumor. Thus, the system can be used to save life of many people as the system classifies and alerts the level in which tumor is present. We use ANN algorithm based on random forest classification method to get more accuracy in the result in comparison to manual method. The objective of using random forest method is to classify the detected tumor region as benign or malignant. Finally, an optimal feature set is extracted from these images which are very important for performance enhancement of the automated system.

2. System Design

There are mainly two phases in our system. Training phase and Testing phase. Overview of proposed system has been shown in Fig. Initially to perform classification on MRI mages, we require image database. After gathering images we have to apply various image processing techniques in both training and testing phase. Techniques followed in these phases are, Image Pre-processing, Feature Extraction, Tumor Detection and Classification. Images are required to be pre-processed for feature extraction process.

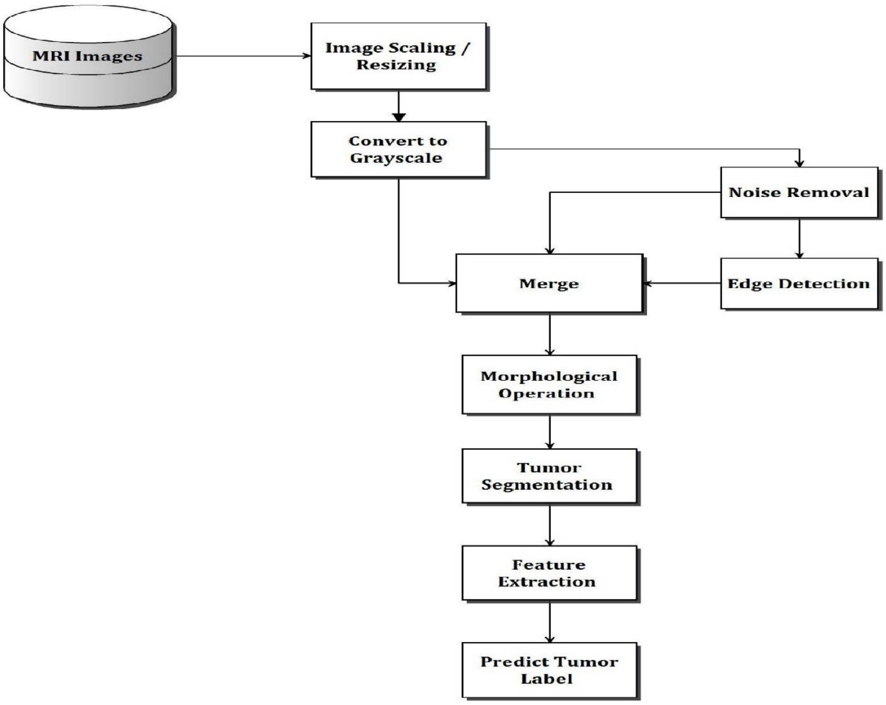


Fig. System Architecture

2.1 Image Pre-processing

The fabric image is collected and stored it into our system. During image pre-processing, there may be artifacts in the images that should be corrected prior to feature measurement and analysis. In this module the data is taken from online source. Further the image is resized for future use. Image resizing, or image scaling, is a geometric image transformation which modifies the image sizebased on an image interpolation algorithm. This image scaling process can increase or decrease the resolution of a target image so that the absolute size of image data is adjusted. The image will be converted to the computer which will assign each pixel a value based on how dark it is. All the numbers are put into an array and the computer does computations on that array. Then the resulting array is given to the next step.

2.1.1RGB to Grayscale Conversion

All grayscale algorithms utilize the same basic two-step process:

- Get the red, green, and blue values of a pixel.
- Use fancy math to turn those numbers into a single gray value.

i. **The lightness method** averages the most prominent and least prominent color:

$$(Max (R, G, B) + min (R, G, B)) / 2$$

ii. **The average method** simply averages the values: (R + G + B) / 3.

International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 06, Volume 6 (June 2019) www.irjcs.com
SPECIAL ISSUE - 5th International Conference - "ACCE – 2019"

iii. **The luminosity method** averages the values, but it forms a weighted average to account for human perception. Everyone are more sensitive to green than other colors, so green is weighted most heavily. The formula for luminosity is 0.21 R + 0.72 G + 0.07 B.

2.1.2 Median Filter Algorithm

Median filter is a nonlinear digital filtering technique, often used to remove noise from an image or signal. Such noise reduction is a typical pre-processing step to improve the results of later processing (for example, edge detection on an image). Median filtering is very widely used in image processing because, under certain conditions, it preserves edges while removing noise[5]. The main idea of the median filter is to run through the signal entry by entry, replacing each entry with the median of neighbouring entries. The pattern of neighbours is called the "window", which slides, entry by entry, over the entire signal. For 1D signals, the most obvious window is just the first few preceding and following entries, whereas for 2D (or higher-dimensional) signals such as images, more complex window patterns are possible (such as "box" or "cross" patterns). Note that if the window has an odd number of entries, then the median is simple to define: it is just the middle value after all the entries in the window are sorted numerically. For an even number of entries, there is more than one possible median.

