Brain Tumor Detection

Department of Computer Science
Gurudas College
Calcutta University
24/07/2021

Detection of tumorous cells using machine learning models

Submitted in partial fulfillment for the requirements for the degree Bachelor of Science (Honors) in Computer Science.
Academic Year: **2018-2021**

Authors:
• Shoptorshi • Brahmajit • Rajarshi • Bhargav

Supervisor: Kajari Bhattacharjeee

Certificate

Department of Computer Science
Gurudas College
Calcutta University

This is to certify that the project entitled "Brain tumor detection using Machine Learning models" is a bona fide work of **Shoptorshi**, **Brahmajit**, **Rajarshi** and **Bhargav** submitted to Gurudas College, University of Calcutta; in partial fulfillment of the requirement for the award of the degree "Bachelor of Science (Honors)" in Computer Science.

Supervisor:

____

Kajari Bhattacharjeee

Principal:

____

Dr. Mausumi Chatterjee

Department
Head:

____

Srijeeta Charkraborty

Acknowledgement

I would like to express my special thanks of gratitude to our supervisor **Kajari Bhattacharjeee**, teachers of our department **Sonali Gupta** and **Srijeeta Charkraborty** (current department head); as well as our principal ma'am **Dr. Mausumi Chatterjee** who gave us the golden opportunity to do this wonderful project on *Brain tumor detection*, which has also helped us in doing a lot of research and we came to know about many new things.

We are really thankful to them.

Secondly we would like to thank the incredible authors of the research papers in our references/citations for providing us with the information.

Table of Content

Brain Tumor Detection

Using various machine learning models to detect brain tumor.

Abstract

Tumors are cancerous or non-cancerous mass or growth of abnormal cells in brain. Tumors can start in brain, or cancer elsewhere in the body can spread to brain. There are many way to control the occurrence of these abnormal cells. A tumor can be denoted as a malformed mass of tissues wherein the cells multiply abruptly and ceaselessly, that is there is no control over the growth of the cells.

The process of Image segmentation is adopted for extracting abnormal tumor region within the brain. In the MRI (magnetic resonance image), segmentation of brain tissue holds very significant in order to identify the presence of outlines concerning the brain tumor. There is abundance of hidden information in stored in the Health care sector. With appropriate use of accurate data mining classification techniques, early prediction of any disease can be effectively performed.

The project examines list of risk factors that are being traced out in brain tumor surveillance systems. Also the method proposed assures to be highly efficient and precise for brain tumor detection, classification and segmentation. To achieve this precise automatic or semi-automatic methods are needed. The project proposes an automatic segmentation method that relies upon *CNN (Convolution Neural Networks)* , *VGG 16* and *Resent 50* , determining small 7 x 7 kernels. By incorporating this single technique, segmentation and classification is accomplished. CNN (a ML technique) from NN (Neural Networks)wherein it has layer based for results classification.

Various levels involved in the proposed mechanisms are:

1. **Data collection**
2. **Pre-processing**
3. **Average filtering**
4. **segmentation**
5. **feature extraction**
6. **CNN (or any other model) via classification and identification. By utilizing the DM (data mining) techniques, significant relations and patterns from the data can be extracted. The techniques of ML (machine learning) and Data mining are being effectively employed for brain tumor detection and prevention at an early stage.**

Introduction

Domain Description

1. **Neurological Examination**: It is a series of test to measures the function of the patients nervous system and also his/her physical and mental alertness.

2. **Machine Learning**: Machine learning approaches address these problems by mainly using hand-crafted features (or pre-defined features). As an initial step in this kind of segmentation, the key information is extracted from the input image using some feature extraction algorithm, and then a discriminative model is trained to recognize the tumor from normal tissues. The designed machine learning techniques generally employ hand-crafted features with various classifiers, such as random forest, support vector machine (SVM), fuzzy clustering. The designed methods and features extraction algorithms have to extract features, edge-related details, and other necessary information—which is time-consuming. Moreover, when boundaries between healthy tissues and tumors are fuzzy/vague, these methods demonstrate poorer performances.

3. **Brain Scan**: Brain scan is a picture of the internal structure of the brain. A specialized machine takes a scan in the same way as a digital camera takes a photograph. Using computer technology, a scan compiles an image of the brain by photographing it from various angles. Some types of scan uses contrast agent (or contrast dye), which helps the doctor to see the difference between normal and abnormal brain tissues.

