Brain Tumor Detection using Convolutional Neural Network

A thesis
Submitted in partial fulfillment of the requirements for the Degree of
Bachelor of Science in Computer Science and Engineering

Submitted by

Mohsena Ashraf	**150104012**
Tonmoy Hossain	**150104032**
Fairuz Shadmani Shishir	**150104082**
MD Abdullah Al Nasim	**150104085**

Supervised by

Mr. Faisal Muhammad Shah
Assistant Professor
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Dhaka, Bangladesh

June 2019

CANDIDATES' DECLARATION

We, hereby, declare that the thesis presented in this report is the outcome of the investigation performed by us under the supervision of Mr. Faisal Muhammad Shah, Assistant Professor, Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh. The work was spread over two final year courses, CSE4100: Project and Thesis-I and CSE4250: Project and Thesis-II, in accordance with the course curriculum of the Department for the Bachelor of Science in Computer Science and Engineering program.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Mohsena Ashraf
150104012

Tonmoy Hossain
150104032

Fairuz Shadmani Shishir
150104082

MD Abdullah Al Nasim
150104085

CERTIFICATION

This thesis titled, **"Brain Tumor Detection using Convolutional Neural Network"**, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for the degree B.Sc. in Computer Science and Engineering in June 2019.

Group Members:

Mohsena Ashraf	**150104012**
Tonmoy Hossain	**150104032**
Fairuz Shadmani Shishir	**150104082**
MD Abdullah Al Nasim	**150104085**

Mr. Faisal Muhammad Shah
Assistant Professor & Supervisor
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Professor Dr. Kazi A Kalpoma
Professor & Head
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

ACKNOWLEDGEMENT

We would like to express our deep gratitude to our honourable thesis supervisor Mr. Faisal Muhammad Shah, Assistant Professor, Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, for his patient guidance, enthusiastic encouragement and constructive critiques of this thesis work. We would also like to thank Professor Dr. Kazi A Kalpoma, the honourable Head of the Department of Computer Science and Engineering of Ahsanullah University of Science and Technology, and all our respected teachers, lab assistants, and friends for their moral support and helpful discussions. Finally, we would like to thank our parents for their constant support and encouragement throughout our study.

Dhaka
June 2019

Mohsena Ashraf

Tonmoy Hossain

Fairuz Shadmani Shishir

MD Abdullah Al Nasim

ABSTRACT

Brain Tumor segmentation is one of the most crucial and arduous tasks in the field of medical image processing as a human-assisted manual classification can result in inaccurate prediction and diagnosis. Moreover, it becomes a tedious task when there is a large amount of data present to be processed manually. Brain tumors have diversified appearance and there is a similarity between tumor and normal tissues and thus the extraction of tumor regions from images becomes complicated. In this thesis work, we developed a model to extract brain tumor from 2D Magnetic Resonance brain Images (MRI) by Fuzzy C-Means clustering algorithm which was followed by both traditional classifiers and deep learning methods. The experimental study was carried on a real-time dataset with diverse tumor sizes, locations, shapes, and different image intensities. In traditional classifier part, we applied six traditional classifiers namely- Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Multi-layer Perceptron (MLP), Logistic Regression, Naive Bayes and Random Forest. Among these classifiers, SVM provided the best result. Afterwards, we moved on to Convolutional Neural Network (CNN) which shows an improvement in performance over the traditional classifiers. We compared the result of the traditional classifiers with the result of CNN. Furthermore, the performance evaluation was done by changing the split ratio of CNN and traditional classifiers multiple times. We also compared our result with the existing research works in terms of segmentation and detection and achieved better results than many state-of-the-art methods. For the traditional classifier part, we achieved an accuracy of 92.42% which was obtained by Support Vector Machine (SVM) and CNN gave an accuracy of 97.87%.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Medical imaging techniques are used to image the inner portions of a human body for medical diagnosis. And medical image classification is one of the most challenging & affluent topics in the field of Image Processing. Medical image classification problems, tumor detection or detection of Cancer is the most prominent one. The statistics about the death rate from brain tumor suggest that it is one of the most alarming and critical cancer types in the Human body. As per the International Agency of Research on Cancer (IARC), more than 1,000,000 people are diagnosed with brain tumor per year around the world, with ever increasing fatality rate. It is the second most fatal cause of death related to Cancer in children and adults younger than 34 years [1]. In recent times, the physicians are following the advanced methods to identify the tumor which is more painful for the patients. To analyze the abnormalities in different parts of the body, CT (Computed Tomography) scan and MRI (Medical Reasoning Imaging) are two convenient methods. MRI-based medical image analysis for brain tumor studies has been gaining attention in recent times due to an increased need for efficient and objective evaluation of large amounts of medical data. Analysis of this diverse range of image types requires sophisticated computerized quantification and visualization tools. So, automatic brain tumor detection from MRI images will play a crucial role in this case by alleviating the need of manual processing of huge amount of data.

