

UNIVERSITY OF HERTFORDSHIRE

Department of Computer Science

**MSc Title**: Data Science and Analytics

**Module Code and Title:** 7COM1039-0509-2020 - Advanced Computer Science Masters Project

**Date:** 26-08-2021

**PROJECT TITLE**: Brain Tumor Detection using Image Processing Techniques

**Name:** Ali Riaz

**Student ID: 19021008**

**Supervisor: Supervisor’s name:** Raimund Kirner

Table of Contents

[Table of Contents 2](#_Toc80651708)

[Table of Figures: 3](#_Toc80651709)

[1. Introduction and Overview: 4](#_Toc80651710)

[1.1 Hypothesis 6](#_Toc80651711)

[1.2 Description 7](#_Toc80651712)

[1.3 Aim of Project 9](#_Toc80651713)

[1.4 Research Questions 11](#_Toc80651714)

[Literature Review 13](#_Toc80651715)

[2. Methodology 20](#_Toc80651716)

[2.1 Dataset Collection 20](#_Toc80651717)

[2.2 Model Selection 20](#_Toc80651718)

[2.3 IDE and Libraries Used 21](#_Toc80651719)

[2.4 Working 24](#_Toc80651720)

[2.4.1 Defining Model 25](#_Toc80651721)

[2.4.2 Model Compile and Training 29](#_Toc80651722)

[3. Results and Discussion 30](#_Toc80651723)

[References 34](#_Toc80651724)

[Appendices 37](#_Toc80651725)

[Appendix 1: 37](#_Toc80651726)

[Appendix 2: 43](#_Toc80651727)

# Table of Figures:

[Figure 1 Brain Pathology 18](#_Toc80651659)

[Figure 2 wavelet 20](#_Toc80651660)

[Figure 3Imported Libraries 25](#_Toc80651661)

[Figure 4 Healthy Tissue 25](file:///C:\Users\Faisal\Downloads\final%20report%20(2)%20(1).docx#_Toc80651662)

[Figure 5 Tumor Tissue 25](#_Toc80651663)

[Figure 6: Model Summary 29](file:///C:\Users\Faisal\Downloads\final%20report%20(2)%20(1).docx#_Toc80651664)

[Figure 7 Model Training 30](#_Toc80651665)

[Figure 8 Accuracy 31](#_Toc80651666)

[Figure 9: Model Accuracy 32](file:///C:\Users\Faisal\Downloads\final%20report%20(2)%20(1).docx#_Toc80651667)

[Figure 10: Model Loss 32](file:///C:\Users\Faisal\Downloads\final%20report%20(2)%20(1).docx#_Toc80651668)

[Figure 11: Prediction Results 33](file:///C:\Users\Faisal\Downloads\final%20report%20(2)%20(1).docx#_Toc80651669)

[Figure 12 Dataset images with no tumor 44](#_Toc80651670)

[Figure 13 Images dataset with tumor 45](#_Toc80651671)

# 

# Introduction and Overview:

The main idea of this research is to design an algorithm of machine learning and its implementation. Deep learning and Image processing techniques are also used. These are inherited from the Major field called Artificial Intelligence. Artificial intelligence is emerging field and it can be seen in every type of industry wither its agriculture, science or business etc. Model is built with some algorithm designed according the requirement and goal need to be achieved. The models are trained on the dataset of images or numeric data accordingly. After the model is trained it is able to identify the new data input or make some predictions based on the previous learning. The datasets are quite large which may take years for a human being to understand and predict rightly or never. For instance if AI algorithms are used in the business then they process the years of large data set and predict the future profits and also suggest the bestselling products. Artificial intelligence algorithms are also being used to recommend the products to the customer to increase the sales and they do really help. Artificial intelligence is also stepped into the field of medical, and in this research we choose the medical field to create the algorithm for. Brain tumors have an effect on the humans badly, attributable to the unusual growth of cells inside the cerebrum. It will damage correct cerebrum operate and can be grievous. 2 kinds of brain tumors are known as benign cancer and malignant cancer. Benign cancer is less injurious than malignant cancer as malignant are quick growing and injurious whereas benign are slow developing and fewer injurious. Medical imaging methodology is employed to make optical illustration of inside of the frame for medical functions, also non-invasive prospects are often diagnosed by this technology. The most goal of medical image process is to spot correct and substantive data exploitation pictures with the minimum error potential. Tomography is principally accustomed to get pictures of the frame and cancerous tissues attributable to its high resolution and higher quality pictures compared with different imaging technologies. Neoplasm identifications through tomography pictures could be a troublesome task attributable to the quality of the brain. Tomography pictures are often processed and also the neoplasms is often metameric. These tumors are often metameric exploitation numerous picture segmentation methodologies. The method of characteristic brain tumors through tomography pictures are often classified into 4 completely different parts pre-processing, image segmentation, feature extraction and image classification.

Artificial Intelligence has reduced the human efforts, all the processing is done using machine s that are capable of diagnosing the tumor cells in brain. Humans just have to read the output of the machine. Basic idea behind the scene is that an algorithm is designed that is capable of training and testing on the dataset and make predictions according to the training of the model. Dataset images containing healthy brain tissues and tumor tissues are fed to the model which then performs necessary operations and the refined data is obtained at the output layer of the model as model comprises of different input, hidden and output layers. The trained model is then tested on some random images and accuracy of the model is calculated which can then be used in practical field after completion of testing phase.

Cerebrum tumor is perhaps the most thorough infections in the clinical science. A powerful and proficient examination is consistently a vital worry for the radiologist in the untimely period of tumor development. Histological evaluating, in view of a stereotactic biopsy test, is the highest quality level and the show for identifying the grade of a cerebrum tumor. The biopsy methodology requires the neurosurgeon to penetrate a little opening into the skull from which the tissue is gathered. There are many danger factors implying the biopsy test, including draining from the tumor and cerebrum causing contamination, seizures, extreme headache, stroke, unconsciousness and even passing. However, the primary worry with the stereotactic biopsy is that it isn't 100% exact which might bring about a genuine symptomatic mistake followed by a wrong clinical administration of the illness. Tumor biopsy being trying for mind tumor patients, non-intrusive imaging methods like Magnetic Resonance Imaging (MRI) have been broadly utilized in diagnosing mind tumors. Consequently, improvement of frameworks for the recognition and expectation of the grade of tumors dependent on MRI information has gotten essential. However, from the start of the imaging methodology like in Magnetic Resonance Imaging (MRI), the appropriate perception of the tumor cells also, its separation with its close by delicate tissues is fairly troublesome undertaking which might be expected to the presence of low light in imaging modalities or its enormous presence of information or a few intricacy and change of tumors-like unstructured shape, practical size and flighty areas of the tumor. Robotized imperfection recognition in clinical imaging utilizing AI has gotten the rising field in a few clinical analytic applications. Its application in the identification of mind tumor in MRI is exceptionally significant as it gives data about unusual tissues which is important for arranging treatment.Studies in the new writing have additionally announced that programmed mechanized location and finding of the illness, in view of clinical picture investigation, could be a decent option as it would save radiologist time and furthermore get a tried exactness. Besides, if PC calculations can give strong and quantitative estimations of tumor portrayal, these robotized estimations will significantly support the clinical administration of cerebrum tumors by liberating doctors from the weight of the manual portrayal of tumors. The AI based methodologies like Deep ConvNets in radiology and other clinical science fields assumes a significant part to analyze the infection in a lot easier way as never done and thus giving a doable option in contrast to careful biopsy for cerebrum tumors. In this undertaking, we endeavored at distinguishing and characterizing the cerebrum tumor and contrasting the aftereffects of parallel and multi class characterization of cerebrum tumor with and without Transfer Learning utilizing Convolutional Neural Network (CNN) engineering.

## 1.1 Hypothesis

If we will identify tumor in brain in early stages, and differentiate between its types that either it's benign or malignant, we will save an awfully valuable time in patient treatment which might end in successful treatment of patient. Artificial Intelligence (AI) is now widely used in practically every other profession, including medicine and law. The medical field uses a variety of machine learning algorithms capable of detecting disorders and assisting clinicians in a quick diagnosis. A Chabot is another example of artificial intelligence being employed. There are algorithms in place in the realm of e-commerce which aid in product selection by displaying recommendations based on the user's searches. It’s been also used in the YouTube searches to enhance the user experience, we start writing the search query and a lot of suggestions comes up instantly to choose from, by this example we can conclude the process is very fast. In a Human resource management departments AI helps to filter out the right candidate for the job, the data of all the candidates is fed to the system and required skills also, the system process the data and then choose the best fit candidate. This saves a lot of time and makes life easier for Human Resource Managers. Additionally, AI affects intelligent cyber security systems. In order to generate the report, algorithms are employed to find irregularities and flaws in the system. There is a firewall that defends against hacking attacks. Algorithms that are trained on the company's past sales records and forecast future sales are developed in business. Artificial Intelligence, in a nutshell, is influencing every field. 'Artificial Intelligence' is a large field with several sub-fields. Convolutional neural networks (CNNs) are utilized in this simulation to train a model that will be used to predict future images based on the model's training data. The following questions will be the basis for our simulation work. Which strategy should be used to get the best results possible? How do we choose the optimal model for our dataset? To enhance the model's efficiency, which layers should be included?

