Detection of Brain Tumor using Machine Learning Approach

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Computer Science

Submitted By
Megha Chadha
(801732029)

Under the supervision of:
Dr. Sushma Jain
(Assistant Professor)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
PATIALA – 147004

June, 2019

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, **"Detection of Brain Tumor using Machine Learning Approach"**, in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science* submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my own work carried out under the supervision of ***Dr. Sushma Jain*** and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Signature
(Megha Chadha)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Sushma Jain)
Assistant Professor,
CSED, TIET, Patiala

Acknowledgement

The successful completion of any task is incomplete without acknowledging the people who made it possible and whose constant guidance and encouragement secured the success.

With the profound sense of gratitude and heartiest regard, I want thank my guide **Dr. Sushma Jain**, Department of Computer Science & Engineering, Thapar University, Patiala, who has been very concerned and has aided for all the essentials for the preparation of thesis work. She not only helped me in exploring the vast topic in an organized manner but also provided me with all the ideas of the work towards a research-oriented venture. I am also thankful to **Dr. Maninder Singh**, Head of Department and **Dr. Ashutosh Mishra**, PG Coordinator, CSE department for the motivation and inspiration in completion of the thesis work. I would be failing in my duty if I do not express my gratitude towards **Dr. S.S. Bhatia**, Dean of Academic Affairs for making provisions of infrastructure such as library facilities, computer equipped labs and latest technology so that we can pursue the research work positively.

Last but not the least, I would like to express a heartfelt thanks to my parents, the almighty and my classmates who with their thought provoking views, veracity and cooperation helped me in successful completion of my thesis work.

Megha Chadha

(801732029)

Abstract

Tumor in brain is one of the most dangerous diseases which if not detected at the early stage using accurate methods can even risk the life. Currently, the methods used by neurologists for analysis are not completely error free and states that manual segmentation isn't a good idea. Tumor is basically divided into two types: Benign tumor and malignant tumor. The study discussed both the types of tumor along with their symptoms and the ways tumor can be treated. The purpose of the thesis is to detect the tumor, segment it from the MR image of the brain and classify it into either benign tumor or malignant tumor.

The study presents machine based approach for segmentation of brain images using thresholding segmentation technique followed by the identification and classification of tumor into its types using CNN classification approach. Comparison between machine learning approaches is shown to compare the results in terms of accuracy and classification output. Results are compared to find a better machine learning approach which improves the performance; minimize the complexity and works on real time data.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1. Introduction to Brain Tumor

Tumor in brain is the uncontrolled mass of tissues on the brain. It is divided into tissues having cancerous cells called benign tumor, and tissues with non-cancerous cells called malignant tumor. Tumors fall under grade I to grade IV according to World Health Organization where grade I and grade II are called low grade tumors whereas grade III and grade IV are called high grade tumors. The study of e-health care systems and evolving technology in the recent times let radiologists provide better health care to the patients.

To detect infected tumor tissues from medical imaging modalities, segmentation technique is employed. Magnetic Resonance of Image (MRI) provides the detailed information about the shape, size and position of the tumor in which a contrast material is used to differentiate the outline boundary. It is done using a large cylindrical tube machine where the patient is made to lie down on a movable table with the head leading towards the magnet. The tube then uses a powerful magnetic field where radio waves are passed and monitor shows the detailed image of the tissues bones or organs of the brain. This procedure provides the basic anatomy of the brain to the radiologists who compare the distributions of bright and dark areas to conclude whether the tissue in brain is healthy or not. This process is painless, invasive and precise than other imaging methods.

The extraction of brain tumor requires the separation of MR images of the brain into two basic regions. The first region contains the tumor cells and the second contains the healthy brain tissues. Following are the types of tumor discussed:

1.1.1. Benign Tumor

These tumors are called grade I or grade II tumors and do not contain active cancer cells in the tissues of the brain. They have a uniform, homogenous structure and are only harmful in the cases where they press the vital structures such as blood vessels, veins or some nerves. It doesn’t generally spread to the other parts of the body, but still said to weaken the bones and overall functioning of the body. The body keeps on forming new cells when required but this tumor doesn't let it kill the old cells which should die [17]. The tumor is shown in Fig 1.1. Not in every case it happens that these tumors are removed. They may be pulverized as well. The tumor itself isn't painful, just a massive extra mass of tissues in the brain. But it may cause severe weakening of the bones leading to breakage which can be excessively painful. Causes of benign tumors are:

- Environmental toxins
- Intake of improper diet
- Genetic disorder
- Local trauma or any injury from infection

In many case, these tumors require no treatment, as doctors suggest. They may suggest watchful waiting to make sure it doesn't cause any damage. Surgery is yet the most common treatment for the brain tumor where the abnormal mass is removed without disturbing the neighboring tissues of the brain.

Other treatments for benign tumor are medication or through radiation. Once removed, the tumor usually doesn't grow back.

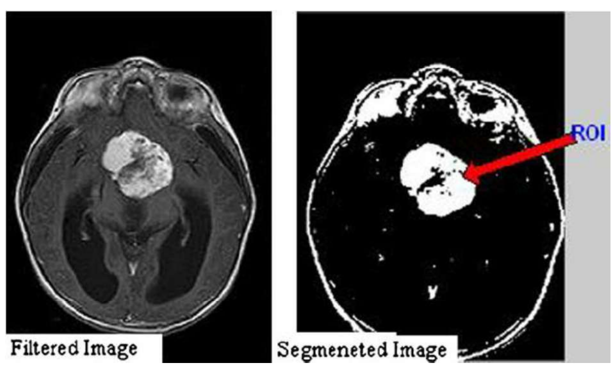

Figure 1.1: Segmented Benign Tumor

1.1.2 Malignant Tumor

Malignant tumors are referred as grade III or grade IV tumors and carry active cancer cells in the brain. They have a non uniform heterogeneous structure and high grade that spreads rapidly in the body [32]. It may also happen that they enter the lymph or the blood infecting other tissues of the body as well. If left untreated, they may also result to death as the abnormal cells multiply at a very fast rate. As early detected and started to be cured, the more the chances that the tumor will be cured. Depending upon the study of the tests of the brain or the MRI image study, radiologists decide the stage of the tumor and whether it will be curable with cancer surgery or not. Major symptoms of malignant tumor:

- Frequent headaches
- Fitz (seizures)
- Excessive vomiting and being sick
- Vision problems
- Change in personality

The tumors can be treated using surgery. In this process, the skull is pre processed and tumor is cut into pieces before it spreads, followed by the radio therapy. Another way is camustine implants where a new way of chemotherapy for some high grade tumor is introduced, and implants are inserted into the brain.

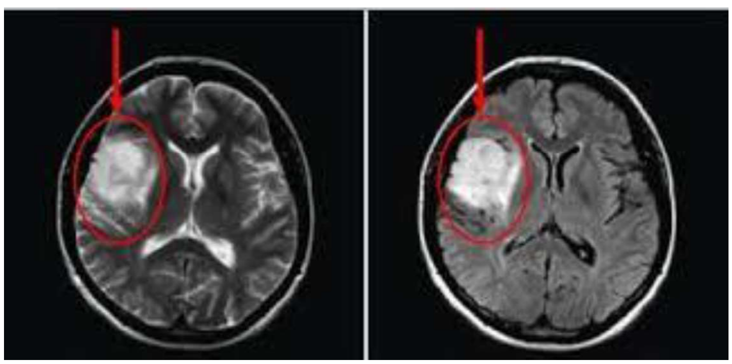

Figure 1.2: Segmented Malignant Tumor

1.2. Segmentation

Segmentation is the process employed to detect the infected tumor tissues by separating the image into different regions which share the common properties in the image such as color, texture, brightness, size, boundaries, gray matter etc. the process also involves separation of dead cells from the healthy brain tissues such as solid mass, white matter (WM), gray matter (GM), cerebral spinal fluid (CSF) with the help of MR images. Manual segmentation of the tumor by neurologists is a very difficult and time consuming task. Hence, segmentation process helps in early detection of tumor which contributes to the improvement in patient's health and survival rate of the patient. Segmentation techniques include manual segmentation, semi-automatic segmentation and fully-automatic segmentation. The technique is useful as:

- This technique is fast and accurate for brain tumor detection, competitive both in terms of accuracy and speed as compared to other methods.
- The method is based on Deep Neural Networks (DNN) that learns the features specific to brain tumor segmentation.

1.2.1. Types of segmentation

- *Edge based segmentation:* It represents a large group of methods based on information about the edges in the image depending on gray level, texture, color etc. It gives us the output as the boundaries of objects present in the image. It works on the basis of discontinuity in the brightness, since images contain a lot of data and most of their information lies in the edges.
- *Region based segmentation:* This method segments the image by splitting the image into smaller regions and further merging them into larger ones. Region based ha further two categories:

➢ Region Growing: It starts with tiny regions which form together to combine neighboring regions that are sufficiently similar to each other.

➢ Region Splitting: It starts with one big region, separating other regions that are not sufficiently homogenous.

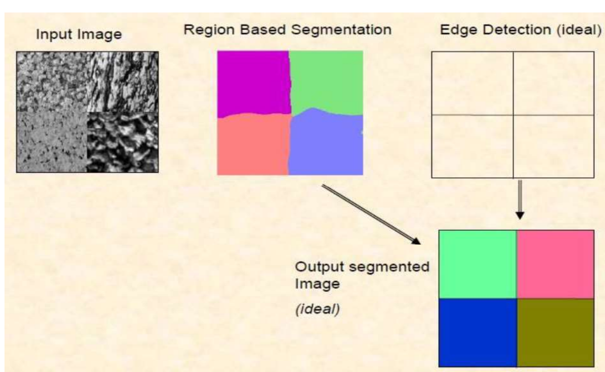

Figure 1.3: Edge vs Region based segmentation

- *Threshold based segmentation*: This segmentation is used to distinguish foreground from the background. We select a threshold value T, after the gray image is converted to binary image. The pixels are partitioned based on their intensity values. Pixels have value greater than the threshold value marks the region as white and the pixels having values less than threshold values marks the region as black.