1.2 Brain Tumor

According to Ilhan et al. [2], a brain tumor occurs when abnormal cells form within the brain. Many different types of brain tumors exist. Some brain tumors are noncancerous (benign), whereas some brain tumors are cancerous (malignant) and some are pre-malignant. Cancerous tumors can be divided into primary tumors that start within the brain, and secondary tumors that have spread from somewhere else, known as brain metastasis tumors [2].

1.3 Classification of Brain Tumor

There are two types of brain tumor. One is Benign Tumor characterized as non-cancerous and the other one is Malignant Tumor- also known as Cancerous Tumor.

1.3.1 Benign Tumor

Benign brain tumors are usually defined as a group of similar cells that do not follow normal cell division and growth, thus developing into a mass of cells that microscopically do not have the characteristic appearance of a cancer [3]. These are the properties of a benign tumor:

- Most benign tumors are found by CT or MRI brain scans.
- Grows slowly, do not invade surrounding tissues or spread to other organs, and often have a border or edge that can be seen on CT scans.
- It can be life threatening because they can compress brain tissues and other structures inside the skull, so the term 'benign' can be misleading.

1.3.2 Malignant Tumor

Malignant brain tumors contain cancer cells and often do not have clear borders. They are considered to be life threatening because they grow rapidly and invade surrounding brain tissues [4]. These are the properties of a malignant tumor:

- Fast growing cancer that spreads to other areas of the brain and spine.
- A malignant brain tumor is either graded 3 or 4, whereas grade 1 or 2 tumors are usually classified as benign or non-cancerous.
- Generally these are more serious and often more fatal threat to life.

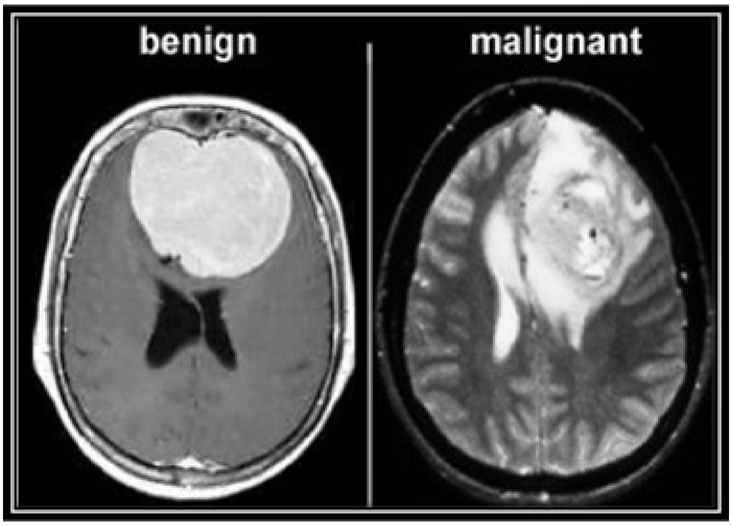


Figure 1.1: Benign Tumor (left) and Malignant Tumor (Right) [5]

1.4 Motivation

Observing the recent statistics of death rate caused by brain tumors, we selected brain tumor detection and classification which belongs to the field of medical image analysis. Tumor detection in medical images are time consuming as it depends on human judgment. The experts in this field, such as radiologists, specialized doctors examine CT scan, MRI, PET scan images and give decisions upon which the treatment depends. This whole process is time consuming. Automated medical image analysis can help to reduce the time and effort taken here and the workload of a human as it will be done by machines.

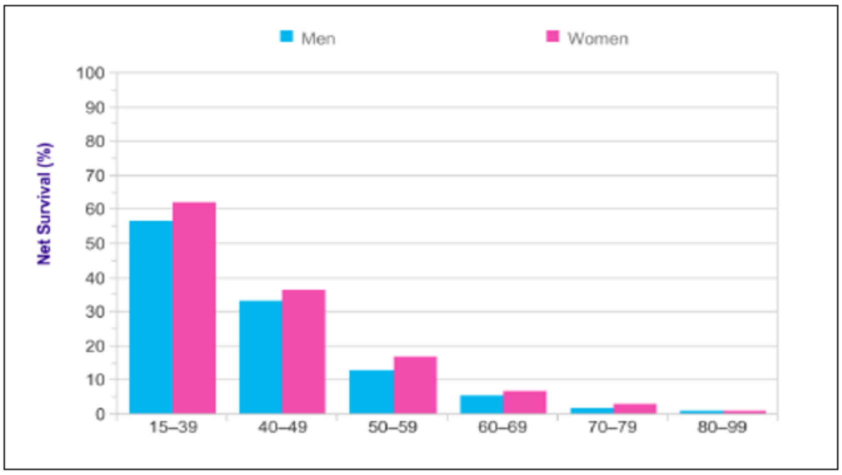


Figure 1.2: New cases and survival rate caused by brain tumor [6]

Figure 1.2 shows that death caused by brain cancer is higher than other types of cancers. Brain tumor detection in an early stage can help to reduce the death rate in this field. For supporting faster communication, where patient care can be extended to remote areas using information technology, automated image analysis will help to a great extent.