## 1.2 Description

Brain tumor detection and analysis is very challenging because of various factors like the size of the tumor is not unified, shape of the tumors are different in each case, tumor presence and overall appearance etc. There is no accurate measure of the tumor that’s why it’s really hard to find it (Qayyum, 2017). Once the tumor is detected at its early stages then it can be cured, the later stages are hard to handle and results in bad results. Currently MRI i.e. Medical Reasoning Imaging is being used to detect the tumors. MRI is the visual representation of the human body obtained from the scanning. This is extensively used methodology in the medical field for that purpose in which the analysis is based on the human experience which can result into the false results. Human beings are not capable of processing the large data as machine can do. Before the advancement of artificial intelligence, brain tumor was difficult to diagnose and it take more time to diagnose. After the artificial intelligence is introduced brain tumor is detected using machine learning techniques that helps in the easy diagnosis of tumor cells in brain. Different algorithms are available through which machine is capable of diagnosing brain tumor cells at early stages before its too late.

An abnormal and cell improvement within the brain is called as cancer. The human psyche is the most open piece of the body. It controls muscle improvements and interpretation of unmistakable information like torture, etc. The human psyche involves Gray Matter (GM), White Matter (WM), and Cerebrospinal Liquid (CSF) and in light of variables like an assessment of tissues, space of inconsistencies, breakdowns, and pathologies, and expressive radiology, a presence of growth is recognized. Growth in the brain can impact such material information and muscle advancements or even results in an extra hazardous situation that joins death. Dependent upon the spot of starting, growth can be orchestrated into fundamental cancers and discretionary cancers. If the cancer is begun inside the skull, then the growth is known as fundamental frontal cortex cancer regardless in the event that cancer's presentation place is somewhere else in the body and moved towards the brain, then such cancers are called discretionary cancers. Psyche growth can be of the going with sorts glioblastoma, sarcoma, metastatic bronchogenic carcinoma dependent on the center plane. While a couple of growths like meningioma can be easily divided, others like gliomas and glioblastomas are extensively more difficult to restrict. World Health Organization (WHO) ordered gliomas into - HGG/high-grade glioma/glioblastoma/IV stage/hurtful and LGG/low-quality glioma/II and III stage/pleasant. Yet most of the LGG cancers have all the more sluggish improvement rates diverged from HGG and are responsive to therapy, there is a subgroup of LGG growths which if not dissected previously and left untreated could incite GBM. In the two cases, the right therapy organizing (tallying an operation, radiotherapy, and chemotherapy autonomously or in the blend) becomes basic, taking into account that an early and suitable acknowledgment of the cancer grade can incite a respectable perception. Perseverance time for a GBM (Glioblastoma Multiform) or HGG patient is especially low for instance in the extent of 12 to 15 months.

Attractive Resonance Imaging (MRI) has become the standard non-intrusive method for mind tumor finding in the course of the most recent couple of many years, because of its further developed delicate tissue contrast that does not utilize hurtful radiations not at all like different techniques like CT(Computed Tomography), X-beam, PET (Position Emission Tomography) checks and so forth The MRI picture is essentially a framework of pixels having trademark highlights. Since glioblastomas are infiltrative tumors, their lines are frequently fluffy and hard to recognize from solid tissues. As an answer, more than one MRI methodology is frequently utilized for example T1 (turn cross section unwinding), T1-differentiated (T1C), T2 (turn unwinding), proton thickness (PD) contrast imaging, dispersion MRI (dMRI), and liquid lessening reversal recuperation (FLAIR) beat groupings. T1-weighted pictures with intravenous differentiation feature the most vascular districts of the tumor (T 1C gives considerably more precision than T1.), called Enhancing tumor' (ET), alongside the ‗tumor center' (TC) that doesn't include peritumoral edema. T2-weighted (T2W) also, T2W-Fluid Attenuation Inversion Recovery (FLAIR) pictures are utilized to assess the tumor and peritumoral edema together characterized as the whole tumor' (WT). Gliomas and glioblastomas are hard to recognize in T1, T1c, T2 and PD. They are better recognized in FLAIR modalities.

## 1.3 Aim of Project

Brain tumors are alarming to the human body, tumors are in fact life threatening and occurs due to the abnormal growth of cells present in the brain. There are two kinds of tumors are detected so far named as malignant tumors and benign tumors. Benign tumors are less harmful as they grow slowly but the other king which is malignant tumors grow rapidly and is more harmful to the humans. The medical imaging methodology is used to visually depict the inside of the form for medical reasons and this technology diagnoses non-invasive options. The fundamental objective of medical imaging is to detect accurate and significant information with the least potential mistake through imaging. Because of its high resolution and superior pictures in comparison to other imaging methods MRI mainly get images of the figure and cancer tissues (Bauer, 2013). MRI pictures may be challenging to identify tumors because of brain architecture. Images of MRI are processed as well as a segmentation of the tumor. These tumors are classified by different imaging methods (Menze, 2015). The method of identification of brain cancers by MRI imaging is generally divided into four separate preprocessing, picture segmentation, extraction and image classification sections.

Neural Networks (NN) structure the foundation of profound learning, a subfield of AI where the calculations are enlivened by the construction of the human mind. NN take in information, train themselves to perceive the examples in this information and afterward foresee the yields for another arrangement of comparable information. NN are comprised of layers of neurons. These neurons are the center handling units of the organization. First we have the information layer which gets the information; the yield layer predicts our last yield. In the middle, exist the secret layers which perform the greater part of the calculations needed by our organization. Our cerebrum tumor pictures are made out of 128 by 128 pixels which compensate for 16,384 pixels. Every pixel is taken care of as contribution to every neuron of the principal layer. Neurons of one layer are associated with neurons of the following layer through channels .Each of these channels is allocated a mathematical worth known as ‗weight'. The sources of info are duplicated to the relating weight and their total is sent as contribution to the neurons in the secret layer. Every one of these neurons is related with a mathematical esteem called the bias' which is then added to the information aggregate. This worth is then gone through a limit work called the activation work'. The aftereffect of the actuation work decides whether the specific neuron will get initiated or not. An initiated neuron communicates information to the neurons of the following layer over the channels. Thusly the information is proliferated through the organization this is called forward engendering'. In the yield layer the neuron with the most elevated esteem fires and decides the yield. The qualities are fundamentally a plausible. The anticipated yield is contrasted against the genuine yield with understand the ‗error' in forecast. The greatness of the RCCIIT, DEPT. OF EE Page 13 mistake provides a sign of the guidance and greatness of progress to diminish the blunder. This data is then moved in reverse through our organization. This is known as back spread'. Presently dependent on this data the loads are changed. This pattern of forward spread and back engendering is iteratively performed with various sources of info. This cycle proceeds until our loads are doled out to such an extent that the organization can anticipate the sort of tumor accurately in a large portion of the cases. This finishes our preparation cycle. NN might require hours or indeed, even a long time to prepare however time is a sensible compromise when contrasted with its degree Several tests show that after pre-handling MRI pictures, neural organization grouping calculation was the best more explicitly CNN(Convolutional Neural Network) when contrasted with Support Vector Machine(SVM),Random Forest Field.

Neurons are linked to form a full brain network. Communication occurs between them. In designing the neural network, scientists took into account how the human brain works. Each neuron in the human brain is linked to form a complex network. In order to improve itself, the human brain is able to self-learn depending on its environment. On a similar note, in deep learning, neural networks are constructed in a way that replicates the human brain's self-learning process and adds that learning to its neural network for later use as a prediction or some other beneficial function.

First and foremost, a dataset is selected for the aim of training the models. In order to prepare the dataset for training the model, a few steps need to be taken. Datasets such as the form of the dataset, the color model, etc., are analyzed in detail. As a result, the data set is correctly organized into classes, each with a suitable class name. If the class name is not relevant to the data it contains, then it should be changed. If you're going to utilize a class folder for training purposes, don't put any unnecessary photos in it! Otherwise, it will have a negative impact on the model's performance and accuracy. To train and test the model, the dataset will be divided into training and testing datasets. After that, the dataset is subjected to a variety of procedures before being sent to the model. For example, you may reshape your dataset according to your chosen model, or resize all your photos to a proportionate size because they are all various sizes.

## 1.4 Research Questions

1. **How will it differentiate between healthy tissues and tumor tissues?**

We have a dataset comprising of images of two different collections named as ‘yes’ and ‘no’. ‘Yes’ class contains images having brain tumor and ‘no’ class contains images of healthy tissues. When the model is trained on both the classes, the model will be able to differentiate tumor tissues and healthy tissues.