2.1.3 Threshold based Segmentation

Thresholding is the simplest method of image segmentation. From a gray scale image, thresholding can be used to create binary images. Binary images are produced from color images by segmentation. Segmentation is the process of assigning each pixel in the source image to two or more classes. In image processing, thresholding is used to split an image into smaller segments, using at least one color or gray scale value to define their boundary. The advantage of obtaining a binary image is that it reduces the complexity of the data and simplifies the process of recognition and classification. The most common way to convert a gray level image to a binary image is to select a single threshold value (T). The input to a thresholding operation is typically a gray scale or color image. In the simplest implementation, the output is a binary image representing the segmentation. Black pixels correspond to background and white pixels correspond to foreground (or vice versa). This method of segmentation applies a single fixed criterion to all pixels in the image simultaneously.

Image Segmentation = divide image into (continuous) regions or sets of pixels. The pixels are partitioned depending on their intensity value.

2.2 Tumor Detection

Once the images are per-processed, these images can then be used to identify the tumor region and then it is classified into images with tumor and images without tumor. Here for detecting the tumor region we use morphological operation, where it analyses the difference between the brain tissues and the non-brain tissues. The basic morphological operators are erosion, dilation, opening and closing. The morphology is developed from the grayscale images and binary images.

2.2.1 Morphological Operation

Amorphological operation is erosion followed by dilation with the same structuring element:

$$A \circ B = (A \ominus B) \oplus B$$

Remember that erosion finds all the places where the structuring element fits inside the image, but it only marks these positions at the origin of the element. By following erosion by dilation, we "fill back in" the full structuring element at places where the element fits inside the object. So, an opening can be considered to be the union of all translated copies of the structuring element that can fit inside the object. Openings can be used to remove small objects, protrusions from objects, and connections between objects [5].

2.3 Feature Extraction

Feature extraction is related to dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be redundant, then it can be transformed into a reduced set of features (also named a feature vector). Determining a subset of the initial features is called feature selection. The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data [6].

2.3.2 GLCM:

A GLCM is a matrix where the number id rows and columns is equal to the number of gray levels, G, in the image. It is mainly used in feature extraction from the image. Transformation of such an input data into sets of features is called feature extraction. In this stage, the important features required for image classification are extracted. The GLCM features are used to differentiate between normal and abnormal brain[7].

2.4 Tumor Classification

Random Forest classification algorithm combines multiple weak classification algorithms to form strong classification algorithm. A single algorithm may classify the objects poorly. But if the multiple classifiers are combined with selection of training set at every iteration and assigning right amount of weight in final voting, then good accuracy score is achieved for overall classifier. Now the dataset is split into train and test dataset where Train dataset will be used for model training. After model construction it is time for model training. An ensemble learning based classifier is built that can recognize the brain image either as normal or abnormal[8].