The developed countries of the world have been introduced to the automation of medical image analysis. But, in Bangladesh, it has not been adopted well yet. We want to build a model which will be efficient and feasible in the perspective of Bangladesh.

If we go through the most recent decade's statistics, it had been shown that there was an estimated 14.1 million cancer cases around the world in 2012. Men comprised of 7.4 million among them, while rest 6.7 million were consisted of female. This number is expected to increase to 24 million by 2035. Among all the form of cancers, Lungs cancer was the most common cancer worldwide contributing 13% of the total number of new cases diagnosed in 2012 [7]. So analyzing these statistics, we wanted to contribute in the field of medical image analysis.

1.5 Objective

The main objective of our thesis is to build a model that can predict whether the medical images contain a tumor or not and find its properties. Primarily, dataset collection is the main task to work on a medical image because brain tumor dataset is scarce as well as very much complicated to acquire. Most of the researchers focused on definitive work like filtering, segmentation, feature selection or skull removing. Here we tried to establish a model which can accomplish all the fundamental and major necessary tasks to find a tumor and its properties. We proposed an efficient and effective method which helps in the segmentation and detection of the brain tumor without any human assistance, based on both traditional classifiers and Convolutional Neural Network. Finally, we compared all the experimental results to find out which model provides better performance in terms of accuracy, sensitivity and other performance metrics.

1.6 Thesis Contribution

- To understand the current status of segmentation techniques of Brain tumor, we have done a statistical review which includes 52 research articles closely from different background. This statistical review of segmenting tumor includes both image processing and Neural Network techniques. It covers almost all types of segmentation methods along with the different type of pre-processing techniques.

- For segmentation of the tumor, we developed a skull stripping method using basic image processing techniques which produces a much better result than the existing skull removal techniques which is discussed in the experimental results section.

- We carried out two types of classifications to detect the tumor. Tumor classification using Traditional Machine Learning classifiers and Convolutional Neural Network were carried out and comparison of performance measures was done between these two models.

- A five-layer convolutional neural network is applied to the dataset which gives an improved result with respect to other research studies. The model is less complex as we detected the tumor using only five-layer CNN. In terms of training time and accuracy, the proposed CNN model gives a better result than most state-of-the-art works. The model is less complex as we detected the tumor using only five-layer CNN.

1.7 Thesis Structure

Our thesis book consists of seven chapters based on our research work. Following, we will briefly discuss about the basis of the chapters.

- **Chapter 1: Introduction** This is the introductory chapter of the thesis book. In this chapter, we will describe shortly the objective, our goal, and contribution of our thesis. We will also characterize Brain Tumor and its classification shortly.

- **Chapter 2: Related Works** We will describe some associated work that had been done before and describe their working approach, advantages and disadvantages.

- **Chapter 3: A Statistical Analysis on Brain Tumor Segmentation Techniques** A statistical analysis is represented in this section along with some depiction of charts, figures, and tables followed by their description.

- **Chapter 4: Background Study** This chapter includes all the prerequisite knowledge related to our thesis topic. We will discuss the Image processing techniques, Traditional Machine Learning Classifiers, and Deep Learning.

- **Chapter 5: Proposed Methodology** Our proposed methodology to segment the tumor using basic image processing techniques and detect the tumor using traditional machine learning classifiers & Convolutional Neural Network have been discussed in this chapter.

- **Chapter 6: Experimental Results & Evaluation** In this chapter, we will explain the performance measures, our proposed algorithm for performance evaluation and we will discuss about the experimental results.

- **Chapter 7: Conclusion** Finally, we will end our work with the conclusion chapter. In this division, we will discuss the limitations of our work and future directions to work on and improve.

1.8 Summary

In this chapter, a short description of Brain tumor and its sub-fields are described. We have discussed different types of Brain tumor and its characteristics. We shortly explained the motivation behind our work and the objective for doing it. At last, our thesis contribution and the structure of our thesis was briefly described.

Chapter 2

Literature Review on Brain Tumor Detection

2.1 Overview

In recent years, numerous and diverse types of work have been carried out in the field of medical image processing. Researchers from the various ground such as- computer vision, image processing, machine learning came into a place in the field of Medical Image Processing. We have studied some of the existing papers to find the most useful and advanced methods that were used in the existing articles in recent times. We worked on a total of 52 research articles. In this chapter, we will discuss thoroughly about these papers and their working procedures which are related to our work.