1. **How can we achieve 95% accuracy detecting tumors in brain?**

The accuracy of the model depends upon the structure of the model and the training data that is being given as input to the model. The structure of the model is decided by the layers we add to the model and the model chosen for training purpose. The missing values should be treated and refrain from unnecessary data. Adding more and accurate data also add value to the accuracy.

1. **Difference between supervised learning and unsupervised learning. Which is adopted and Why?**

Machine Learning can be supervised, unsupervised, and reinforcement learning. In supervised learning, the algorithm of Machine Learning is trained on the labeled dataset which means we need to feed the algorithm with the proper dataset on which it is trained and make the predictions. For that purpose, an extra effort of dataset collection is required and the dataset is named accordingly and properly. As we allow to train algorithms on new data, the efficiency and working of the model improve. In unsupervised learning, there is no need to collect the labeled dataset like in supervised learning. The unsupervised learning algorithm is capable of working with unlabeled data. While is reinforcement learning the algorithm is capable enough to learn from new things like the human brain. The algorithm improves itself and learns from new situations and can make its own decision according to its learning and situation and store it in the memory for future use. In our research, we will be using a supervised learning approach as we have labeled datasets for model training. The dataset is divided into proper classes and named accordingly which helps in the accurate predictions. Our dataset is comprised of a tumor dataset on which model will be trained.

# Literature Review

The length of medical image repositories be growing speedily by utilization of digital imaging knowledge in medical centers. These causes issue in managing and querying these massive databases resulting in the requirement of content primarily based medical image retrieval (CBMIR) systems. The learned options and also the classification results square measure won’t to retrieve medical pictures. A significant challenge in CBMIR systems be that the linguistics gap that exists between the low level optical data captured by imaging devices and high level linguistics data recognized by human. For retrieval of medical images of different body parts, the proposed method could be used with high efficiency (Qayyum, 2017). Medical image analysis is known as the clinical problem solution by analyzing the images generated during the medical practice aiming to extract the effective information from image. Bio-medical imaging is the most advancing field in the present era. Deep learning is used for machine learning with contrast of neural networks. As feature selection and calculation is a demanding job, deep convolutional networks is best for use as it includes different application area such as deformity identification, segmentation, disease categorization and computer support identification and retrieval (Anwar, 2018). Brain affected with the Gliomas have no survival rate and Gliomas are the most frequent affecting the cerebrum. Accurate identification of such malignancy are most pivotal part to remain alive and tumor purposes. Tumors are tend to have irregular shapes and spatially located in brain, so the detection of these tumors are difficult task. The most challenging part is the accurate segmentation. For this purpose an automated segmentation algorithm is designed using deep convolutional neural network. When data is sparse over-fitting is always an issue, to lessen the possibility of over-fitting max-out and drop-out are used. The suggested algorithm performs preprocessing during which pictures square measure normalized and bias field corrected and post process wherever tiny positives square measure removed exploitation morphological operations (Hussain, 2017). In past few years, deep learning has become popular for solving problem in various fields as well as medical imaging analysis. For diagnose purpose of Alzheimer’s disease and its stages using magnetic reasoning imaging deep convolutional neural network is used. Alzheimer’s disease cause damage to memory and cells of brains. The diagnosis of Alzheimer’s in elderly people is quite difficult and requires a highly discriminative feature representation for classification due to similar brain patterns and pixel intensities. The proposed technique results in a prediction accuracy of 98.8%, which is a noticeable increase in accuracy as compared to the previous studies and clearly reveals the effectiveness of the proposed method (Farooq, 2017). The essential indicator for diagnosing of cerebrovascular diseases are Cerebral Micro bleeds (CMBs). Physical identification of CMBs liable to flaws and for to reduce chances of errors this technique is presented for CMB identification. Three phases of this method are brain extraction, initial candidates extraction on threshold and filter dimension and feature extraction and classification affected one from healthy one with help of Support vector machine (SVM), Quadratic discriminant analysis (QDA) and ensemble classifiers. The validation of the method is done on dataset of 10 subject with CMBs in 6 for training and 4 for testing (Ateeq, 2018). Different types of brain tumors occur in brain like Primary cerebrum cancer, Gliomas cerebrum cancer, benign cerebrum cancer and malignant cerebrum cancer. Among these tumors, Gliomas brain tumor is most common and leading to short span of life for human. For the detection purposes of brain tumor, Medical Reasoning Imaging technique is used which produces large amount of data and manual segmentation is prevented in limited time which could result in death of patient (Pereira, 2016). The challenging thing in the identification of cerebrum cancer is segmentation, registration and modelling as tumor causes morphological changes in the brain (Bauer, 2013). Different segmentation algorithms work best for different sub-parts but not a single algorithm devise for the purpose that could be implemented and work best for each sub-region (Menze, 2015). Image segmentation always remain a challenging problem due to the cluttered object, image noise, image contents and other factors. K-means clustering integrated with Fuzzy C- means algorithm helps us in minimal computation time and in accuracy of detection (Abdel-Maksoud, 2015). Brain tumor could appear anywhere in brain in different size, shape and contrast. Convolutional Neural Network is highly efficient which exploits both logical and contextual feature (Havaei, 2017). Recognition of cerebrum cancer from MRI is a demanding job due to its location, shape and size. Intensity similarities between brain injuries and traditional tissues, few create use of multi-spectral structural imaging scans. Time and value for multi-spectral anatomical imaging scans area unit high. So, the proposed technique should segmenting successfully brain lesions with high correctness and low computational time (Nabizadeh, 2015). Image segmentation demonstrated to be a strongest tool as it separate the objects from the background. A better edge detection algorithm for brain tumor detection is powerful technique for segmentation as it appear to distinguish between different objects on basis of their edges. For this purpose Sobel edge detection combines with thresholding method to find different regions using closed contour algorithm (Aslam, 2015). The target is to detect tumor in the brain using image processing techniques as detection from MRI is quite time taking and difficult. Major steps in image processing includes image pre-processing, segmentation of pathological tissues, normal tissues and gray matter and cerebrospinal fluid, feature extraction from every segmentation and classification is done with Neural Networks (Damodharan, 2015). One image processing technique for the identification of brain tumor is hybrid self-organizing map(SOM) with Fuzzy K-means(FKM) which is efficient in term of specificity, sensitivity, mean square error, time of computation and memory demand (Vishnuvarthanan, 2016). Brain tumor detection become apparent in biomedical imaging. Detection is critical as tumor morphological changes remain subtle. Watershed segmentation is utilized to distinguish unusual tissues from normal surrounding tissues with connected component labeling for the better identification of brain tumor (Dhage, 2015). To overcome limitations faced by clinical experts and criticality of image segmentation as image segmentation proved to be difficult part in brain tumor detection Berkeley wavelet transform based image segmentation is used. For the precision and standard rate of support vector machine based classifier used to extract relevant features from every segmented tissue (Bahadure, 2017). For brain tumor detection, a hybrid method based on combination of support vector machine (SVM) and Fuzzy c-means is used. Detection of brain tumor is done in few major steps which includes image enhancement using contrast improvement, double thresholding for skull striping, Fuzzy c-means for segmentation to identify suspicious regions in brain MRI and Grey level run length matrix for the feature withdrawal and support vector machine is utilized for classification (SIngh, 2015). In medical imaging, image segmentation plays an important role and in past few years it become a popular area of research. Thresholding is common approach with the morphological operations for the identification of tumor. Each type of tumor cannot be identified by this methodology so the region growing technique is also used for this purpose. Non negative Matrix Factorization is another technique which has been used for text mining pattern analysis is used for detection too. It is primarily used as uninterruptable decomposition approach for detection of neoplasm and to any classify into varied sorts and additionally for feature extraction (Prajapati, 2015). Segmentation is one of the main method to identify brain tumor detection while using this technique many of the work is done threshold segmentation is type of segmentation is also used along with watershed segmentation and morphological operators. Threshold segmentation is type of segmentation in which we portioned the pixel values depending on their intensity values global thresholding and variable thresholding is further types used in threshold segmentation. Morphological operations apply a structure element to the image we take as an input which will create the output image of the same size then it will compare each value of input pixel value with output pixel value along with its neighbors with the help of this you can construct morphological that is sensitive to the particular shape in input image. Dilation and erosion is further types of morphological operators we are talking about. By apply above techniques we will able to take the image of humans brain MRI images and apply these segmentation techniques to detect the tumor in human brain. These techniques will help us to give us better result than previous techniques results of the experiments which are conducts using these techniques are outstanding as all they are easy to execute (Mustaqeem, 2012). Sometimes there is problem that occur that using some brain tumor detection techniques they will not differentiate between healthy tissue and the growing tissue which cause tumor in brain for this type of problem as a solution automatic magnetic resonance (MR) segmentation by using this type of technique we will be able to overcome many problems we face like segmenting localization and measure the volume of different brain tissues. Healthy brain tissues which we want to differentiate from faulty tissues are classified in to three categorize on the basis of broad tissue types these are grey matter white matter and d cerebra-spinal fluid the classification can be done manually but if you are doing automatically then it will be much more efficient (Fletcher-Heath, 2001). If we are able to understand the texture of brain and able to analyze it we will be able to get much more efficient and better result it have different application including segmentation and differentiate between healthy and faulty tissues.