Figure 1.4: Threshold based segmentation

- *Cluster based segmentation:* This technique aims at segmenting the image into clusters having pixels with similar characteristics. It is applicable to images with multispectral or grey level. Methods can be applied easily and are also extendable to high dimensional data. Clustering can be of two types:

➢ *Hard Clustering*: It is a type of clustering technique that divides the image into set of clusters where one pixel belongs to only one cluster.

➢ *Soft Clustering:* The pixels are partitioned into clusters based on partial membership where one pixel can belong to more than one clusters.

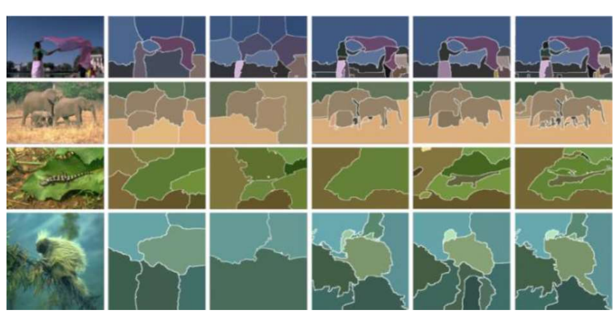

Figure 1.5: Cluster based segmentation

1.3. Machine Learning

1.3.1. Support Vector Machine (SVM):

SVMs are supervised learning models which are used to analyze the data for classification as well as regression. It was invented by Vladimir N. Vapnik which was later modified by Carpis and Vapnik in 1993. If a training dataset is given, the purpose of SVM is to divide the non-linear transformation into a linear transformation using kernel functions of SVM. SVM has kernels namely linear non linear, sigmoid type, polynomial, Gaussian, RBF etc. kernels takes the data in the raw form and transforms them into useful outputs. In the implementation, the kernel that I've used is Gaussian kernel which has made the classification very easy and convenient [15]. SVM works with hyper planes dividing the dataset into two parameters which maximizes the margin between the non overlapping parameters. The performance in SVM is measured by accuracy, sensitivity and specificity. We've chosen SVM because it works best even when the data is not linearly separable. It is basically defined by a distinguishing hyperplane. The training dataset gives us the output as a hyper-plane which finds a boundary between the two possible outputs.

1.3.2. Convolutional Neural Network (CNN):

When a device takes in an image, device sees it as an array combined of pixels which depends on the image resolution, typically the height, width and dimension. The array of matrix size 6x6x3 has RGB values whereas an array of matrix size 4x4x1 refers of an image with grey scale. CNN is used to identify images and the objects in an image, used for classification of images, detection etc. The process depends on passing each image through a series of convolution layers with kernels which act as filters, pooling and a series of fully connected layers.

- *Convolution layer:* It is the prior layer out of all the layers which is essential to extract the features from the image. In this layer, the convolution matrix is multiplied with the filter matrix referred as feature map.
- *Pooling layer:* When the image size is too large, this section of layer helps in reducing the parameters, retaining all the important information intact.
- *Fully connected layer:* In this section of layer, we flatten our matrix in the form of a

vector and pass it to the fully connected layer in the form of a neural network and all the features are combined to form a model.

1.4. Performance Parameters

The parameters which we use to achieve the performance in our model are described below:

1.4.1 Sensitivity

Sensitivity is defined as the degree of total true positive or positive cases which are predicted as true. It is also known as recall. It is given as the ratio of True Positive (TP) with the sum of True Positive (TP) and False Negative (FN). True Positive means patient having cancer is actually suffering from cancer and False Negative means person having no cancer are predicted as having cancer.

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

1.4.2 Specificity

Specificity is defined as the degree of actual negative value and is predictive as negative. It is the ratio of True Negative with the sum of True Negative (TN) and False Positive (FP). True Negative means person not suffering from cancer has no cancer in actual. False Positive means the person predicted as having cancer has actual no cancer.

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

1.4.3 Precision

Precision is defined as the percentage of actual results which are true or relevant. It is given by the ratio of True Positive (TP) with the sum of True Positive (TP) and False Positive (FP).

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

1.4.4 Accuracy

Accuracy is defined as the percentage of Predicted results which are predicted correctly with respect to actual values and is given by following eq 13.

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

1.5. Chapterization

The rest of the thesis is organized as given below:

Chapter 2 gives a detailed view of related work done based on brain tumor detection using machine learning algorithms.

Chapter 3 defines research gaps collected from literature survey and problem statement of the work. The methodology adopted is also discussed in this section.

Chapter 4 describes the proposed work for brain tumor segmentation and classification using extreme machine learning.

Chapter 5 describes implementation and result obtained by applying machine learning algorithm for detection of tumor as well as the future scope of the research.

1.6 Summary

This chapter provides detailed information about brain tumor detection, types of brain tumor, their diagnosis and treatment, suitable machine learning approaches and different parameters used.

Chapter 2
Literature Survey

Kaur *et al.* [1] proposed a research to partition the image into meaningful parts having same features and properties. Author explained 4 segmentation techniques for the research in the field of medical image applications. It is clear that not every method is suitable of any particular image dataset. Different datasets require different segmentation techniques as per their type. Some useful information that has been concluded is:

- Threshold segmentation is the simplest method as it doesn't require any previous information.
- Edge based segmentation is good for images having better contrast between the objects and isn't suitable for image having too many edges.
- Region based segmentation is more immune to noise but it is an expensive method in terms of money and memory as compared to other methods.
- Cluster based method is more useful for real time data problems. Partial Differential Based method is the fastest method and best for time critical problems.

Ism *et al.* [2] proposed an approach for layer by layer segmentation and then applying filters on them. Comparison is shown between manual segmentation process, fully automatic segmentation process and semi automatic segmentation process. These are applied on the dataset and calculated whether the patient has whole tumor, core tumor or active tumor in the form of percentages. This was done by taking the patches of the skull and then using trained filters on them. The classification is done using some DNN techniques and traditional methods. The modern method of CNN helps in automatically learning complex features for both normal and abnormal brain tissues simply from the MRI image. However, traditional methods used prior information and then translated it into probabilistic maps which was a very time consuming and a difficult task.

Yuheng *et al.* [3] concluded some image algorithms that are a combination of multiple segmentation techniques very effectively. It makes full use of the advantages of different algorithms on the basis of multi-feature fusion, so as to achieve better segmentation effect. A single segmentation algorithm cannot be adapted to work with all the images. Study shows the comparison of various segmentation techniques and algorithms and concludes that it is still not perfect and has some practical problems in the applied research. The study also suggested proper parameter selection methods using machine learning for the analysis, such as finding the value of k in k-means algorithm.

Kaur *et al.* [4] presented different image detection and segmentation techniques in depth for the segmentation of an image. The application of image engineering can be medical imagine, image recognition, etc.

Vaithegi *et al.* [5] worked on image binarization which is the process of separation of pixel values into two groups, black as background and white as foreground. According to this categorization, thresholding segmentation is done. The study discusses various thresholding methods which are categorized as global thresholding methods and local thresholding methods. This technique was implemented on medical images that remove background by using local mean and standard deviation.

- Global thresholding methods are used when the intensity distribution between the foreground and background of the object is very prominent and clear. Simply a numerical value can be taken and it is sufficient to distinguish the foreground and the background. This value of the threshold usually depends on the pixel value or the grey scale level of the image.

- Local thresholding methods aims at calculating the threshold of each image separately by taking in account features like mean, variance or intensity of the pixels. This method is a little time consuming and complex as the technique depends on contrast, pixel values, individual image characteristics, region size of the image etc.

Kumari *et al.* [8] proposed a research on the classification and feature extraction techniques. Support Vector Machine (SVM) classifier is a binary classifier which is used to classify the extracted features on an image into further classes. In this study, SVM was used on 110 brain images, where out of them, mean value was calculated and which gave a rough idea if the patient has healthy brain tissues or not, along with the type of abnormality if it exists. The ratio of normal and abnormal brain images taken were 60 and 50, where in the 50 images consisted of some cuts or bleed showing the abnormal activity in the MR image of the brain. SVM classifier produced better results with improved sensitivity, specificity and accuracy. The execution time was improved as well. The study further proposes that better results can be achieved with increased dataset and by extraction of more features.

Damodaran *et al.* [10] presented a comparison of classification techniques using ANN, k-means and NN individually. The study was implemented on MATLAB where in a dataset of total 20 images was taken out of which 10 images of brain have tumor and rest 10 don't have tumor. The method segments and detects tumor inside the brain, including normal tissue (WM and GM part) and fluid from the images of the brain. Best accuracy of 80% was given by validating the results of k-NN technique and Bayesian classification all together separating the tissues of the brain into WM, GM and CSF.

Devi *et al.* [11] proposed a technique of extreme machine learning for classification of brain tumor from 3D MR images. Basic differential evolution is adopted to extract appropriate features, which is further classified under refined group search optimizer to carry out the classification process. The 3D process was carried out and study showed that this method produces minimum number of features with higher classification accuracy when compared with conventional 2D and 3D feature extraction methods. This method obtained an improved accuracy with good sensitivity and specificity rate.

Alwan *et al.* [12] proposed difference between two techniques that are Arnold Scrambling and Berkeley Wavelet Transformation in digital image processing. These techniques produced the difference between the original image and watermarked image. The Berkeley wavelet transformation method is used for images with grey scale. To boost the security purposes, the image was first scrambled using Arnold transformation which showed very good results in terms of peak signal and noise ratio.