   MRI (Magnetic Resonance Imaging): It is a scanning device that uses magnetic field and computer to capture images of the brain on films. It does not use x-rays. It provides pictures from various planes, which permits doctor to create a three-dimensional image of the tumor. The MRI detects signals emitted from normal and abnormal tissues, providing clear images of almost all tumors.

Motivation

The motivation is to develop a software with better segmentation capability for use in medical imaging to detect diseases like brain tumor. Image segmentation has been identified as the key problem of medical image analysis and remains a popular and challenging area of research. Image segmentation is increasingly used in many clinical and research applications to analyze medical imaging datasets; which motivated us to present a snapshot of dynamically changing field of medical image segmentation.

CT (Computed Tomography), MRI (Magnetic Resonance Imaging), PET (Positron Emission Tomography) etc. generates a large amount of image information. With the improved technology, not only does the size and resolution of the images grow but also the number of dimensions increases. In the future, we would like to have algorithms which can automatically detect diseases, lesions and tumors, and highlight their locations in the large pile of images.

The motivation of this work is to increase patient safety by providing better and more precise data for medical decision.

Scope of Work

Deliverables

- Working program to take an MRI scan as input and predict presence of tumorous cells with $\geq 90\%$ accuracy.

Scope

- The working program has external dependencies (libraries) and it's expected to have a them installed for the program to work.

Timeline

- **April 27, 2021** Project Assigned
- **May 2, 2021** Project finalized by supervisor, and group is divided into groups of two.
- **May 3, 2021** Data collection started.
- **May 12, 2021** Project Repository created and coding is started.
- **July 7, 2021** Coding is finished, documentation is started.
- **Just 21, 2021** Documentation complete.

Reports

- Constantly updating and pushing code to repository.
- Both teams staying in touch with each other to keep up with each others progress.
- Report back to supervisor every once a week.

Background

Natarajan [1] proposed brain tumor detection method for MRI brain images. The MRI brain images are first preprocessed using median filter, then segmentation of image is done using threshold segmentation and morphological operations are applied and then finally, the tumor region is obtained using image subtraction technique. This approach gives the exact shape of tumor in MRI brain image. Joshi [2] proposed brain tumor detection and classification system in MR images by first extracting the tumor portion from brain image, then extracting the texture features of the detected tumor using Gray Level Co- occurrence Matrix (GLCM) and then classified using neuro-fuzzy classifier. Amin and Mageed [3] proposed neural network and segmentation base system to automatically detect the tumor in brain MRI images. The Principal Component Analysis (PCA) is used for feature extraction and then Multi-Layer Perceptron (MLP) is used classify the extracted features of MRI brain image. The average recognition rate is 88.2% and peak recognition rate is 96.7%. Sapra [4] proposed image segmentation technique to detect brain tumor from MRI images and then Probabilistic Neural Network (PNN) is used for automated brain tumor classification in MRI scans. PNN system proposed handle the process of brain tumor classification more accurately. Suchita and Lalit [5] proposed unsupervised neural network learning technique for classification of brain MRI images. The MRI brain images are first preprocessed which include noise filtering, edge detection, then the tumor is extracted using segmentation. The texture features are extracted using Gray-Level Co-occurrence Matrix(GLCM) and then Self-Organizing Maps (SOM) are used to classify the brain as normal or abnormal brain, that is, whether it contain tumor or not. Rajeshwari and Sharmila [6] proposed preprocessing techniques which are used to improve the quality of MRI image before using it into an application. The average, median and wiener filters are used for noise removal and interpolation based Discrete Wavelet Transform (DWT) technique is used for resolution enhancement. The Peak Signal to Noise Ratio (PSNR) is used for evaluation of these techniques.