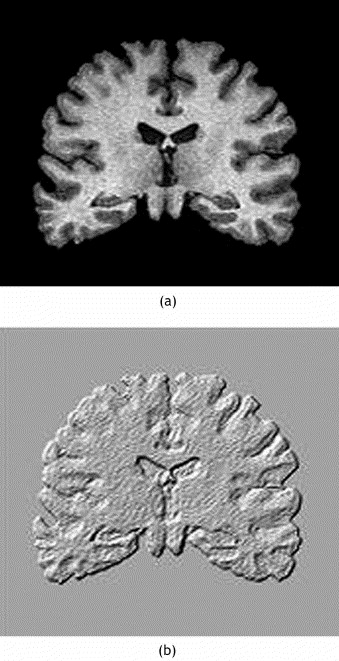


Figure 1 Brain Pathology

Many methods are used to construct the texture. The structural methods using well define primitives for example square shame will be describe as straight lines the main advantage of this method is very helpful in representation of figurative explanation of the picture**.** The model based methodsusing sophisticated approach to show appearance in an picture mathematical model such as fractal or stochastic these replica are used to analyze the picture only drawback of this type of method is mathematical complexities to estimate these parameters

**The transform method**

In this method we Will able to find texture of an image using frequency and scale space these methods comprise on Fourier Gabor and wavelet. Wavelet transform is that frequently used attributable to the benefit with that it should be adjusted to the matter in question. There are sudden variables which are used for finding parameters it is in vast amount used in clinical practice. If we will use the magnetic resonance image technique the problem is that it will not provide us microscopic information which will help us to examine the problem visually. Most commonly used parameters to find texture information’s are following

**Histogram**

in any digital image the pixel are some integer value ranging from 0 to 2a-1 a is stand for number of bits in picture in histogram image grey-level value span from 0 to 255 but in medical image it utilize 12 bits fpr this the range of grey-level is from 0 to 4095

**Absolute Gradient**

The measurement of spatial variation of grey-level is known to be gradient if there is sudden change in any image from white to black the value of gradient will be high at that point similarly if intensity value change from black to white smoothly to light grey the gradient value will be low. The value of gradient can also be positive and negative which depend mainly on if the value is shifting from light to dark or dark to light

**Wavelets**

It deals in terms of frequency means if one dimensional signal changes quickly with respect to time the frequency will be high similarly if it changes slowly its frequency will be low. If there is an grey scale two dimensional image and it changes fast then there will be many small differences within small chunk of picture we will link excessive spatial frequency to this chunk if the grey-level values changes slowly than the region has low spatial frequency this depend on the scale of image the larger the variation in the image the greater the possibility to find spatial frequency also the dimensions of the image is also very important.

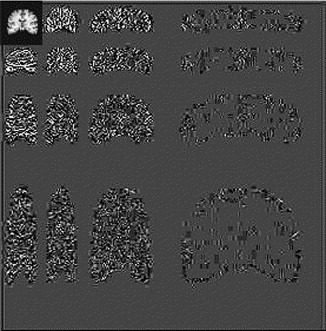


Figure 2 wavelet

In wavelets we analyze the frequency of the image along with divergent scales of the that particular picture after this inspection we will able to find contrast wavelet co-efficient which will correspond to divergent scales and also divergent frequency directions when we will calculate the wavelet transform of the picture we will connect the every pixel to set of digits which will able to distinguish the frequency contents of the picture at that point along the whole scale. In wavelets we analyze the frequency of the image along with divergent scales of the that particular picture after this explanation we will able to find contrast wavelet co-efficient which will correspond to divergent scales and also divergent frequency directions when we will calculate the wavelet transform of the picture we will connect the every pixel to set of numbers which will able to characterize the frequency contents of the picture at that point along the whole scale (Castellano, 2004). Image fusion has also a very important role in finding a tumor in human brain which help us to find and also differentiate between cancerous and non-cancerous region of the brain. This is the type of technique in which we take many image of the brain and diffuse them to single image to have a single view in this view well will reduce all the redundancies and extract only the useful portion of the brain. Once we create the single view then will apply the feature extraction techniques (Hagargi, 2018).

# Methodology

Artificial intelligence is greatly impacting almost every domain of life and eases the task for us to much extinct. For our current research, we have used the Machine Learning technique of Artificial Intelligence. Convolutional Neural Network is used to train and test the model and the required results are obtained. Convolutional Neural Network comprises of different convolution layers through which dataset is passed and each layer performs some operations on the dataset and pass data to its next connecting layers. Artificial intelligence (AI) is the recreation of human intelligence measures by machines, particularly PC frameworks empowering it to try and copy human conduct. Its applications lie in fields of Computer Vision, Natural Language Processing, Robotics, Speech Acknowledgment, and so forth Benefits of utilizing AI are further developed client experience, speed up market, foster modern items, empower cost advancement, improve representative usefulness also, work on functional effectiveness. AI (ML) is a subset of AI which is modified to think all alone, perform social communication, take in new data from the given information and adjust just as improve with experience. Despite the fact that preparation time through Deep Learning (DL) techniques is more than Machine Learning strategies, it is repaid by higher exactness in the previous case. Likewise, DL being programmed, huge space information isn't needed for acquiring wanted outcomes not at all like in ML.

## 2.1 Dataset Collection

The selected dataset for this research is the brain tumor dataset which comprises two different classes of brain tumors. The dataset folder with class name ‘no’ contains images dataset having no brain tumor and the dataset with folder name ‘yes’ contains images dataset with a brain tumor. Each dataset folder contains many relevant images like the ‘yes’ folder contains different images having brain tumors from different dimensions. The more images we have, the more efficient results we have. We have almost more than 300 images in the dataset on which we will train the model.

## 2.2 Model Selection

Before the discussion of model selection, it is important to understand what the model actually is. Model is a network of different layers through which the input data is passed and at the output layer, the data is in the refined form. Basically, there are three types of layers named the input layer, hidden layer, and output layer. The input data is feed to the input layer which can be in any form like CSV or images. The input layer is further connected to the hidden layers. The input data at the input layer is then passed to the hidden layer. There can be as many hidden layers in the network as required according to the nature and type of data. The hidden layers behave as filter layers as each layer filters out specific properties from the data and pass it to the next connected layer which further carries out the process (Bakes, 2018). Finally, in the output layer the input data is in the refined form with all the specific properties filtered out.

A sequential model is used to train the model on the brain tumor dataset. In a sequential model, all the layers are sequentially interconnected with each other. Like other models, the sequential model also contains as many hidden layers as required which are sequentially interconnected with each other through which input dataset will be passed. In short, the sequential model is a linear stack of layers.

## 2.3 IDE and Libraries Used

Python is the most used language nowadays for coding different Artificial Intelligence algorithms. The benefit of using Python is that there are different built-in libraries that are compatible with python for different purposes, we just need to import these libraries and use their functionalities within the code, and get the job done. There are different IDE’s (Integrated Development Environment) available for python but we are using Jupyter Notebook, the benefit of using Jupyter Notebook is that we can divide the code into smaller blocks and each block will be executed separately. We just need to focus on the specific block of the code which is executing. In addition to that, we do not need to execute all the blocks again in order to run the code. Moreover, we can also add notes blocks in the Jupyter Notebook. In a nutshell, Jupyter Notebook is a complete package of managing the code and notes in the form of blocks and we just need to focus on the currently executing block rather than executing the code from start every time we run the code.

As mentioned earlier, libraries make the task easier to perform. Libraries contain built-in functions that are coded by the programmers, we just need to import the libraries and use the built-in functionality of the library rather than code that functionality from scratch, which saves time and effort. There are different libraries that are being used in the code, the purpose of each library is defined below:

* Numpy library is used to deal with arrays i.e. 2D and multi-dimensional arrays and their manipulation. Numpy also deals with linear algebra, Fourier transformation.
* Another library that is used is Pandas that is used for data manipulation and data analysis purposes. As our dataset is in raw form like all the images do not have the same dimensions, so before feeding the data to the model, the dataset is first analyzed and manipulated to the same dimensions for linear results that will be done using Pandas.
* Math library is also imported as it provides some common math functions that will be used throughout the code in order to transform the data. Complex mathematical computations will be done using the Math library.
* Open CV library is also added in the row of imported libraries that mainly deals with image manipulation techniques like reading the image, change the color model of the image. Open CV is mostly used for Machine Learning, Computer Vision, and digital image processing. As our dataset is in the form of images so Open CV library plays its part.
* Matplotlib library is used to visualize the data in data in the form of charts or graphs. This library contains different types of built-in graphs and charts that can be used to visualize the data.
* OS library is also added in the project which is the short form of Operating System. OS deals with the directories like reading from the directory, creation, and deletion of the directories, etc.
* UMAP library is used for the purpose to reduce the dimensions of the image dataset that will be given as input to the model for training purposes.
* PIL is the short form of Pillow library that supports images of different formats that are not supported in the Open CV libraries. Moreover, PIL is also used for opening, manipulating, and saving images.
* Scipy library is used for the computation of mathematical and technical problems. This library is used in combination with Numpy.
* Sklearn library is also the most commonly used library for machine learning as it contains different efficient tools for regression, clustering techniques.
* TensorFlow is designed by Google that serves the purpose of a compiler for machine learning algorithms that helps in the execution.
* Keras library is used in combination with TensorFlow as it contains different models and layers that can be added to the model for model training.