Yin *et al.* [13] presented a study on the classification of images in which images are detected and classified on the basis of object tracking, behavior analysis, computer vision; image segmentation etc. research concluded that the existing image classification methods can be useful for every single feature individually but it can lead to large number of redundant information and low accuracy comparatively. Hence, use of an integrated method using SVM might give improved results in less time.

Chadda *et al.* [15] proposed a technique for automatic feature extraction for brain tumor detection based on Gaussian mixture model using MR images. The goal was to implement three classifier techniques that are naïve bayes, SVM and probabilistic neural network. Naïve bayes was implemented using the kernel method approximated the compled distribution of data, further SVM was implemented using Gaussian base function and finally the fastest technique was applied which is probabilistic neural network. Overall, the method worked on the principal component analysis and wavelet based features, the performance of the GMM based feature extraction is enhanced. The accuracy achieved was improved according to the study.

Demirhan *et al.* [16] presented a new tissue segmentation algorithm using wavelets and neural networks, which claims effective segmentation of brain MR images into the tumor, WM, GM, edema and CSF. The goal was to achieve higher classification rate to diagnose normal and abnormal tissues in the brain. The steps followed are as follows:

- The image was firstly pre-processed to set mean intensity of the pixels at the same level. This helps in lightening the darker images to maintain the intensity.
- Wavelet transformation is then adopted to extract features from the image. It helps in analysis of image at various levels of resolution.
- The number of features is to be reduced with the help of principal component analysis. It also gives the result as reduced discrimination between among the classes by reducing the dimensionality of data thus reducing the computational cost of the analysis.

The proposed model was appropriate for the diagnosis of tumor through MR images showing correct classification of more than 93% cases in k-means clustering and 98.5%

in case of Probabilistic neural network [19].

Tohka *et al.* [17] presented research work on partial volume effect (PVE) on segmentation of an MR image of the brain. It stated the concept of voxels in the brain as each voxel contains multiple types of tissues inside it, such that a single voxel of the brain may contain various different kinds of tissues. Due to the acute and complex structure of the brain, PVE is required as it generalizes the segmentation process where each tissue type is solved within each voxel. This provided a systematic and symmetric way of segmenting the image. This approach provided improved accuracy.

Roslan *et al.* [18] examined a methodology where mean was calculated for each image sequence for around 90 MR images.

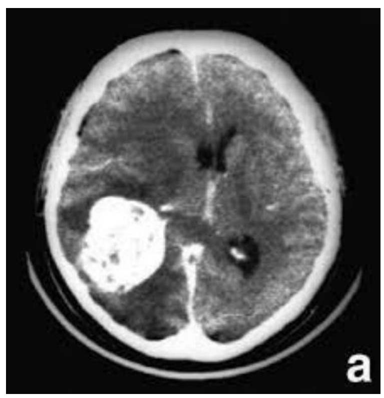

Figure 2.1: Seed point selection

The study was done to compare the results of region growing segmentation and mathematical morphology; further expanding research to thresholding with Otsu's thresholding segmentation. It was done using thresholding method which produced more accurate and robust results. The study resulted in qualitative evaluations as skull stripping using mathematical morphology was better than region growing segmentation at a good acceptance rate. The false positive and false negative ratio were also calculated giving improved results.

Zanaty *et al.* [21] presented a theory on brain tumor segmentation based on hybrid type of approach, combining Fuzzy c-means (FCM), seed region growing and Jaccard similarity coefficient algorithm to measure segmented GM and WM tissues from MR images.

- FCM is basically a data clustering algorithm in which each data point in the cluster belongs a cluster with an associated degree such that dissimilarity measure is minimized.
- Seed growing algorithm aims at merging neighboring pixels which have similar properties to the region. It follows homogeneity criteria, and if it is satisfied, the candidate chosen pixel is merged into the main region.

The proposed technique used FCM for finding the correct seed point as a tool for pre - segmentation, them seed growing algorithm was applied on the seed to obtain close regions, and further results were refined. A performance measure was calculated for merging regions and extraction of final segmentation. It gave an advantage as the results of both techniques. Hence, FCM gave quick results, but seed growing algorithm needed initial seed point to produce faster and better and highly accurate results. The final results obtained gave an average improved segmentation with less noise level.

Seetha *et al.* [22] examined machine learning classification schemes such as FCM segmentation which separates brain area into tumor and non-tumor region. The second scheme evaluated was bio physio-mechanical tumor growth to analyze steps by steps measure to show tumor growth of patients. The computation time for this technique is very high. Another scheme is multi model segmentation technique which combines various segmentation techniques to achieve high performance. It gave better results as compared to Adaboost classification in which both the complexity and accuracy calculated were low. The study proposed by the author using CNN gave high accuracy results with low complexity. The normal brain image had the lowest probability score, tumor image had the highest and non-tumor brain image had the lowest.

Mohsen *et al.* [23] presented a study for classifying a dataset of 66 images of brain into 4 different classes namely normal, glioblastoma, sarcoma and metastatic bronchogenic carcinoma tumors. The DNN classifier was combined with wavelet transformation with principal component analysis which gave a good response as overall performance measures. The segmentation method used was fuzzy c means, suitable features were extracted and then DNN was used as a classification scheme. The results showed 22 images as normal brain images and 44 as abnormal brain images.

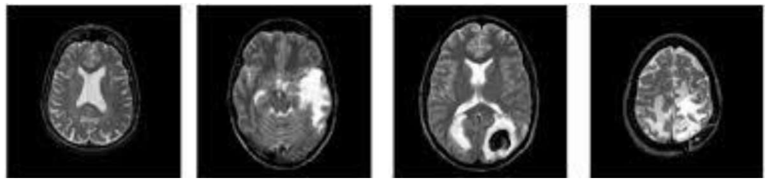

Figure 2.2: Normal, glioblastoma, sarcoma, metastatic bronchogenic carcinoma tumor

Kong *et al.* [24] surveyed the process of automatic tissue image segmementation based on image processing and deep learning techniques. Pre-processing is done to remove the complicity of the brain. Multiscale transformation is done to remove the noisy signals by the process of wavelet domain denoising to get accurate contours of different tissues such as CSF, WM, GM etc. Then CNN is applied for image classification on a dataset of 800 images. The concept of parallet computing has been introduced in the study as it greatly redcues the processing time as compared to manual and semi automatic segmentation. The speed, accuracy and run time were improved due to parallel computing.

Jangid *et al.* [25] proposed semantic image segmentation using deep convolutional networks and use of super pixels. Initially, semantic image segmentation has been defined followed by choosing an image from the dataset. Pixel labelled images were loaded and classes were defined at each step accordingly. The semantic segmentation has been identified and then plotted at the correct segmentation region. The more weights were varied, the better accuracy was achieved. Study showed good segmentation results as deep segmentation model predicted better color labels.

Blaschke *et al.* [26] presented different image strategies that have been compared with different techniques which are used to deal with different image semantics. Most of the understanding about an image is done using pixels. The results obtained from region or object based segmentation are easier to understand and further work on than pre-pixel classification. The study suggested that some aspects of geo-information cannot be achieved using information from pixels, but can be achieved using the information from the neighborhood. A combination of strategies and techniques can provide meaningful results with robust and reliable information as achieved by a human interpreter.

Kamavisdar *et al.* [27] examined different image classification techniques. Classification includes image sensors, pixel information, image pro-processing, segmentation, extraction of features and then classification. The main method followed is taking the digital data, pre-processing of the data, extraction of features, selection of training and testing data and then classification. After classification, post processing is done on the data and accuracy is calculated. The basic techniques discussed are supervised and unsupervised, parametric or non-parametric, object oriented or sub-pixel per-pixel and spectral classifier in which SVM had the maximum number of advantages over fuzzy c-means, decision tree and artificial neural network.

Lu *et al.* [28] presented a survey on image classification methods and techniques for improving the classification performance. Various methods of classification along with the best method to improve the accuracy and performance have been discussed. The technique has been applied on remotely sensed data which includes both airborne and spaceborne sensor data. Image processing has been done on the data that for the detection and even restoration of the image pixels, lines, radio calibration and topographical correction. Hence, the study suggests than each classifier is suitable in its own way depending on the type of data, and classifier can be chosen according to that. Post processing is done and later classification scheme is applied on the data.

Wang *et al.* [29] discussed the idea of CNN for image classification in the study. Basically, neural network is made up of layers. Each layer of the network has neurons connected to the neurons of the layer above and below them, having different weight and biases. The concept has been defined as the top layer being called as single vector, last layer being called as and the middle layers are called hidden layers. A dataset of 10000 images has been taken and by dividing them into 10 classes, data is acquired. By following this technique, the performance and accuracy were improved by 20%. In addition, the training time was reduced with better consistency.

Anthony *et al.* [30] presented classification of data using SVM. SVM has been known for applied machine fields to study characters, handwritten digits and texts, image recognition. It is known to work with a kernel function while working on the non linear data. For each kernel type such as linear, quadratic, polynomial and RBF, pixels have been examined and further accuracy is calculated. Due to mix and effect of unclassified pixels, the accuracy and performance varies. The accuracy reduced for linear and RBF kernel type, remain the same for polynomial type kernel but was increased for quadratic classifier. The choice of kernel though depends on the preference and the result can also vary depending on the type of data as high classification results can depend on the type of data as well as chosen kernel type.

Basavaprasad *et al.* [31] highlighted the importance of image processing and it's applications in various fields. Techniques used in digital image processing such as pixilation, self- organizing maps, linear filtering, Hidden Markov Model (HMM) and partial differential equations have also been discussed. Processes that are being followed are image enhancement, image restoration, image compression, character recognition, signature verification, biometrics work as application of digital image processing. Image processing is developing a fast pace in terms of multimedia, virtual interacting and medical image processing. It also helps in the growth of industrialized industries including medical, IT, biochemical etc.