Methodology

As per literature survey, it was found that automated brain tumor detection is very necessary as high accuracy is needed when human life is involved. Automated detection of tumor in MR images involves feature extraction and classification using machine learning algorithm. In this paper, a system to automatically detect tumor in MR images is proposed as shown in Figure 1

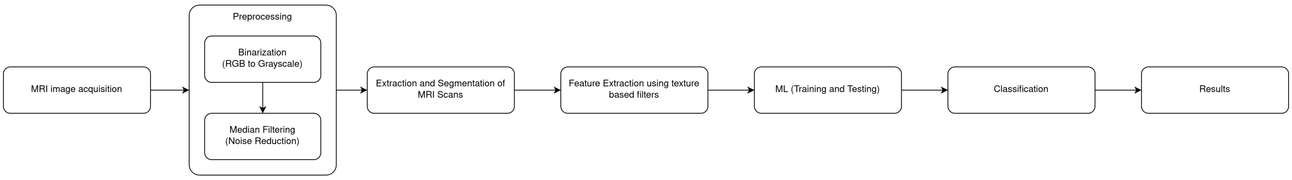
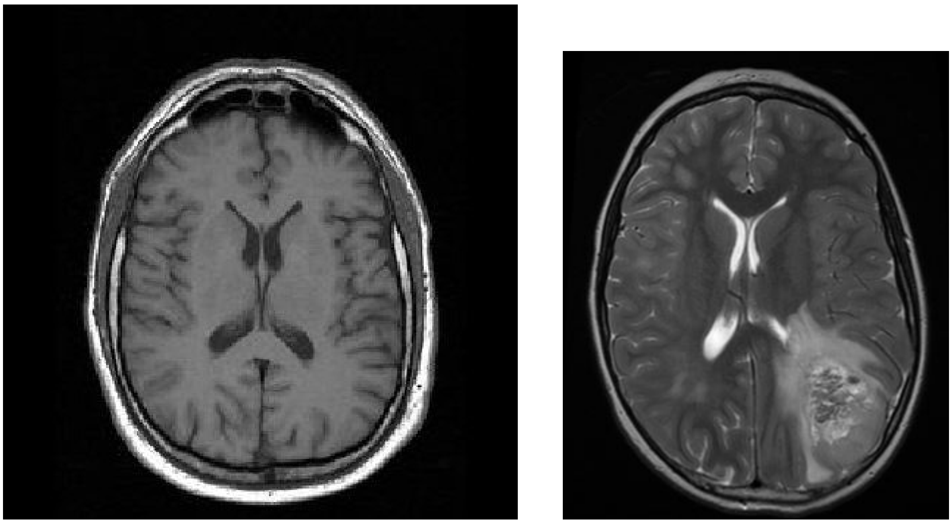


Figure 1: Proposed Methodology

Image Acquisition

The MRI brain images are acquired and are given as input to pre-processing stage. The sample brain MR images are shown in Figure 2.



(a) MRI scan shown no presence of tumor (b) MRI scan of a tumorous cell

Figure 2: MRI Scans

Preprocessing

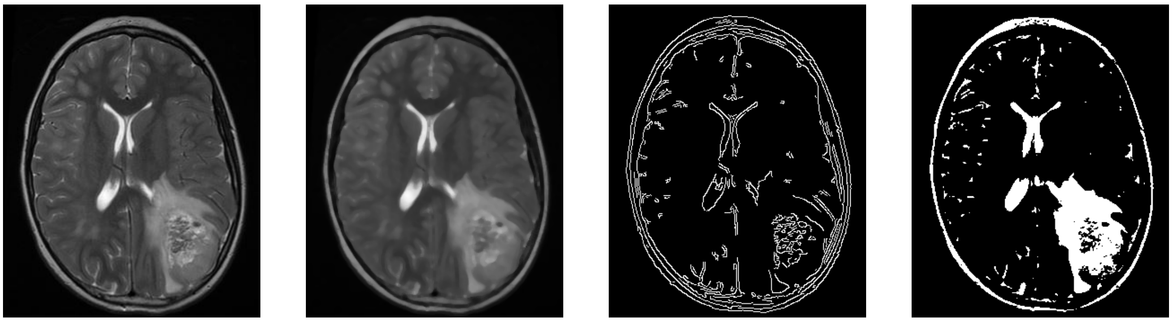
Preprocessing is needed as it provides improvement in image data which enhances some of the image features which are important for further processing. The pre-processing steps that are applied to MR image are as follows:

1. The RGB MR image is converted to gray scale image and then median filter is applied for noise removal from brain MR images as shown in Figure 3b. The noise is to removed for further processing as high accuracy is needed.

2. Then edges are detected from filtered image using canny edge detection as shown in Figure 3c. The edge detected image is needed for segmentation of the image

3. Then watershed segmentation is done for finding the location of the tumor in the brain image as shown in Figure 3d. Segmentation is the process of dividing an image into multiple

segments. The aim of segmentation is to change representation of image into something which is more easy to analyze. The result of watershed segmentation is label image. In label image, all the different objects identified will have different pixel values , all the pixels of first object will have value 1, all the pixels of second object will have value 2 and so on.