When working with arrays, the numpy library is utilised, as this library is especially designed to deal with arrays. The Numpy library has a number of built-in methods for manipulating arrays. A matrix of arrays is created from the pictures, and numerous functions are performed to the matrix in order for it's appearance and behaviour to be altered as a result of this treatment. To play around with data and extract the desired findings, the Pandas library is used for data science purposes. Matplotlib is a library that specialises in data visualisations in many forms, such as bar charts, pie charts, etc. Directories are managed by the os library. Similar to Matplotlib, the Seaborn library provides several advanced visualisation methods that have proven to be useful when displaying data as graphs and charts. Using the Python Imaging Package (PIL) library, you may work with a variety of image formats. Image formats such as jpg, png and jpeg may be handled by the PIL library. Image manipulation and storing are also performed by this library. Torch and Tensorflow are both machine learning libraries. There are a variety of deep learning techniques in these libraries that may be used directly on the dataset without any further code required. Only the needed parameters must be sent into functions or methods offered by both libraries. For this simulation, skleran's skleran library is utilised. As for statical modelling, Sklearn library is responsible for a variety of regression and clustering algorithms. Sklearn also provides us with the layers that we need to add to the model; we simply need to provide them the appropriate parameters.

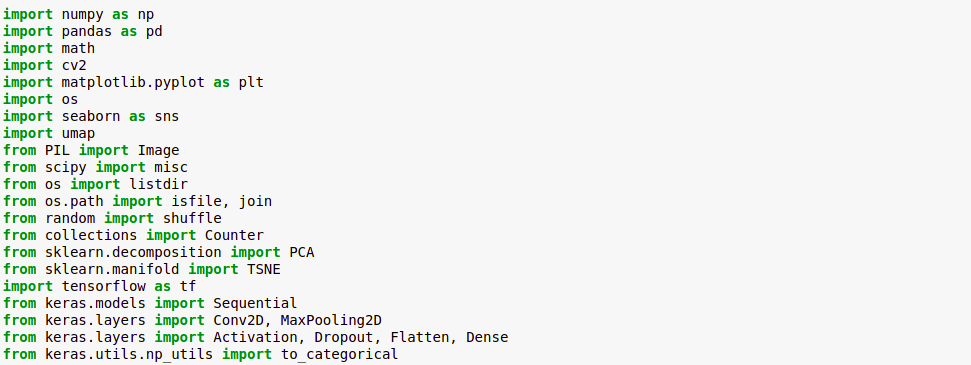
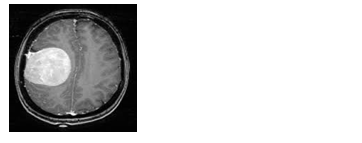


Figure 3Imported Libraries

## 2.4 Working

The working process begins with the dataset collection, as discussed in the section earlier regarding the selection and class names in the dataset. After the dataset is all set, the next step is to import all the necessary libraries that will be used in the code in order to get the job done. Detailed discussion on libraries is also done in the prior section. After all the libraries are imported, the dataset is also imported with all the directories and class names. Sample images are also shown in order to confirm that the dataset is imported correctly.

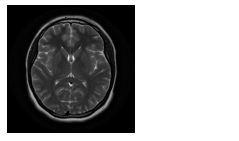


Figure 4 Healthy Tissue

Figure 5 Tumor Tissue

Images with tumors and without tumors are shown in the images above that are displayed from the collected dataset. All the images in the ‘yes (with brain tumor)’ class are stored in the separate variable and all the images with ‘no (without brain tumor)’ class are stored separately. All the images from both classes are read and stored separately. All the images from both the classes are converted into matrix form using Keras to\_categorical built-in function so that the manipulation operations can be easily performed on the data. After that, all the images are resized to the same shape as all the images have different dimensions and shapes.

### 2.4.1 Defining Model

Model is the most important part of machine learning. It is the model that is trained on the given dataset and make future predictions according to the training of the model. A model comprises of different layers which are categorized as input layer, hidden layer and output layer. Python Keras library contains built-in models and diverse layers which can be used to design a model according to the nature of the dataset. Keras sequential model is selected for training purpose. Sequential model is the linear stack of layers. Each layer in the sequential model is interconnected linearly and there is no crossing of layer in between the two layers which means the output of one layer will be the input of next layer connected to the previous layer. A neural network comprised of neurons that impersonates human being cerebrum like the person mind comprises layers of neurons that are connected with each other and shapes a neural network. Additionally, a neural network comprises neurons that are interconnected with one another and structure an entire network. Neural networks react to changing information that is given to them as picture or text document and the network yields the outcome with no prerequisite of updating the model. A neural network comprises three kinds of layers called as input, hidden and output layers. The data is passed to the input layer as an input which is then passed to the hidden layers which performs necessary required operations on the data and pass it to the output layer. There can be different hidden layers in one neural network depends on the sort of model utilized and the dataset. Every hidden layer extricates the particular features from the data according to the requirements and purpose of adding these layers. In the wake of separating the particular provisions from input information, the information is then passed to the following layer as info and the next layer removes different components. Toward the end, the refined information is gotten at the yield layer and that information can be additionally utilized for forecast purposes.

In a neural network, there are diverse initiation works that are being utilized to actuate the hub or not relying on the weighted total. For instance, in sync work, we set the edge esteem. Assuming the weighted total passes the boundary esteem, the hub will be initiated and yield is created for the following layer on the off chance that the weighted aggregate is underneath the edge esteem, the hub won't be actuated and there will be no yield for the following layer in the neural network. Likewise indirect capacity, we get a scope of yield esteems from the layer. The most well-known and every now and again utilized initiation in neural networks is sigmoid capacity. Sigmoid capacity is most ordinarily called as non-linear initiation work with the worth reaches from zero to one. Presently many datasets are accessible on the web for various purposes like AI, profound learning, and so on Cifar-10 and Cifar-100 datasets are utilized for research and the outcomes are deduced eventually. CNN has been introducing a usable class of models for better comprehension of substance present in a picture, in this way bringing about better picture acknowledgment, division, location, and recovery. Convolutional Neural Network (CNN) has a wide scope of utilizations that are being utilized for various purposes like picture acknowledgment, division, identification, signal acknowledgment, object recognition, computerized picture handling. Due to that reason model is firstly prepared for example dataset. The excess amount of images the dataset have, the more productive and exact the forecasts are detected by the model. First and foremost, the dataset is handled like resizing the pictures similarly, re-figuring the dataset pictures, and so on After the pictures are in refined form and fit to be taken care of to the model as information, the model or network is planned that comprises of various layers named input, covered up and yield layers containing distinctive actuation capacities in it. At every layer, the elements are removed and the refined dataset is passed to the preceding layer where more features extraction continue to happen. At the output layer, we have the filtered dataset separated, and the model is prepared to predict the images. Some additional layers that are included in the model are the convolution layers, these additionally added layers act as sifting layers that channel a portion of the components of the dataset. Additionally, another layer that is added is the pooling layer. Enormous size pictures are given as a contribution to the pooling layer and the pooling layer is liable for contracting of the huge measured pictures, however, the significant data in the pictures are safeguarded (Srinivas, S., Sarvadevabhatla, R. K., Mopuri, K. R., Prabhu, N., Kruthiventi, S. S., and Babu, R. V. (2016)). Contrasted and the consequences of other profound Convolutional Neural Networks (U-net, SegNet, Deeplab-V3), the aftereffects of the profound Convolutional Neural Networks CCN technique worked on a great deal, and th generally precision and mean F1 score came to 80% and 86%, separately. Via cautious determination of the layers we can enhance the expense work and work on the effectiveness and exactness of the model. After that, the pictures are formed and standardized utilizing change create and standardize capacities bypassing the legitimate contentions to the techniques. In the wake of normalizing the pictures, the dataset is separated into preparing, testing, and approval segment, with the goal that a portion of the dataset pictures are utilized for preparing the model, a few pictures for testing of the model, and a few pictures for approval reason. After effective control of the dataset pictures, the respective model is selected and different layers that are added to the model utilizing sklearn library is imported in the starting. The model chose for preparing reason for existing is consecutive. The advantage of utilizing consecutive models is dataset is gone through the layers in a succession and eacch layer performs its required actions on the dataset. Different layers that are being added to our sequential model and dataset needs to go through it in a clarified straightaway manner. Convolution layer is also added to the model, convolution layer (Conv2D) contains various channels that are liable for separating various components. Relu is an initiation work that is added to the consecutive model. Relu enactment work is a straight capacity that straightforwardly yields the info if the information is positive and yields zero if the information is negative. ReLU actuation work is the most regularly utilized default work in each model of profound learning as a result of its better productivity and improved execution that aids in the preparation of the model. The model additionally contains max pooling layer called as MaxPool2D work that down examples the contribution as indicated by the most extreme qualities in the lattice. Steps are changed along each course in the grid and supplant with the most extreme worth as indicated by the cushioning and refine the state of the information.