Azhari *et al.* [32] proposed a technique for image segmentation in the field of medical imaging. The basic method showed in the study for the detection of tumor is image acquisition as the first step followed by CT scan of the brain image. The detection of gray part in the image is followed by extraction of useful parameters. The comparison is made after classification whether the patient has tumor or not. The infected area in the brain is not just at one spot which can be found easily but it is spread and is diagnosed after various examinations. Further segmentation of the tumor region is done using nodules concept using morphological operations. The noise has been removed using a series of filters namely Gaussian filter, linear filter and average filter. This way the accuracy is easily calculated along with the change in performance.

Girod *et al.* [33] examined the uses of processing digital images and how these images can be made useful in the field of technology. Image processing is useful in various fields such as machine learning, artificial intelligence, computer vision, robotics, visual perception, imaging, image coding and statistic, display technology. After an image is acquired of correct aperture and balance or else image is restored from imbalance, it is prepared for printing or display by gamma correction, color mapping or half toning, and then picture storage is facilitated contributing to the enhancement of the image. Meaningful information is extracted from the image by the process of character recognition or bar code method and finally image is processed.

Coelho *et al.* [34] indicated the diagnosis of benign tumors in human brain. The study has been clearly stating that most of the tumors that occur in human body are benign and curable upto a certain extent. The difference between the benign and malignant tumor is basically done by the clinical data and imaging studies. There may be a few cases where tumor is detected after the growth of hepatic mass. These tumors might require waitful watching but once they start to grow, they should be treated immediately before they result in an incurable abnormal mass of brain tissue.

Nagpal *et al.* [35] presented a study on various diagnostic tools for the detection of tumor and tumor markers. Tumor markers are the product of cancerous cells or the fluids that respond to cancer. The markers help in determining the stage of tumor and the wide presence of tumor in different regions. They reflect as how cancer is likely to progress and affect the system of the body. Once the tumor is detected, tumor markers are used to monitor patients as if the level of tumor is increasing or going low with time. If the level increases, it means the treatment is not responding and there should be some suitable change in the treatment. Certainly, suitable drugs can also be verified and changed according to the response of the tumor markers. But study has been stating that one cannot completely rely on these markers. Certain clinical tests are highly recommended if tumor markers state anything serious in the human body.

Dolecek *et al.* [36] reported the statistical reports of primary and central brain tumors diagnosed in UN. The data isn't collected from the medical record of the patients. It is collected from the pertinent information on their residents and further the files are reported to the government agencies taking care of this. Certain factors that are mentioned to be taken care of are how the tumor has occurred, how it was analyzed, whether the diagnosed tumor is benign or malignant, and if benign, how it can be cured. The study showed that in human body, the percentage of affected brain tissues was detected more in females than males.

Zacharaki *et al.* [37] proposed a study on classification of brain tumor and its grade using MRI texture and shape in machine learning. The study indicates the different types of tumor in the brain such as primary gliomas from metastases including the grade of the abnormal mass. The features have been selected using SVM to satisfy the further classification process. Some feature selection methods discussed are ranking based method, feature subset selection method, constrained LDA. The method predicts the level of malignancy on the basis of intensity and texture which states the overall ability and performance of the classifier. The data of total 98 patients has been taken and process was done on them. Results showed that accuracy was better when method was performed using SVM classifier than when performed using k-test.

Devkota *et al.* [38] presented a study on early detection of brain tumor using morphological reconstruction. The study states that after the pre processing of image, the noise is removed and image is segmented to classification as whether the brain has tumor or not. The use of segmentation is that it helps to achieve better accuracy in less computation time. The new method proposed for early detection is that image is processed using median filter; segmentation is done using fuzzy c-means method. Both statistical and textural features are extracted and reduction is done. If the accuracy achieved is high, it means the test of MRI image done is correct.

Razzak *et al.* [39] proposed a model on deep learning for medical images and their proceedings. Various techniques of machine learning and artificial intelligence have been playing a very important role in the field of medical image processing, computer aided diagnosis, interpretation of images, fusion and registration of images. Deep learning has more benefits over machine learning because of accurate diagnosis with great and high resolution pictures. Deep neural network has a layer of neurons which step by step performs various methods and they help in better performance. The study proved that deep learning methods are better than machine learning methods.

Chapter 3
Research Problem

3.1 Research Gaps

Segmentation of tumor represents the actual extraction of tumor present in the brain. Earlier, radiologists used manual segmentation techniques to extract the abnormal part from the brain. These techniques were often time consuming as a lot of time was wasted on viewing the temporal and spatial results and then examining the profile of pixels and lesion boundary. But with the growing era, the use of fully automatic segmentation techniques has been giving better results.

The data extracted was not enough to have appropriate results of classification of the tumor into benign or malignant. Hence, data was augmented using several methods to increase the data size. In CNN based machine learning technique, more the training data, better will be the results.

Most of the previous works have used cluster based segmentation method which aims at following clusters of the pixels depending on the type of pixels. This technique has a drawback that it only works well for the pixels which are majorly different from each other. We have used Berkeley Wavelet Transformation (BWT) in threshold segmentation technique which works well for gray scale images. The classifier used after the segmentation process is CNN compared with the results of SVM classifier. Although, SVM is good with just real time data, but the overall accuracy and performance was much less because of the variability in the selection of kernel in SVM. CNN has a network of learning units called neurons. The basic advantage of using CNN was that it automatically detects the important features without any human supervision with better accuracy than other machine learning approaches. CNN works with fully connected layers where each neuron in each layer is connected to all the neurons in the next layer making the data free to over-fitting. Moreover, the more the neurons are mapped to images, the better it is for CNN to identify unmapped images. This is how training of neurons in CNN is done.

3.2 Problem Statement

A lot of research is being carried out in the area of brain tumor detection. The test performed to get the results comprises of various factors such as thickness of the clump, size of tumor, variability in the tumor and so on which is very difficult for the neurologists to interpret the results. So, machine learning is need for early detection of tumor in the brain.

The models are created in relevance to the type of data and knockdown by studying daily basis and continuous enhancements in the medical and healthcare field. These studies incorporate with individual's relationships with others in terms of different parameters of human behavior. The preprocessing steps such as normalization of images and similarity measurement are carried out to make the image set suitable for analysis. A lot of research on image segmentation has been proposed but there is lack of effective study to analyze and utilize these techniques on the type of data. No specific segmentation technique is totally suitable for a specific type of data when we want to segment a part out of the image. Techniques such as edge detection method, region-based method, hybrid method, cluster based method etc are suitable in their own way.

The data on health, specifically the images of MR of the brain needs to be rectified so that noise can be removed from the dataset. Images need to be set in one size properly before any technique is applied on them. They can either be synthetic images or even realistic images of tumor inside the human brain. Pixel analysis and effective visualization of the analyzed images is another main concern in the area of healthcare as the operations done on any kind of image can only be processed with the help of pixel analysis. My research work will focus on the pre-processing of the image followed by the skull stripping method. It is very useful in extracting the extra tissues and determining the appropriate skull boundaries. It is then followed by the use of best segmentation techniques to strip out the tumor part out of the image. For the detection of abnormal tissue in the brain, a classifier with adequate performance will be chosen for classifying the dataset into type of tumors.

3.3 Research Objectives

In the light of above discussed research gaps following objectives have been formulated.

- To study existing segmentation and detection approaches related to brain tumor.
- To pre-process the data and perform data augmentation.
- To segment the tumor using threshold segmentation.
- To design a classification model by using CNN and validate it.

3.4 Research Methodology

In my research work, I will work with python language, which is an excellent scripting language for manipulating text. The dataset required for analysis will be extracted using MICCAI BraTS. The authorization credentials for the software were obtained from the college only which will be used for the implementation.

Step 1: BraTS dataset of 23 patients have been taken and it is extracted using MICCAI BraTs package in Python. Since, the dataset is too small, for augmentation of the dataset of images is done.

Step 2: The pre-processing step involving data cleansing and removal of noise is performed. There's still some noise left which is removed using k-means clustering process.

Step 3: Segmentation is done using threshold segmentation by keeping the threshold value as 128, distinguishing the pixels. The pixels having value higher than the threshold value are white and the pixels having value less than threshold are black.

Step 4: Feature Selection is performed for selecting the best informative features. It is performed using FSelector. The statistical values are calculated for the same. The total no of features extracted from the dataset.

Step 5: SVM is implemented using Gaussian kernel followed by the implementation of CNN classifier on the data using keras machine learning model.

Step 6: Performance is evaluated using various performance parameters comprising true negative, true positive, false positive, false negative, sensitivity, specificity and accuracy

Step 7: Finally, we get results in terms of accuracy.

3.5 Summary

In this, firstly the research gaps in the study are discussed and then the problem statement is defined. After that the objectives to achieve are discussed. Along with that the methodology followed to perform the proposed work is explained.

Chapter 4
PROPOSED WORK

For analysis and visualization of images from BraTS dataset, the first step was pre-processing of the images. The images were extracted and checked first horizontally and then vertically. Rotation and flipping were performed on the images. It was done by adding noise to the dataset using sp_noise. For some of the image dataset, it was augmented by flipping, creating the mirror image of the original image, and continuous rotations were performed for the images to have augmented data.

Of the complete dataset for rotation process, images were taken and rotated at 90 degrees, 180 degrees and 270 degrees. Further, data is also augmented according to the size and texture of the images. Then, segmentation process on these images is done by the method of threshold segmentation.