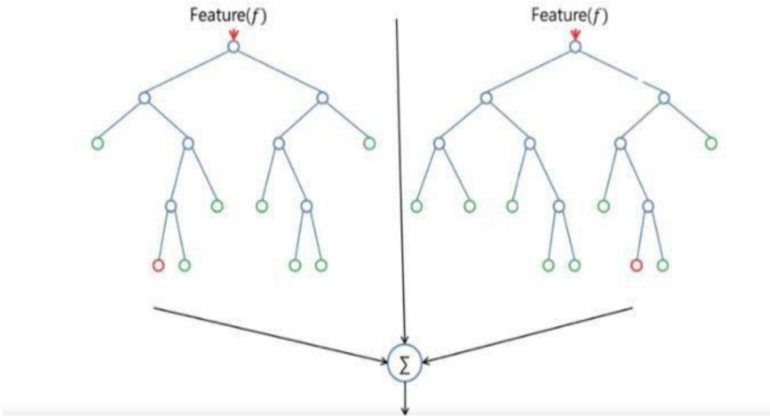


Fig. Random Forest

2.4.1 Classification Algorithm (Random forest Algorithm)

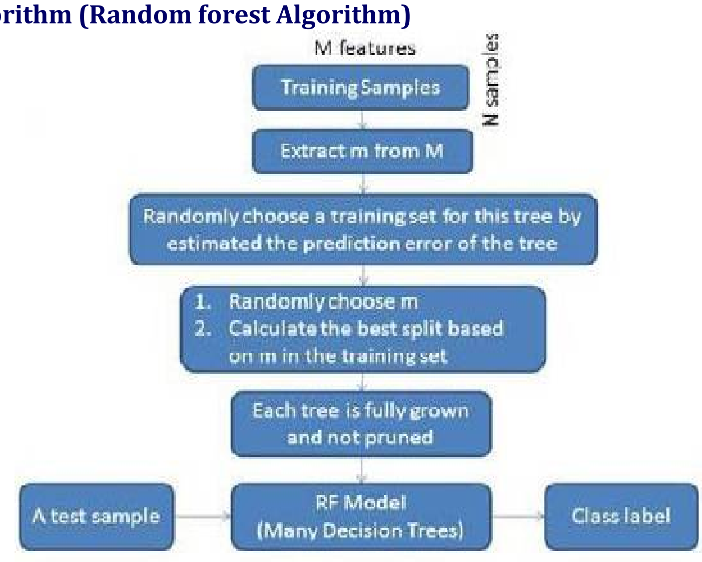


Fig. Random Forest algorithm

Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because its simplicity and the fact that it can be used for both classification and regression tasks. Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

III. EXPERIMENTAL RESULTS

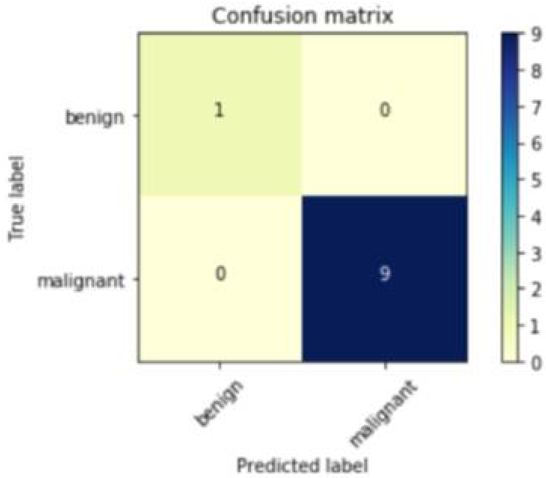


Fig. original and detected tumor image

The confusion matrix can be used to determine the performance of the proposed method. Here classification algorithms used is Random Forest. This matrix describes all possible outcomes of a prediction results in table structure. The possible outcomes of a two class prediction be represented as True positive (TP), True negative (TN), False Positive (FP) and False Negative (FN). The normal and abnormal images are correctly classified as True Positive and True Negative respectively. A False Positive is when the outcome is incorrectly classified as positive when it is a negative. False Positive is the False alarm in the classification process. A false negative is when the outcome is incorrectly predicted as negative when it should have been in fact positive.

In our system consider,
TP= Number of Abnormal images correctly classified
TN= Number of Normal images correctly classified
FP= Number of Normal images classified as Abnormal
FN= Number of Abnormal images classified as Normal.

Precision: The fraction of abnormal images with correct results.

$$\frac{TP}{TP+FP}$$

Sensitivity (Recall): The probability of the test finding the abnormal case among all abnormal cases.

$$\frac{TP}{TP+FN}$$

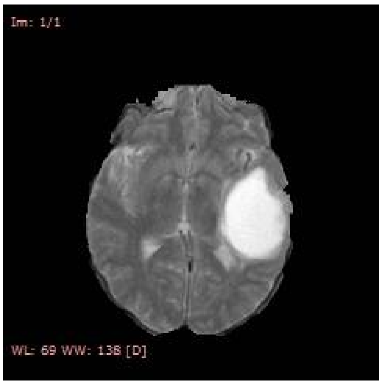
Specificity: The probability of the test finding the normal case

$$\frac{TN}{TN+FP}$$

Accuracy: The fraction of test results those are correct.

$$\frac{TP+TN}{TP+FN+TN+FP}$$

Using these equations, we can analyse which classification method gives better performance. In our system we have analysed 250 MRI images. From 250 images, we have used 220 images for training phase and remaining 30 images for the testing phase.
Results using Random Forest Classification:
Precision: 100%
Sensitivity: 93%
Specificity: 100%
Accuracy: 93%



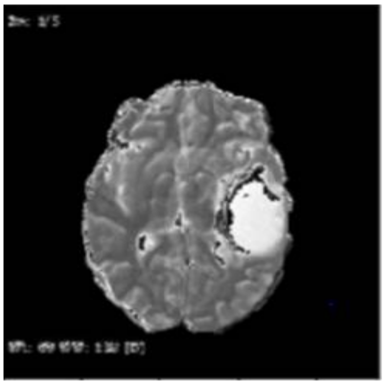
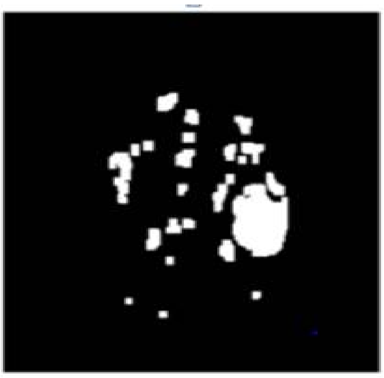