Keras library also provide built-in layers that can be added to the model as per requirement. The layer in the sequential model is named as Conv2D which has the purpose to create convolution kernel which helps in producing tensors of output (Erihov, 2015). After Con2D here comes the MaxPooling2D, in MaxPooling2D different pools of the matrix are created according to the given pool size. After the pools are created maximum values from that pool are separated into another matrix. The main purpose of MaxPooling2D is to reduce the dimensions and specific features from the image are obtained. After that dropout layer is added, the main purpose of adding this layer is that the model will be over fitted. These mentioned layers are added as many times as required with different parameter values like filters, padding, kernel size, activation function, input size, etc. At the end of the model flatten layer function is added which has the purpose to convert the pooled featured map to a single column that is then passed to the fully connected layer (Simonyan, 2014). The purpose of a dense layer is to feed all the outputs of a layer to the fully connected next layer. There are also different activation functions that are also added to the layers, the purpose of adding these activation functions is to make the model learn fast. Different activation functions include relu, sigmoid etc. The summary of the model is shown below:

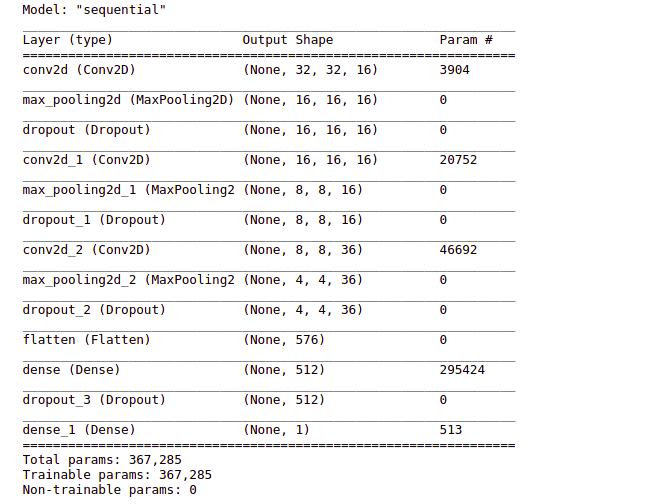


Figure 6 Model Summary

### 2.4.2 Model Compile and Training

After defining the model, the next step is to compile and train the model on the given dataset. Model is compiled using compile function by passing loss parameter of binary\_crossentropy, the optimizer is adam and metrics of accuracy. The model is then trained with a batch size of 128, epochs of 150. As the model training gradually progresses, model accuracy increases gradually after every epoch. The model training summary is shown:

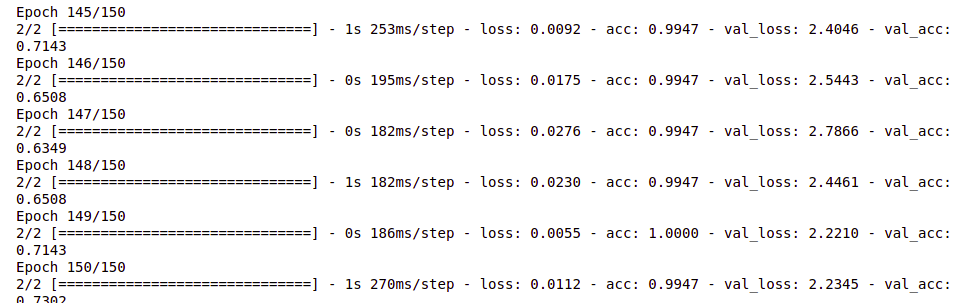


Figure 7 Model Training

The profound learning Convolutional Neural Network (CNN) calculation is utilized for our reenactment. CNN is utilized in light of the fact that it is explicitly intended for picture characterization, picture acknowledgment, face identification, and item location proficiently. As we are managing pictures data so CNN model meets our reenactments. Working of CNN model incorporates the model preparing segment where the model is prepared first and based on the prepared model it predicts the info picture being given and guides it on the classes on which the model is prepared. For instance in the event that we prepared the model on the classes of felines and canines, and we then, at that point give the picture of the feline to it. It effectively maps the picture to one of the classes of felines. On the off chance that we give the picture of a pony, it neglects to plan the picture on the prepared classes or guides the picture on the closest conceivable prepared class.

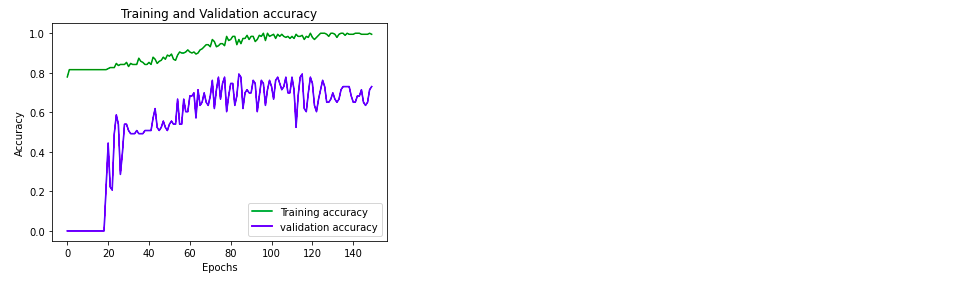
# 3. Results and Discussion

The dataset has been collected and arranged properly in the form of classes and proper naming conventions for classes have been adopted that can be easily understood by model and humans too. ‘Yes’ class name contain images with tumor brain tissues and ‘no’ class name contains healthy brain tissues images. Supervised learning approach is adopted as we have properly labeled dataset with proper classes and names. Different python classes are used in order to get the job done which includes numpy, matplotlib, tensorflow, sklearn, keras etc. All the libraries are imported and used in the code with proper understanding as discussed in earlier sections. Before training of model, the images in the dataset are first need some cleaning and manipulation so that data can be processed easily. The images are converted into matrix form after that necessary actions are performed like reshape and resize etc. The processed data is then fed to model input layer and refined data is obtained at the output layer. The model is then defined and dataset is given as input for model training. Sequential model is used for training purpose as it is the linear stack of input, hidden and output layers. All the layers are also defined while focusing on model selection. Each added layer serves the purpose of specific features extraction from input data and pass it to the preceding layer which is then processed further. Data at the end is in refined form at output layer. After the model is trained, it is is then put to test on some testing images to check the efficiency and competency of model. After model training is completed the accuracy of the model is found to be more than 70% that can be shown in the image below:

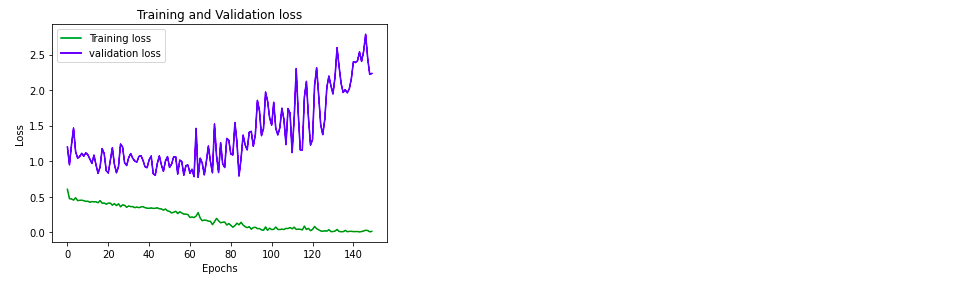


Figure 8 Accuracy

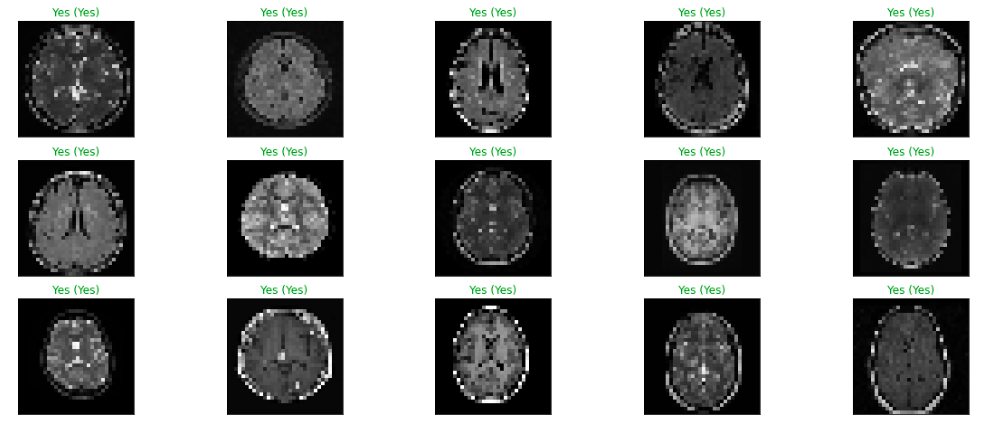
The model accuracy curve is shown in the image below:

Figure 9: Model Accuracy

The model loss curve is shown in the image below:

Figure 10: Model Loss

After the model is completely trained, the model accuracy is then tested on some sample images of brain tumors from the same dataset. The testing results are shown in the image below:

Figure 11: Prediction Results

The code can be practically implemented in the field of medical where brain tumor diagnosis will be more efficient and can be prevented from growing. In this proposed work distinctive clinical pictures like MRI cerebrum malignancy pictures are taken for recognizing tumor. The proposed approach for mind tumor discovery upheld convolution neural organization classifies into multi-facet perceptron neural network. The proposed approach uses a combination of this neural organization strategy and comprises of a few stages counting preparing the framework, pre-handling, execution of the tensor stream, arrangement. Later on, we'll take an out sized data set and look at to offer more exactness which can chip away at any kind of MRI cerebrum tumor. The recreation work is finished and the outcomes are acquired and imagined in various arrangements. Subsequent to examining the outcomes it is tracked down that the dataset is a tad buggy. Besides, the exactness of the model is likewise above 70%.

In the wake of accomplishing the reproduction work and examining the outcomes got from recreation, the outcomes are fulfilling as the exactness found to be above 70% and the plots reveals model loss and model precision. The chose CNN calculation of profound learning has produced delightful outcomes. CNN proves to be the good the methodology of profound training on the dataset as it turns out best for picture acknowledgment, object location, picture order and so on These all procedures depend on the picture dataset so my recreation work is the best illustration of that. Something else is that you need to choose the model layers as per the idea of images data and the provisions you need to extricate from the info picture information. As the model controls the information so the information is in a refined structure gotten at the yield layer that is additionally utilized for expectation purposes. Along these lines, the choice of layers in the model preparation is critical. The wrong choice of the layers may at last outcomes into helpless exactness of the model which further influences the model expectation gravely. While accomplishing the reproduction work I have seen something else that is significant which is the appropriate dataset that you will take care of the model with. The dataset ought to be kept up with appropriately, for example, there ought to be legitimate classes in the dataset and each class ought to have the information applicable to that class as it were. On the off chance that the dataset is ill-advised, the model will be prepared and it won't produce any blunder however it will influence the expectation by anticipating some unacceptable class due to the wrong model preparation on the dataset. Choice of profound learning CNN ends up being productive for the given dataset. As the dataset is made do with the appropriate classes and with applicable pictures, the expectation of the model turns out to be more proficient. The right selection of model for data training and required layers assist the selected model while preparing the model on the given dataset. The model precision and forecast of CNN are best generally, yet CNN has one disadvantage which is it will likewise show the outcomes regardless of whether it isn't pertinent to any of the classes in the dataset. On the off chance that the information picture is superfluous of the given dataset classes even, it predicts the class name which is closer to the information picture. It won't ever show any sort of blunder for insignificant pictures. Basically, the general presentation of CNN is awesome, yet the downside of CNN can never be overlooked as it might prompt the failure of the model. The overall performance of the model is acceptable and predicting the tumor with great efficiency.

# References

Abdel-Maksoud, E. a. E. M. a. A.-A. R., 2015. Brain tumor segmentation based on a hybrid clustring technique. *Egyptian Informatics Journal,* 16(Elsevier), pp. 71--81.

Anwar, S. M. a. M. M. a. Q. A. a. A. M. a. A. M. a. K. M. K., 2018. Madical image analysis using convolutional neural networks: a review. *Journal of medical systems,* Volume 42, p. 226.

Aslam, A. a. K. E. a. B. M. S., 2015. Improved edge detection algorithm for brain tumor segmentation. *Procedia Computer Science,* 58(Elsevier), pp. 430--437.

Ateeq, T. a. M. M. N. a. A. S. M. a. M. M. a. R. Z.-u. a. L. J. W. a. M. K. a. W. S. a. B. S. W. a. M. I., 2018. Ensemble-Classifiers-assisted detection of cerebral microbleeds in brain MRI. *Computes \& Electrical Engineering,* Issue Elsevier.

Bahadure, N. B. a. R. A. K. a. T. H., 2017. Image analysis for MRI based brain tumor detection and feature extraction using biologically inspired BWT and SVM. *International Journal of biomedical imaging,* 2017(Hindawi).

Bakes, 2018. *Identifying the best machine learn-ing algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the Brats challenge..* s.l.:s.n.

Bauer, 2013. *A survey of MRI-based medical image analysis for brain tumor studies.* s.l.:s.n.

Castellano, G. a. B. L. a. L. L. a. C. F., 2004. Texture analysis of medical images. *Clinical radiology,* 59(12), pp. 1061--1069.

Damodharan, S. a. R. D., 2015. Combining Tissue Segmentation and Neural Network for Brain Tumor Detection. *Intrernational Arab Journal of Information Technology (IAJIT),* 12(1).

Dhage, P. a. p. M. a. S. S., 2015. Watershed segmentation brain tumor detection. In: *Pervasive Computing (ICPC), 2015 International Conference on.* s.l.:IEEE, pp. 1--5.

Erihov, 2015. *A cross saliency approach to asymmetry-based tumor detection. In: medical image computing and computerassisted intervention–miccai.* s.l.:s.n.

Farooq, A. a. A. S. M. a. A. M. a. R. S., 2017. A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: *Imaging Systems and Techniques (IST), 2017 IEEE International conference om.* s.l.:IEEE, pp. 1--6.

Fletcher-Heath, L. M. a. H. L. O. a. G. D. B. a. M. F. R., 2001. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artificial intelligence in medicine,* 21(1-3), pp. 43--63.

Hagargi, P. A. a. S. D., 2018. Brain Tumour Detection and ART Classification Technique in MR Brain Images using RPCA QT Decomposition. *Brain,* 5(04).

Havaei, M. a. D. a. W.-F. D. a. B. A. a. C. A. a. B. Y. a. P. ,. C. a. J. P.-M. a. L. H., 2017. Brain tumor segmentation with depp neural nerworks. *Medical image analysis,* 35(Elsevier), pp. 18--31.

Hussain, S. a. A. S. M. a. M. M., 2017. Brain tumor segmentation using cascaded deep convolutional neural network. In: *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE.* s.l.:IEEE, pp. 1998--2001.

Menze, 2015. *The multimodal brain tumor image segmentation benchmark (BRATS).* s.l.:s.n.

Mustaqeem, A. a. J. A. a. F. T., 2012. An efficient brain tumor detection algorithm using watershed \& thresholding based segmentation. *International Journal of Image, Graphics and Signal Processing,* 4(10), p. 34.

Nabizadeh, N. a. K. M., 2015. Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features. *Coputers \& Electrical Engineering,* 45(Elsevier), pp. 286--301.

Pereira, S. a. p. A. a. A. V. a. S. C. A., 2016. *Brain Tumor segmentation using convolutional neural netorks in MRI images,* s.l.: s.n.

Prajapati, S. J. a. J. K. R., 2015. Brain tumor detection by various image segmentation techniques with introduction to non negative matrix factorization. *Brain,* 4(3), pp. 600--3.

Qayyum, M., 2017. *Medical Image retrival using deep convolutional neural network.* s.l.:s.n.

Simonyan, 2014. *Very deep convolutional networks for large-scale image recognition.* s.l.:s.n.

SIngh, A. a. o., 2015. Detection of brain tumor in MRI images, using combination of fuzzy c-means and SVM. In: *Signal processing and integrated networks (SPIN) , 2015 2nd International conference on.* s.l.:IEEE, pp. 98--102.

Vishnuvarthanan, G. a. R. M. P. a. S. P. a. V. A., 2016. An unsupervised learning method with clustering approach for tumor identification and tissue segmentation in magnetic resonance brain images. *Applied Soft Computing,* 38(Elsever), pp. 190--212.