4.1. Threshold Segmentation

This segmentation is used to distinguish foreground from the background. We select a threshold value T, after the gray image is converted to binary image. The method to select threshold is as follows:

- An initial threshold value is selected, usually an 8-bit value of the original image.
- The original image is divided into two portions.
- Calculate the mean average value of the newly created images.
- Calculate the new threshold value by finding an average of the two means.

The threshold value T is 128 in this case. The pixels are partitioned based on their intensity values. Pixels have value greater than the threshold value marks the region as white and the pixels having values less than threshold values marks the region as black. Thresholding techniques can be categorized into local thresholding and global thresholding:

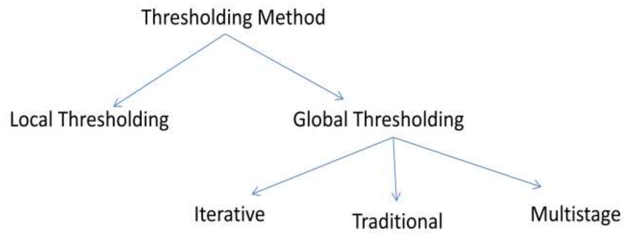

Figure 4.1: Thresholding techniques

- **Local Thresholding:** In this technique, threshold value is calculated for each pixel, based on factors such as range, intensity, variance etc. The intensity distribution between the background and the foreground is not very distinct. Some limitations of this approach are that they are edge and region bound. The techniques that fall under local thresholding are Bernsen's technique, Niblack's technique etc.
- **Global Thresholding:** It is a technique where the intensity distribution between the background and the foreground of objects is very distinct. In this case, a single value of threshold can simply be used to differentiate both objects apart. The value of the threshold depends on the characteristic of pixel and the grey level of the image. The main techniques that fall under global thresholding are Otsu method and entropy based method.

Following are the ways to choose the appropriate threshold value:

- *Entropy:* The methods result in algorithms that use the entropy of the foreground and background regions, the cross-entropy between the original and binary image.
- *Histogram-shape:* The methods, where, for example, the peaks, the valleys, curvatures of the smoothed histogram are analyzed.
- *Object-attribute:* The methods search a measure of similarity between the gray-level and the binary images, such as fuzzy shape similarity, edge coincidence.
- *Local* methods adapt the threshold value on each pixel to the local image characteristics. In these methods, a different T is selected for each pixel in the image.

4.2. Feature Extraction Methodology

In order to collect high level of information such as shape, size, contrast, texture etc. feature extraction is done. By just selecting the prominent features, the accuracy of the diagnosis becomes easy. Gray Level Concurrence Matrix (GLCM) is the most used technique for feature selection in image analysis. First of all the GLCM is calculated and then texture based on GLCM is calculated. It helps in reducing the data and speed ups the algorithm. Not only has the accuracy improved, but the overall performance as well. Some useful features and they methods to calculate them are listed below:

- *Mean (M):* By adding all the pixels values of the image and then dividing it by the total number of pixels, we obtain the mean of an image.

$$\mathrm{M}=\left(\frac{1}{a\times b}\right)\sum_{x=0}^{a-1}\sum_{y=0}^{b-1} f\left(x,y\right) \tag{1}$$

- *Standard Deviation (SD):* In probability distribution, standard deviation is the second most central point. It is the measure of non-uniformity and the more will be the value, more will be the contrast in the images.

$$SD\left(\sigma\right)=\sqrt{\left(\frac{1}{a\times b}\right)\sum_{x=0}^{a-1}\sum_{y=0}^{b-1}(f(x,y)-T)^2} \tag{2}$$

- *Entropy (E):* The degree of randomness in an image is called average information or entropy of an image. A certain image has certain entropy which can describe the texture of the image.

$$E=-\sum_{x=1}^{a-1}\sum_{y=1}^{b-1} f\left(x,y\right)\log f(x,y) \tag{3}$$

- *Skewness (S_k):* It is the measure of non symmetric behavior in an image. If an image is considered from the central point, it has to be similar from both left and right sides in order to be symmetric.

$$S_k(A)=\left(\frac{1}{a\times b}\right)\frac{\sum(f(x,y)-S)^3}{SD^3} \tag{4}$$

- *Kurtosis (K_t):* The probability of random variable is subjected to some shape whether it is flat or peaky which is represented by kurtosis of an image.

$$K_t(X) = \left(\frac{1}{a \times b}\right) \frac{\sum (f(x,y) - M^4 |}{SD^4} \quad (5)$$

- *Energy (En):* Energy is described as the amount of repetition in pixels. It can also be termed as the similar kind of pixels in the image.

$$En = \sqrt{\sum_{x=0}^{a-1} \sum_{y=0}^{n-1} f^2(x,y)} \quad (6)$$

- *Contrast (C):* It is the measure of the color and brightness of an image with respect to its neighbor over the image.

$$C = \sum_{x=0}^{a-1} \sum_{y=0}^{b-1} f^2(x,y) \quad (7)$$

4.3. Classification methodology

4.3.1. Support Vector Machine (SVM):

SVMs are supervised learning models which are used to analyze the data for classification as well as regression. It was invented by Vladimir N. Vapnik which was later modified by Carpis and Vapnik in 1993. If a training dataset is given, the purpose of SVM is to divide the non-linear transformation into a linear transformation using kernel functions of SVM. In the implementation, the kernel that I've used is Gaussian kernel which has made the classification very easy and convenient. SVM works with hyper planes dividing the dataset into two parameters which maximizes the margin between the non overlapping parameters. The performance in SVM is measured by accuracy, sensitivity and specificity. We've chosen SVM because it works best even when the data is not linearly separable. It is basically defined by a distinguishing hyperplane. The training dataset gives us the output as a hyperplane which finds a boundary between the two possible outputs.

4.3.2. Convolutional Neural Network (CNN):

When a device takes in an image, device sees it as an array combined of pixels which depends on the image resolution, typically the height, width and dimension. The array of matrix size 6x6x3 has RGB values whereas an array of matrix size 4x4x1 refers of an image with grey scale. CNN is used to identify images and the objects in an image, used for classification of images, detection etc. The process depends on passing esach image through a series of convolutional layers with kernels which act as filters, pooling and a series of fully connected layers.

- *Convolution layer:* It is the prior layer out of all the layers which is essential to extract the features from the image. In this layer, the convolution matrix is multiplied with the filter matrix referred as feature map.
- *Pooling layer:* When the image size is too large, this section of layer helps in reducing the parameters, retaining all the important information intact.
- *Fully connected layer:* In this section of layer, we flatten our matrix in the form of a vector and pass it to the fully connected layer in the form of a neural network and all the features are combined to form a model.

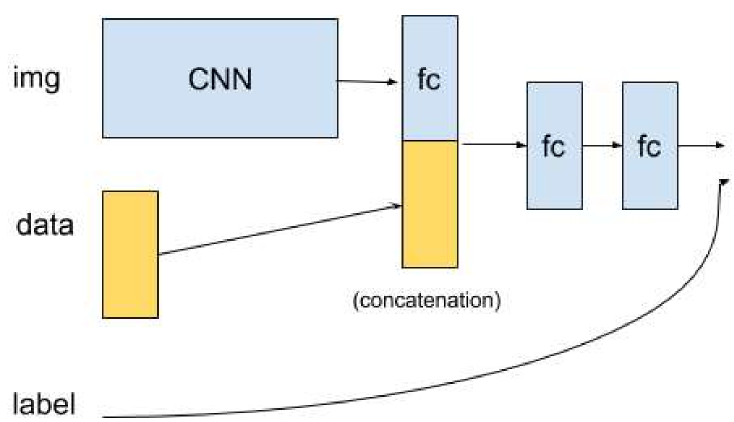

Figure 4.2: CNN layer model

4.4 Implementation Methodology

The implementation methodology is described in Figure 11. The first step is *pre processing.* In this step, the image is pre processed where the raw data is converted into the consistent one. It aims at suppressing unwanted distortions and enhances the features of the image. Data augmentation is also performed by resizing the images, flipping the images and performing mirroring.

The second step is *skull stripping* which is a process of removal of the skull from an image. It is very useful in extracting the extra tissues and determining the appropriate skull boundaries.

In third Step, *segmentation* is done in the infected portion of the brain is segmented out of the image. This step is followed by morphological operations.

Morphological operations are done which aim at removing the unwanted noise or any abnormalities if they are still present after pre-processing in step 4. The image becomes noise free up-to some extent. K-means clustering algorithm is performed which forms clusters of the pixels and separates the unwanted noise.

The step is followed by *feature selection* in step 5 in which features are extracted such as mean intensity, variance, kurtosis, contrast etc.

Then, in step 6, *CNN based classification* is done which gives the desired accuracy, sensitivity and specificity of the image dataset. It gives the output as whether the tissue in the MR image is normal or abnormal and classifies it into benign or malignant.