(a) Original Image (b) Filtered Imaage (c) Edge Detection (d) Segmentation

Figure 3: preprocessing operations

Feature Extraction

When input to an algorithm is very large and redundant to be processed, it is transformed into reduced representative set of features called feature vector. Transformation of input data into set of features is called feature extraction. In this step, the important features needed for image classification are extracted. The segmented brain MR image is used and texture features are extracted from the segmented image which shows the texture property of the image. These features are extracted using Gray Level Co-occurrence Matrix (GLCM) as it is robust method with high performance.

The GLCM features are extracted as follows:

1. Energy: It gives a measure of textural uniformity, that is, measure of pixel pair repetitions.

$$E = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j)^2 \text{ here, range = [0, 1]} \tag{1}$$

2. Contrast: It gives a measure of intensity contrast between a pixel and its neighbor over the whole image.

$$Con = \sum_{n=0}^{N_g-1} n^2 \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j)^2 \text{ here, range = [0, 1]} \tag{2}$$

3. Correlation: It gives a measure of how correlated a pixel to its neighbor over the whole image.

$$C = \frac{1}{\sigma^x \sigma^y} \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i,j)p(i,j)^2 - \mu_x \mu_y \text{ here, range = [-1, 1]} \tag{3}$$

4. Homogeneity: It gives a measure of closeness of distribution of elements in GLCM to GLCM diagonal.

$$H = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{p(i,j)}{1 + \pmod{i, j}} \text{ here, range = [0, 1]} \tag{4}$$

Classification

The Machine learning algorithms are used for classification of MR brain image either as normal or abnormal. The major aim of ML algorithms is to automatically learn and make intelligent decisions. The feature set formed by above specified method was applied to Multi-Layer Perceptron (MLP) and Naive Bayes for classification. MLP [3] is a feed forward artificial neural network model that maps sets of input data into a set of appropriate output. It is known as feed forward because it does not contain any cycles and network output depends only on the current input instance. In MLP, each node is a neuron with a nonlinear activation function. It is based on supervised learning technique. Learning take place by changing connection weights after each piece of data is processed, based on the amount of error in the target output as compared to the expected result. The goal of the learning procedure is to minimize error by improving the current values of the weight associated with each edge. Because of this backward changing process of the weights, model is named as back-propagation.

Naive bayes is a supervised learning as well as statistical method for classification. It is simple probabilistic classifier based on Bayes theorem. It assumes that the value of a particular feature is unrelated to the presence or absence of any other feature. The prior probability and likelihood are calculated in order to calculate the posterior probability. The method of maximum posterior probability is used for parameter estimation. This method requires only a small amount of training data to estimate the parameters which are needed for classification. The time taken for training and classification is less.

References

[1] P. Natarajan, N Krishnan, Natasha Sandeep Kenkre, et al. "Tumor detection using threshold operation in MRI brain images". In: *2012 IEEE International Conference on Computational Intelligence and Computing Research*. 2012, pp. 1–4. DOI: 10.1109/ICCIC.2012.6510299.

[2] Dipali Joshi, N.K. Rana, and V.M. Misra. "Classification of Brain Cancer using Artificial Neural Network". In: June 2010, pp. 112–116. DOI: 10.1109/ICECTECH.2010.5479975.

[3] Mohammed Abdel-Megeed Mohammed Salem. "Brain Tumor Diagnosis Systems Based on Artificial Neural Networks and Segmentation using MRI". In: Jan. 2012.

[4] Pankaj Sapra, Rupinderpal Singh, and Shivani Khurana. "Brain Tumor Detection Using Neural Network". In.

[5] Suchita Goswami and L.K.P. Bhaiya. "Brain Tumour Detection Using Unsupervised Learning Based Neural Network". In: Apr. 2013, pp. 573–577. ISBN: 978-1-4673-5603-9. DOI: 10.1109/CSNT.2013.123.

[6] S. Rajeshwari and T. Sharmila. "Efficient quality analysis of MRI image using preprocessing techniques". In: Apr. 2013, pp. 391–396. ISBN: 978-1-4673-5759-3. DOI: 10.1109/CICT.2013.6558127.