Fig. Original image Fig. Edge detected image



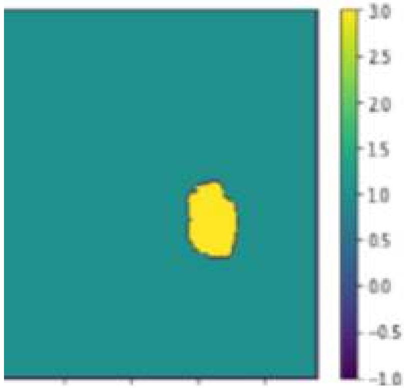
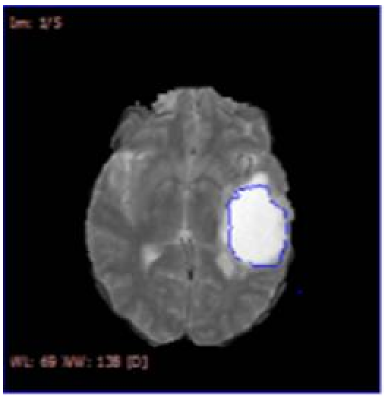


Fig. Morphological operation Fig. Detected tumor region



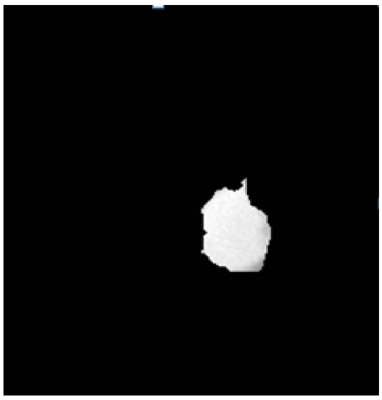


Fig. Detected tumor image Fig. tumor region

IV. CONCLUSION

"Brain Tumor Detection and Classification using Machine Learning Approach" is used to get efficient and accurate results. With the use of Random Forest classification technique tumor has been detected as well as classified into benign or malignant class. All the requirements specified during the analysis and design phase will be fully met. The database collection is one of the significant aspects. Both real-time images and simulated images are used in this project which is an added advantage. Secondly, an extensive pre-processing technique is employed to remove the unwanted noises. The success rate of this step is high which has guaranteed the overall accuracy of the system. Finally, an optimal feature set is extracted from these images which are very important for performance enhancement of the automated system. The developed brain tumor classification system is expected to provide valuable diagnosis techniques for the physicians.

V. FUTURE ENHANCEMENTS

For further enhancement, larger dataset can be considered so that the model can be trained with many data images so as to increase the efficiency and accuracy in predicting the tumor from a MRI images. The videos along with the MRI images can also be included for better and fast detection of tumor which also helps in reduction of the human or manual effort in detecting the tumor and can also be less time consuming. Microscopic observation may also be included so as to analyse the live cell growth of tumor which is a major parameter in detection. Microscopic observation can help in understanding the rate of the rapid growth of the tumor sooner so that the risk of tumor development can be detected in an early stage.

REFERENCES

1. Tom M. Mitchell, Machine Learning, India Edition 2013, McGraw Hill Education.
2. EthenAlpaydm, Introduction to Machine Learning, Second Edition, MIT press.
3. Brain Tumor Detection based on Machine Learning Algorithms, Komal Sharma Research Scholar, AkwinderKaur Assistant Professor, ShrutiGujral Assistant Professor – International Journal of computer application.
4. Machine Learning approach for Brain Tumor Detection, Mahmoud-Al Ayyoub, Ahmad Alabed, GhaithHusari, Omar Darwish – Research Gate publications.
5. R.C Gonzalez and R.E. Woods, Digital Image Processing, Third edtion Prentice-Hall,2009.
6. HanchuanPeng, Fubui Long, and Chris Ding.: Feature Selection based on mutual information: Criteria of Max dependency, Max_relerance and Min_redundancy.:IEEE Transaction on Pattern Analysis and machine Intelligence,Vol. 27 , No. 8, pp . 1226-1238,2005.
7. S. Peckinpaugh, "An Improved Method for Computing Gray-Level Coocurrence Matrix Based Texture Measures", Computer Vision, Graphics, and Image Processing; Graphical Models and Image Processing, Vol. 53, pp. 574-580, 1991.
8. Baskaran.R, Deivamani.M, Kannan.A, 2004. "A multi agent approach for texture based classification and retrieval (MATBCR) using binary decision tree." International journal of computing and information sciences, Vol. 2, No.1, 13-22.