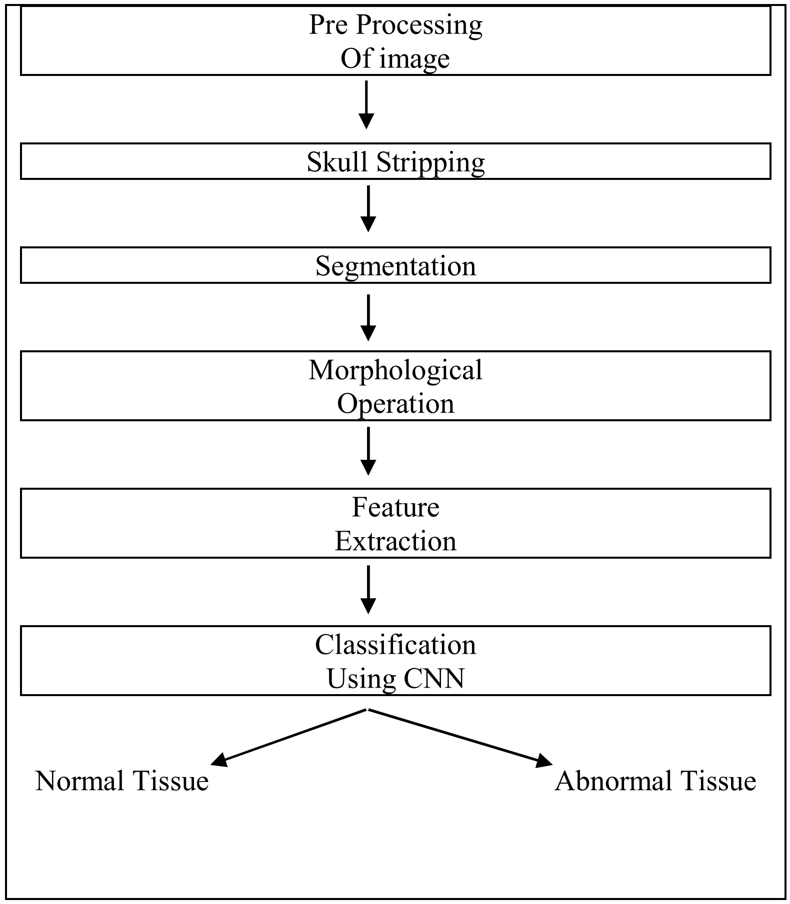

Figure 4.3: Implementation flowchart

Finally, the classification divides the tumor into either benign or malignant tumor as shown in the Fig 11. Further, necessary steps can be followed to cure the tumor depending on the stage of the tumor and its treatment.

4.4.1 Tools and Technology used

The following tools and technology have been used in the implementation of this work.

- Tenser Flow

64-bit system, Windows 10

- IDE

Spyder

- Programming languages

Python (free and open source)

- Important libraries used

- *Numpy:* This library provides support for large, multi-dimensional arrays. It is also called the replacement of matlab as when combined with packages like SciPy and Mat-plotLib, which also makes python a complete language in itself.
- *MatplotLib:* This library is a mathematical extension of numPy. It is used for the plotting of images on graph, creating 2D images using python scripts and further to check the exact accuracy of the output.
- *Tensorflow:* It is a library used for mathematical operations on python for dataflow. It is also used for some machine learning applications to be applied on neural network making them easier and faster. It can train and run deep neural networks for word embeddings, computations, digit recognition, machine translation etc.
- *Pathlib:* This library provides an interface between the file systems and the folders in the device which help in the proper interaction amongst them.
- *SciKit- learn:* This library contains a lot of useful tools for the computation of machine learning techniques such as SVM, random forest technique etc and statistical modeling which includes classification, regressions, reduction as well as clustering.

4.4.2 Assumptions made

- The detection is based on checking the left-right symmetry of the brain, which is assumption of the healthy brain.
- The assumption for an input was an image containing a tumor, so it did not deal with detection of images containing a brain with a tumor. The images of healthy brain are filtered out and could be used as a preprocessing step for the previous work.
- The assumption is made that the head is not rotated while MRI and skull is approximately symmetrical.
- As the head is assumed to be symmetric, the symmetry axes are set to be parallel to the vertical axis and to divide the detected rectangle into two parts of same size.

4.5. Summary

In this, methodology is explained for segmentation technique, feature extraction technique and then classification technique. The tools and technology used in the research have been explained.

Chapter 5
Testing and Results

This chapter discuss about the experimental setup for the proposed framework along with the implementation and results.

5.1 Experimental setup

5.1.1 Minimum Software and Hardware requirements

Table 5.1 H/W and S/W Requirements

1.	Processor	64 bit
2.	RAM	11 GB
3.	Hard Disk	80 GB
4.	Operating System	Windows 10
5.	Programming Language	Python
6.	Platform	Spyder

5.1.2 Segmentation using Berkeley Wavelet Transformation

We implemented machine learning algorithm with the help of python language. For the implementation of threshold segmentation, the threshold has a cut-off of 128. The pixel values greater than the selected threshold were mapped white whereas the pixel values lesser than the selected threshold are mapped as black. Because of this two separate regions are formed around the abnormal tissue section which is segmented out. The second step is to eliminate the white portion of the new image having pixel values higher than the selected threshold. Morphological operations are performed to eliminate it. Wavelet method is performed to work well with functions, data, operators and information into components of different frequencies, which lets the study of each element very easy as studying it as a whole. By scaling and translation processes, $\Psi(t)$ is the basic wavelet which is the cause of all wavelets. The morphological operations are performed to extract the boundary areas of the brain images by rearranging the pixels of the binary images. Dilation process is used when we aim to add pixels in the boundary region whereas erosion process is when we want to remove pixels from the boundary region.

Table 5.2 Steps performed in the segmentation process

Input Image Convert Image to gray scale Convert Image to Binary Image By Thresholding Find the number of connected Objects Assign 1 to inside and 0 to outside of the object in brain region Multiply calculated thresholds to get the final skull stripped image

5.1.3 Code snippet in python for Threshold Segmentation Implementation

```python
import cv2
import feature_extraction as fe
import numpy as np
from skimage import measure

class Segment:
    def __init__(self,segments=5):
        #define number of segments, with default 5
        self.segments=segments

    def noise_removal(self, image):
        image_noiseless = cv2.medianBlur(image, 3)
        image_noiseless=cv2.GaussianBlur(image,(7,7),0)
        return image_noiseless

    def kmeans(self,image):
        image=cv2.GaussianBlur(image,(7,7),0)
        vectorized=image.reshape(-1,3)
        vectorized=np.float32(vectorized)
        criteria=(cv2.TERM_CRITERIA_EPS + cv2.TERM_CRITERIA_MAX_ITER, 10, 1.0)
        ret,label,center=cv2.kmeans(vectorized,self.segments,None,criteria,10,cv2.KMEANS_RANDOM_CENTERS)
        print(center)
        res = center[label.flatten()]
        segmented_image = res.reshape((image.shape))
        return label.reshape((image.shape[0],image.shape[1])),segmented_image.astype(np.uint8)


    def extractComponent(self,image,label_image,label):
        component=np.zeros(image.shape,np.uint8)
        component[label_image==label]=image[label_image==label]
        return component

    def image_gray_scale(self, image):
        image_gray = cv2.cvtColor( image, cv2.COLOR_RGB2GRAY)
        return image_gray
```

Figure 5.1: Code Snippet for segmentation process

In Fig. 12, after defining the number of segments taking the default segments as 5, method for noise removal is applied this returns the image with less noise. If there's still some noise left in the dataset, k-means clustering algorithm is applied which returns noiseless image. In this algorithm, a centre is chosen and clusters are formed based on similarity and dissimilarity of pixels in the image. The final component from the image is extracted and finally the output is the gray scale image after the segmentation process.

5.1.4 Code snippet in python for Feature Extraction

```
from skimage import measure
from scipy import stats
from scipy.stats import moment
import scipy.stats as ss


def features_set1(image_segmented,original_image):
    regions = measure.regionprops(image_segmented, intensity_image = original_image)
    label = [r.label for r in regions]
    mean_intensity = [r.mean_intensity for r in regions]
    area = [r.area for r in regions]
    perimeter = [r.perimeter for r in regions]
    solidity = [r.solidity for r in regions]
    print(label)
    print(mean_intensity)
    print(area)
    print(perimeter)
    print(solidity)

def flattenArray(twodimage):
    arra = flattened_list = [y for x in twodimage for y in x]
    return arra

def statistics_features(image_data):
    image_data = flattenArray(image_data)
    skewness = ss.skew(image_data)
    print(skewness)
    Kurtosis = ss.kurtosis(image_data)
    print(Kurtosis)
    Entropy = ss.entropy(image_data)
    print(Entropy)
    variance = stats.tvar(image_data)
    print(variance)
    standard_deviation = stats.tstd(image_data)
    print(standard_deviation)
```

Figure 5.2: Code Snippet for feature extraction process

In Fig. 13, features mentioned in section 3.2 are being extracted. CNN automatically detects the important features without any human supervision with better accuracy than other machine learning approaches. The feature extraction process helps in improving the speed and effectiveness of the method. The features extracted during the implementation are mean, standard deviation, variance, entropy, skewness, kurtosis and contrast. The csv file for some of the images is attached in the Figure 14 below.