# Appendices

## Appendix 1:

import numpy as np

import pandas as pd

import math

import cv2

import matplotlib.pyplot as plt

import os

import seaborn as sns

import umap

from PIL import Image

from scipy import misc

from os import listdir

from os.path import isfile, join

from random import shuffle

from collections import Counter

from sklearn.decomposition import PCA

from sklearn.manifold import TSNE

import tensorflow as tf

from keras.models import Sequential

from keras.layers import Conv2D, MaxPooling2D

from keras.layers import Activation, Dropout, Flatten, Dense

from keras.utils.np\_utils import to\_categorical

os.listdir('brain\_tumor\_dataset')

im =Image.open('brain\_tumor\_dataset/no/1 no.jpeg').resize((128,128))

im

im =Image.open('brain\_tumor\_dataset/yes/Y1.jpg').resize((128,128))

im

yes=os.listdir('brain\_tumor\_dataset/yes')

no=os.listdir('brain\_tumor\_dataset/no')

data=np.concatenate([yes,no])

len(data)==len(yes)+len(no)

target\_x=np.full(len(yes),1)

target\_y=np.full(len(no),0)

data\_target=np.concatenate([target\_x,target\_y])

len(data\_target)==len(target\_x)+len(target\_y)

len(data\_target)==len(data)

data\_target

data

yes\_values=os.listdir('brain\_tumor\_dataset/yes')

no\_values=os.listdir('brain\_tumor\_dataset/no')

X\_data =[]

for file in yes\_values:

img = cv2.imread('brain\_tumor\_dataset/yes/'+file)

face = cv2.resize(img, (32, 32) )

(b, g, r)=cv2.split(face)

img=cv2.merge([r,g,b])

X\_data.append(img)

for file in no\_values:

img = cv2.imread('brain\_tumor\_dataset/no/'+file)

face = cv2.resize(img, (32, 32) )

(b, g, r)=cv2.split(face)

img=cv2.merge([r,g,b])

X\_data.append(img)

len(X\_data)==len(data)==len(data\_target)

X = np.squeeze(X\_data)

X.shape

# normalize data

X = X.astype('float32')

X /= 255

categorical\_labels = to\_categorical(data\_target, num\_classes=2)

data\_target

(x\_train, y\_train), (x\_test, y\_test) = (X[:190],data\_target[:190]) , (X[190:] , data\_target[190:])

(x\_valid , y\_valid) = (x\_test[:63], y\_test[:63])

model = tf.keras.Sequential()

# Must define the input shape in the first layer of the neural network

model.add(tf.keras.layers.Conv2D(filters=16,kernel\_size=9, padding='same', activation='relu', input\_shape=(32,32,3)))

model.add(tf.keras.layers.MaxPooling2D(pool\_size=2))

model.add(tf.keras.layers.Dropout(0.45))MaxPooling2DMaxPooling2D

model.add(tf.keras.layers.Conv2D(filters=16,kernel\_size=9,padding='same', activation='relu'))

model.add(tf.keras.layers.MaxPooling2D(pool\_size=2))

model.add(tf.keras.layers.Dropout(0.25))

model.add(tf.keras.layers.Conv2D(filters=36, kernel\_size=9, padding='same', activation='relu'))

model.add(tf.keras.layers.MaxPooling2D(pool\_size=2))

model.add(tf.keras.layers.Dropout(0.25))

model.add(tf.keras.layers.Flatten())

model.add(tf.keras.layers.Dense(512, activation='relu'))

model.add(tf.keras.layers.Dropout(0.15))

model.add(tf.keras.layers.Dense(1, activation='sigmoid'))

# summary

model.summary()

model.compile(loss='binary\_crossentropy',

optimizer=tf.keras.optimizers.Adam(),

metrics=['acc'])

history = model.fit(x\_train,

y\_train,

batch\_size=128,

epochs=150,

validation\_data=(x\_valid, y\_valid),)

# Model loss

loss\_train = history.history['loss']

loss\_val = history.history['val\_loss']

epochs = range(0,150)

plt.plot(epochs, loss\_train, 'g', label='Training loss')

plt.plot(epochs, loss\_val, 'b', label='validation loss')

plt.title('Training and Validation loss')

plt.xlabel('Epochs')

plt.ylabel('Loss')

plt.legend()

plt.show()

# Model accuracy

loss\_train = history.history['acc']

loss\_val = history.history['val\_acc']

epochs = range(0,150)

plt.plot(epochs, loss\_train, 'g', label='Training accuracy')

plt.plot(epochs, loss\_val, 'b', label='validation accuracy')

plt.title('Training and Validation accuracy')

plt.xlabel('Epochs')

plt.ylabel('Accuracy')

plt.legend()

plt.show()

# Evaluate the model on test set

score = model.evaluate(x\_test, y\_test, verbose=0)

# Print test accuracy

print('\n', 'Test accuracy:', score[1] \* 100)

labels =["Yes", # index 0

"No", # index 1

]

y\_hat = model.predict(x\_test)

# Plot a random sample of 10 test images, their predicted labels and ground truth

figure = plt.figure(figsize=(20, 8))

for i, index in enumerate(np.random.choice(x\_test.shape[0], size=15, replace=False)):

ax = figure.add\_subplot(3, 5, i + 1, xticks=[], yticks=[])

# Display each image

ax.imshow(np.squeeze(x\_test[index]))

predict\_index = np.argmax(y\_hat[index])

true\_index = np.argmax(y\_test[index])

# Set the title for each image

ax.set\_title("{} ({})".format(labels[predict\_index],

labels[true\_index]),

color=("green" if predict\_index == true\_index else "red"))

plt.show()s

## **Appendix 2:**

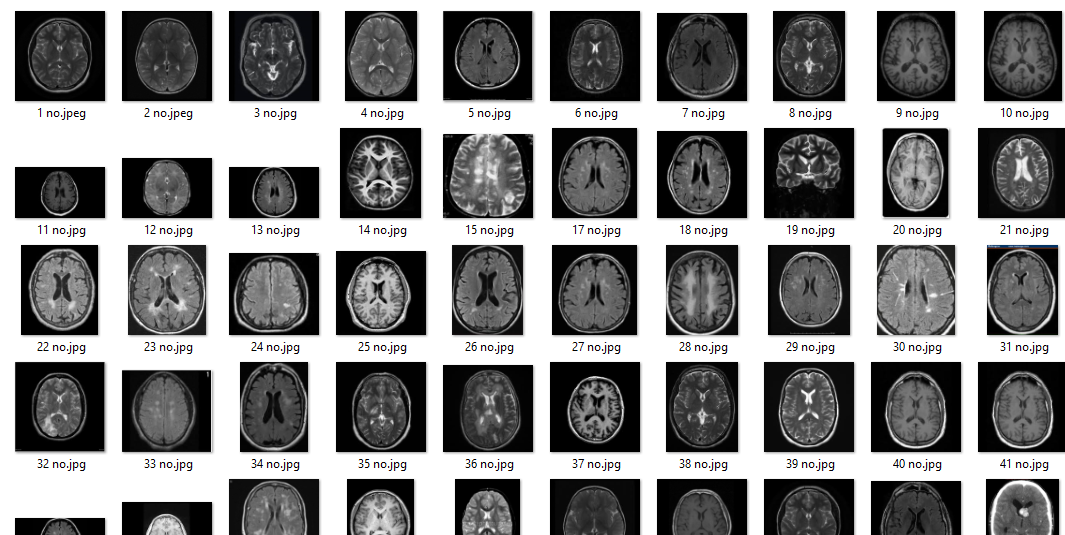
This appendix includes some dataset images downloaded from public resource kaggle.com   


Figure 12 Dataset images with no tumor

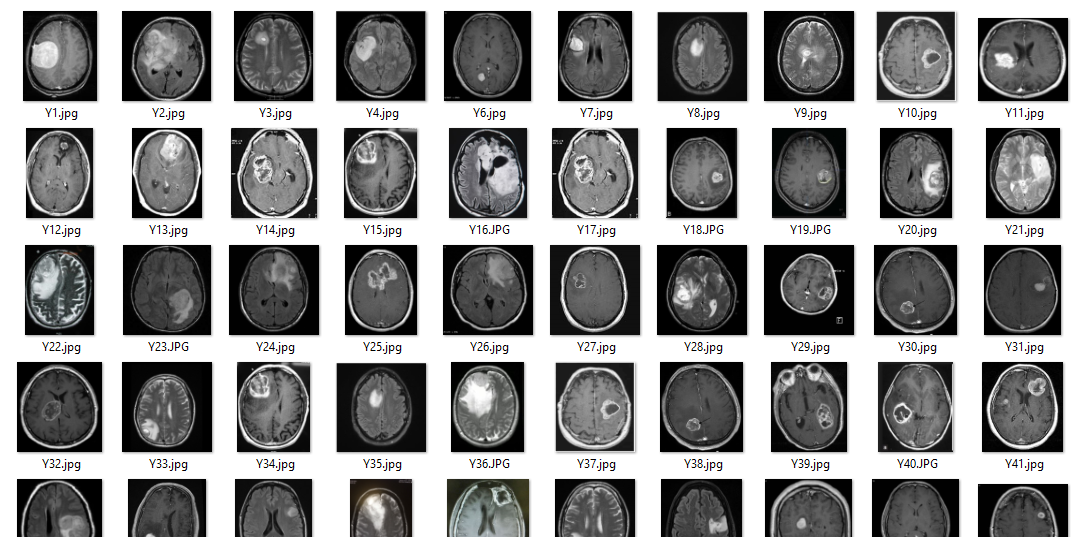


Figure 13 Images dataset with tumor