Tumor id	Skewness	Kurtosis	Entropy	Mean_intensity	Contrast	Variance	Standard Deviatio	Classification
B1	2.923	6.63	8.156	214.86	0.659	3777.631	61.462	1
B2	2.817	6.57	8.57	148.208	0.153	2254.841	47.485	1
B3	3.238	8.874	8.305	177.32	0.098	2421.21	49.2057	1
B4	2.647	5.176	8.57	222.7	0.151	4816.222	69.399	1
B5	2.36	3.819	8.757	168.42	0.163	3265.896	57.148	1
B6	3.104	7.894	7.648	212	0.1845	3619.99	60.166	1
B7	2.27	3.27	8.778	219.1641	0.129	5500.985	74.1686	1
B8	2.665	5.34	8.584	208.144	0.173	4292.296	65.515	1
B10	3.496	10.723	8.187	160.75	0.136	1805.929	42.496	1
B11	2.915	6.695	8.439	219.567	0.157	4397.497	66.3136	1
M1	2.865	6.566	8.48	164.904	0.587	2450.313	49.5	2
M2	2.292	3.34	8.755	224.105	0.1697	5618.048	74.953	2
M3	3.22	8.42	7.35	243.49	0.25	4197.907	64.791	2
M4	4.48	18.61	7.75	210.252	0.114	2037.806	45.142	2
M5	2.77	5.79	8.03	227.625	0.2253	4609.055	67.89	2
M6	2.472	4.212	8.66	233.14	0.138	2661.799	51.5926	2
M7	2.671	5.316	8.567	218.33	0.1406	4582.683	67.6955	2
M9	4.445	18.05	7.74	190.008	0.129	1597.842	39.973	2
M10	2.536	4.617	8.593	239.005	0.147	5502.796	74.18	2
M11	3.927	14.604	6.729	148.287	0.121	1549.902	39.368	2
M12	3.732	12.469	8.1027	203.303	0.127	2751.638	52.456	2

Figure 5.3: CSV file for feature extraction process

5.1.5 Code snippet in python for Classification process

```python

from keras.preprocessing.image import ImageDataGenerator
from keras.models import Sequential
from keras.layers import Conv2D, MaxPooling2D
from keras.layers import Activation, Dropout, Flatten, Dense
from keras import backend as K
import matplotlib.pyplot as plt

# dimensions of our images.
img_width, img_height = 256, 256

train_data_dir = "train-sub-val"
train_data_dir = train_data_dir[1:]
validation_data_dir = 'val'
nb_train_samples = 145
nb_validation_samples = 37
epochs = 75
batch_size = 5

if K.image_data_format() == 'channels_first':
    input_shape = (3, img_width, img_height)
else:
    input_shape = (img_width, img_height, 3)

```

Figure 5.4: Code Snippet for CNN model Implementation

In Fig. 15, keras model is implemented which is best used for the organization of neural network layers. The dimensions of the image are set to 256x256. First the epoch value is set to 70 and accuracy and loss are calculated for 70 epochs. Then the same is done for epoch value 200 and accuracy and loss is calculated. The number of samples for training is taken as 147.

5.2 Results

In order to predict whether the tissue in the brain as normal or abnormal and predict the outcome of various machine learning models we use BraTs dataset. Initially, we have dataset in Raw form. So, we need to preprocess it by using different techniques. The preprocessing of data is explained in section 3.4. Out of the 23 main MR images of the brain, 144 images were made through flipping and rotation methods.

After the implementation of threshold segmentation, threshold being selected as 128 distinguishing the white part and the black part in context to the pixels. The tumor is segmented out of the MR image in Fig 16 and gives us the output as shown in the Fig 17.

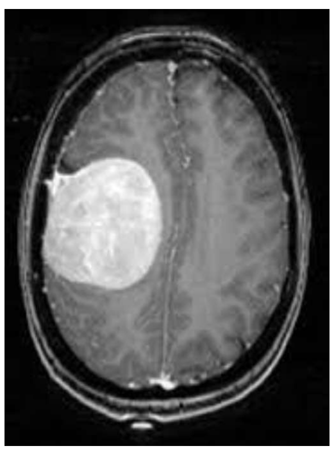

Figure 5.5: MR Image of the Brain Tumor **Figure 5.6: Segmentated**

5.2.1 Methodology

Step 1: BraTS dataset of 144 patients have been taken and it is extracted using MICCAI BraTs package in Python.

Step 2: The pre processing step involving data cleansing and removal of noise is performed. There's still some noise left which is removed using k-means clustering process.

Step 3: Segmentation is done using threshold segmentation by keeping the threshold value as 128, distinguishing the pixels. The pixels having value higher than the threshold value are white and the pixels having value less than threshold are black.

Step 4: Feature Selection is performed for selecting the best informative features. It is performed using FSelector. The statistical values are calculated for the same. The total no of features extracted from the dataset for few images is described in table 5.2 below:

Data	Mean	Standard Deviation	Entropy	Skewness	Kurtosis	Energy
Image 1	8.66	43.99	0.65	0.0053	2.89041E-06	10.94
Image 2	11.81	49.11	0.94	0.0065	2.74079E-06	16.37
Image 3	39.4	75.59	3.03	0.01054	1.8506E-06	65.99
Image 4	6.83	39.45	0.45	0.00517	3.3368E-06	8.11
Image 5	11.9	38.81	2.09	0.02002	1.35422E-05	33.17
Image 6	5.33	28.95	1.12	0.01647	2.05493E-05	13.87

Step 5: SVM is implemented using Gaussian kernel followed by the implementation of CNN classifier on the data using keras machine learning model.

Step 6: Performance is evaluated using various performance parameters comprising true negative, true positive, false positive, false negative, sensitivity, specificity and accuracy

Step 7: Finally, we get results in terms of accuracy.

Table 5.3. Summary of Evaluated Algorithms

Evaluation Parameter	**SVM Classifier**	**CNN**
True Negative	63	65
False Positive	18	4
True Positive	112	129
False Negative	8	3
Sensitivity (%)	93.3	97.72
Specificity (%)	77.77	94.2
Accuracy (%)	83.3	89

The proposed algorithm performed segmentation process, feature extraction process and classification process. The area under curve value and other performance parameters are calculated. The curves for 70 epochs and 200 epochs are shown below. The solid blue line in the curve depicts for the training data and solid orange line depicts for the testing data. It determines the curve between epochs-accuracy and epochs-loss. It determines that how many patients having cancer is actually have tumor in the brain specifying whether the tumor is benign or malignant.

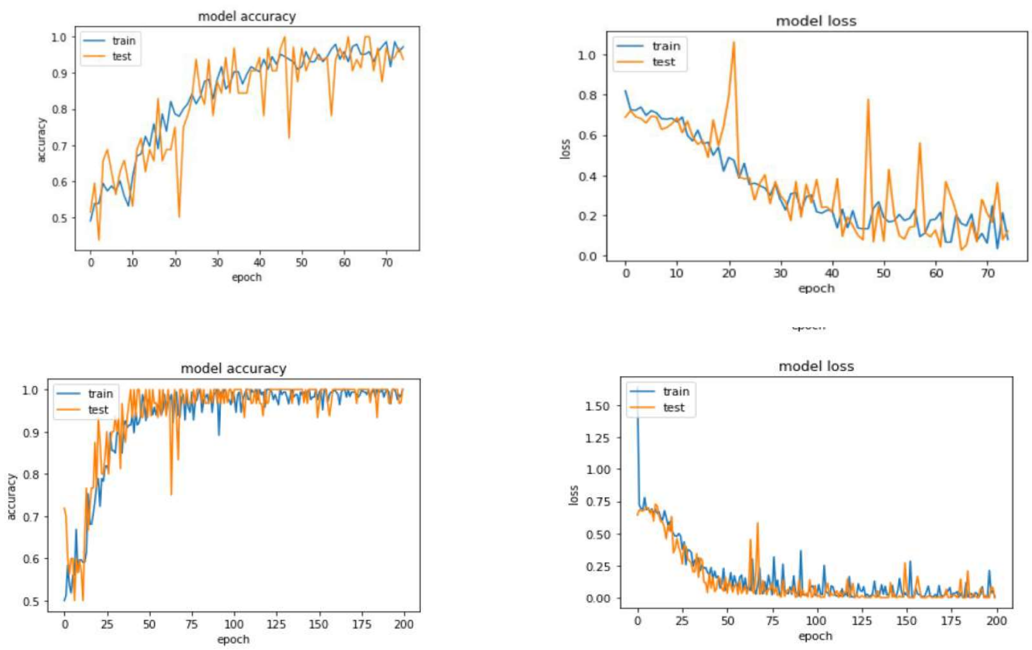

Figure 5.7: Graphs for accuracy v/s epochs and loss v/s epochs

In the Fig. 5.8, we have plotted the graph for accuracy v/s epochs and loss v/s epochs. The first and second graph depicts the graph for model accuracy and epochs for 70 epochs and graph for model loss and epochs for 70 epochs respectively. The third and fourth graph depicts the graph for model accuracy and epochs for 200 epochs and graph for model loss and epochs for 200 epochs respectively. We also plot the values based on the classification of tumor into benign tumor and malignant tumor to show the difference in results graphically.

Predicted	B	M	All
Actual			
B	18	2	20
M	3	23	26
All	21	25	46

Figure 5.8: Classification of tumor tissue into Benign and Malignant

Fig. 5.9 shows the final output of the as the classification of tumor into Benign or Malignant. The image dataset is segmented, the features are extracted and the image is hence classified.

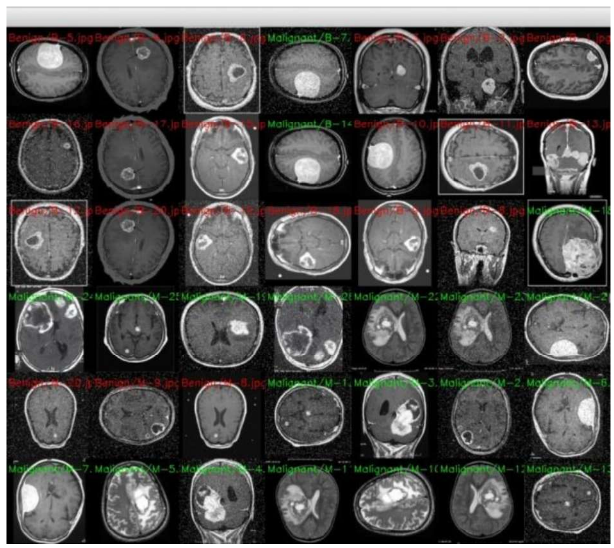

Figure 5.9: Final output of the classification

5.3 Summary

In this chapter, I've discussed the experimental setup of the research work and how the implementation is done along with the code. The results are discussed along with the conclusion of the work and future work that I can proceed with.

Chapter 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

As brain tumor is leading cause of deaths in people these days, so focus is early detection of tumor and it's cure by changing the methods of detection. Manual detection done by neurologists is time consuming, hence segmentation techniques namely fully automatic, or semi -automatic techniques are being promoted. Due to small size of image dataset, data is augmented to have better results of classification. Pre-processing of data is very important as it removes the noise from the image making it easier for detection. Threshold segmentation technique is applied for stripping the tumor out of the MR image of the brain using Berkeley Wavelet Transformation. SVM and CNN machine learning techniques are used for the classification of tumor into benign and malignant. After the data augmentation, the results are shown in terms of classification, accuracy, sensitivity and specificity. SVM works on real time data and is less accurate as compared to CNN [1]. CNN gives us better accuracy because it detects the best features using neural network. This method can also be applied to predict another type of cancers for clinical data, geonomic data as well as images or even integration of these.

6.2 Future Work

From the techniques I've followed, and from the implementation results, it is clear that analysis for brain tumor detection is fast and effective as compared to manual detection of tumor done by neurologists. The proposed research is suitable for the clinical data. However, in the present paper, the size of the dataset used is very limited. I am planning to increase the size of the dataset. Also, working on a technique for timely detection of tumor in the brain along with the exact location of the tumor. My future work will also include combining more than one classifier and feature extraction technique and then comparing the results with the use of single classifier.

References

[1] Kaur, Dilpreet, and Yadwinder Kaur. "Various image segmentation techniques: a review." *International Journal of Computer Science and Mobile Computing* 3.5 (2014): 809-814.

[2] Işın, Ali, Cem Direkoğlu, and Melike Şah. "Review of MRI-based brain tumor image segmentation using deep learning methods." *Procedia Computer Science* 102 (2016): 317-324.

[3] Yuheng, Song, and Yan Hao. "Image segmentation algorithms overview." *arXiv preprint arXiv:1707.02051* (2017).

[4] Kaur, Amanpreet. "A review paper on image segmentation and its various techniques in image processing." *Intern. J. Sci. Res* 3.12 (2012): 12-14.

[5] Senthilkumaran, N., and S. Vaithegi. "Image segmentation by using thresholding techniques for medical images." *Computer Science & Engineering: An International Journal* 6.1 (2016): 1-13.

[6] Gonzalez, Rafael C., and Paul Wintz. "Digital image processing(Book)." *Reading, Mass., Addison-Wesley Publishing Co., Inc.(Applied Mathematics and Computation* 13 (1977): 451.

[7] M. Jamali and H. Abolhassani, "*Different Aspects of Social Network Analysis*", In Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence, pp. 66-72, 2006.

[8] Kumari, Rosy. "SVM classification an approach on detecting abnormality in brain MRI images." *International Journal of Engineering Research and Applications* 3.4 (2013): 1686-1690.

[9] Abdalla, Hussna Elnoor Mohammed, and M. Y. Esmail. "Brain Tumor Detection by using Artificial Neural Network." *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*. IEEE, 2018.

[10] Damodharan, Selvaraj, and Dhanasekaran Raghavan. "Combining Tissue Segmentation and Neural Network for Brain Tumor Detection." *International*

Arab Journal of Information Technology (IAJIT) 12.1 (2015).

[11] ARUNADEVI, Baladhandapani, and Subramaniam Nachimuthu DEEPA. "Texture analysis for 3D classification of brain tumor tissues." *Przegląd Elektrotechniczny* (2013): 342-348.

[12] Alwan, Iman MG, and Enas Muzaffer Jamel. "Digital image watermarking using Arnold scrambling and Berkeley wavelet transform." *Al-Khwarizmi Engineering Journal* 12.2 (2016): 124-133.

[13] Liao, Yin-Yin, Po-Hsiang Tsui, and Chih-Kuang Yeh. "Classification of benign and malignant breast tumors by ultrasound B-scan and Nakagami-based images." *J. Med. Biol. Eng* 30.5 (2010): 307-312.

[14] Torheim, Turid, et al. "Classification of dynamic contrast enhanced MR images of cervical cancers using texture analysis and support vector machines." *IEEE transactions on medical imaging* 33.8 (2014): 1648-1656.

[15] Chaddad, Ahmad. "Automated feature extraction in brain tumor by magnetic resonance imaging using gaussian mixture models." *Journal of Biomedical Imaging* 2015 (2015): 8.

[16] Demirhan, Ayşe, Mustafa Törü, and Inan Güler. "Segmentation of tumor and edema along with healthy tissues of brain using wavelets and neural networks." *IEEE journal of biomedical and health informatics* 19.4 (2014): 1451-1458.

[17] Tohka, Jussi, Alex Zijdenbos, and Alan Evans. "Fast and robust parameter estimation for statistical partial volume models in brain MRI." *Neuroimage* 23.1 (2004): 84-97.

[18] Roslan, Rosniza, Nursuriati Jamil, and Rozi Mahmud. "Skull stripping magnetic resonance images brain images: Region growing versus mathematical morphology." *International Journal of Computer Information Systems and Industrial Management Applications* 3 (2011): 150-158.

[19] Vishnuvarthanan, G., et al. "An unsupervised learning method with a clustering approach for tumor identification and tissue segmentation in magnetic resonance brain images." *Applied Soft Computing* 38 (2016): 190-212.

[20] Benson, C. C., and V. L. Lajish. "Morphology based enhancement and skull stripping of MRI brain images." *2014 International Conference on Intelligent*

Computing Applications. IEEE, 2014.

[21] Zanaty, E. A. "An approach based on fusion concepts for improving brain Magnetic Resonance Images (MRIs) segmentation." *Journal of Medical Imaging and Health Informatics* 3.1 (2013): 30-37.

[22] Seetha, J., and S. S. Raja. "Brain Tumor Classification Using Convolutional Neural Networks." *Biomedical & Pharmacology Journal* 11.3 (2018): 1457-1461.

[23] Mohsen, Heba, et al. "Classification using deep learning neural networks for brain tumors." *Future Computing and Informatics Journal* 3.1 (2018): 68-71.

[24] Kong, Zhenglun, et al. "Automatic Tissue Image Segmentation Based on Image Processing and Deep Learning." *Journal of healthcare engineering* 2019 (2019).

[25] Bullock, Joseph, Carolina Cuesta-Lázaro, and Arnau Quera-Bofarull. "XNet: a convolutional neural network (CNN) implementation for medical x-ray image segmentation suitable for small datasets." *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*. Vol. 10953. International Society for Optics and Photonics, 2019.

[26] Blaschke, Thomas, Charles Burnett, and Anssi Pekkarinen. "Image segmentation methods for object-based analysis and classification." *Remote sensing image analysis: Including the spatial domain*. Springer, Dordrecht, 2004. 211-236.

[27] Kamavisdar, Pooja, Sonam Saluja, and Sonu Agrawal. "A survey on image classification approaches and techniques." *International Journal of Advanced Research in Computer and Communication Engineering* 2.1 (2013): 1005-1009.

[28] Lu, Dengsheng, and Qihao Weng. "A survey of image classification methods and techniques for improving classification performance." *International journal of Remote sensing* 28.5 (2007): 823-870.

[29] Wang, Chen, and Yang Xi. "Convolutional Neural Network for Image Classification." *Johns Hopkins University Baltimore, MD* 21218.

[30] Anthony, Gidudu, Hulley Greg, and Marwala Tshilidzi. "Classification of images using support vector machines." *arXiv preprint arXiv: 0709.3967*

(2007).

[31] Basavaprasad, Bl, and M. Ravi. "A study on the importance of image processing and its applications." *IJRET: International Journal of Research in Engineering and Technology* 3 (2014).

[32] Azhari, Ed-Edily Mohd, et al. "Tumor detection in medical imaging: A survey." *International journal of Advanced Information Technology* 4.1 (2014): 21.

[33] Girod, Bernd, and Frank HD Hartung. "Watermarking method and apparatus for compressed digital video." U.S. Patent No. 5,809,139. 15 Sep. 1998.

[34] Coelho, Júlio Cezar Uili, et al. "Indication and treatment of benign hepatic tumors." *ABCD. Arquivos Brasileiros de Cirurgia Digestiva (São Paulo)* 24.4 (2011): 318-323.

[35] Nagpal, Madhav, et al. "Tumor markers: A diagnostic tool." *National journal of maxillofacial surgery* 7.1 (2016): 17.

[36] Dolecek, Therese A., et al. "CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005–2009." *Neuro-oncology* 14.suppl_5 (2012): v1-v49.

[37] Zacharaki, Evangelia I., et al. "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme." *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 62.6 (2009): 1609-1618.

[38] Devkota, B., et al. "Image segmentation for early stage brain tumor detection using mathematical morphological reconstruction." *Procedia Computer Science* 125 (2018): 115-123.

[39] Razzak, Muhammad Imran, Saeeda Naz, and Ahmad Zaib. "Deep learning for medical image processing: Overview, challenges and the future." *Classification in BioApps*. Springer, Cham, 2018. 323-350.

List of Publications

[1] M. Chadha, S. Jain, "*Detection of Brain Tumor using Machine Learning Approach*", In Proceedings of Springer 3rd International Conference in the area of Advanced Computing and Data Sciences (ICACDS-19), Inderprastha Engineering College Ghaziabad, April 12 – 13, 2019. [**Accepted**]

th

ORIGINALITY REPORT

14%	6%	8%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Nilesh Bhaskarrao Bahadure, Arun Kumar Ray, Har Pal Thethi. "Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM", International Journal of Biomedical Imaging, 2017 Publication	2%
2	www.scilit.net Internet Source	1%
3	Pavel Dvorak, Walter Kropatsch, Karel Bartusek. "Automatic detection of brain tumors in MR images", 2013 36th International Conference on Telecommunications and Signal Processing (TSP), 2013 Publication	1%
4	Submitted to Chandigarh University Student Paper	1%
5	aircconline.com Internet Source	1%

Submitted to University of Wolverhampton