2.2 Reviews of the related papers

Shehzad et al. [8] proposed an algorithm to detect brain tumor from the MR images and calculate the area of the tumor. The designed algorithm claims to detect and extract the tumor of any shape, intensity, size, and location.

Working Approach:

- MRI images are converted to gray-scale images.
- Gaussian low pass filter is used to blur the image and then the blurred image is added to the original image.
- Median filter is used to remove noise.

- The morphological gradient is computed by dilation and erosion.
- Morphological gradient image and filtered image are added for image enhancement.
- The threshold value is calculated with the help of standard deviation and mean of the filtered image.
- Image is binarized by comparing the threshold value with every pixel value of the image.
- Erosion is done again for thinning the image and dilation is done again to get removed (caused by erosion) part of the tumor back.
- Tumor is extracted by comparing the original image with the dilated image.
- Erosion is done to shrink any noise remaining in the resultant tumor extracted image.
- The area of the tumor is calculated.

Sankari et al. [9] along with the other researchers came up with a model for cancer diagnosis for a brain tumor which is the toughest task. Most researches has been done in this field using PCA, Route set theory and Wavelet method. The authors here used Convolutional neural networks to solve the problem.

Working Approach:

- Respected authors proposed image de-noising, intensity normalization and bias-field correction method for the image pre-processing task.
- The bilateral filter is used to remove the noise from the MRI.
- Histogram Equalization is used for enhancing and feature extraction of the image.
- And finally CNN is used to classify the images.

Borase et al. [10] used computer-based procedures to detect tumor blocks and classify the type of tumor by using Artificial Neural Network. They used MRI images for their training and testing stage.

Working Approach:

- High pass filter is used for noise removal and pre-processing.
- For segmentation, region growing method is used.

- After segmentation, every set of connected pixels having the same gray-level values are assigned the same unique region label.
- Artificial Neural Network is applied for classification.
- K-means clustering technique could be a better option for pre-processing of the images.
- Erosion is done again for thinning the image and dilation is done again to get removed (caused by erosion) part of the tumor back.
- The tumor is extracted comparing the original image with the dilated image.
- Erosion is done to shrink any noise remained in the resultant tumor extracted image.

In Ilhan[11], Morphological operations, pixel subtraction, threshold-based segmentation, and image filtering techniques are used. In the pre-processing stage, Pixel subtraction is used to extract the skull from the brain image in a more efficient manner and administered with their developed thresholding approach from the traditional one and then filtering is used. There are some issues that can be improved in the system- the first one is the inaccurately classified images that have a tumor. Here maybe different methods of classification like neuro-fuzzy or support vector machine should be tried. Another one is that the proposed method is not an automatic procedure.

Working Approach:

- The image obtained from the opening operation is subtracted from the original image, thus separating the skull from the original image. The image obtained from the opening operation is subtracted from the skull image which removes the undesired gray pixels from the skull image. Lastly, the cleaned skull image is subtracted from the original image to get the resulting image which is excluding the skull.
- For segmentation, they proposed a threshold approach in which the threshold value was calculated by the sum of unique pixel values excluding zeros (black pixels) divided by the count of unique pixel values.
- The median filter is used in the filtering stage after segmentation.
- Accuracy of the proposed threshold method is 96%.

Joseph et al. [12] used K-means as well as SVM method for segmentation and to maintain the pattern for future uses respectively. They defined a relation between SVM and skull masking technique. They combine the K-means segmentation and SVM technique with skull masking for getting a better result. Maintaining the tumor pattern is also a computationally complex task which they did and also justified with the projected idea. They modified the feature extraction technique and existing K-means technique in a way that the amount of skull tissue is obscured and a proper tumor-detecting image is diversified. Hence we could do more by distinguishing the type of tumor and the location as well as the stage which is not yet identified properly.

Working Approach:

- They apply K-means segmentation with pre-processing on the MRI images.
- Image is converted to gray-scale. A 3*3 median filter is applied on brain MR image to reduce the noise and then high pass filter is applied to detect the edges.
- Skull Masking is used to find the region of interest.
- Support Vector Machine (SVM) is used in an unsupervised manner which will use to create and maintain the pattern for future use.

In hunnur et al.[13], the authors proposed a method to detect brain tumors mainly based on Thresholding approach and morphological operations. It also calculates the area of the tumor and displays the stage of the tumor.

Working Approach:

- MRI images of the brain are used as input and the images are converted into grayscale images.
- High pass filter is used to remove the noise present in the converted images.
- The median filter is applied to remove the impulse noise.
- Thresholding is used to extract the object from the background by selecting a threshold value.
- Morphological operations- dilation and erosion are done.
- Tumor region is detected and then the image is shrunk to remove the unwanted details present in the images.
- The tumor area is calculated and at last, the stage of the tumor patient is displayed.

A hybrid segmentation technique including Watershed and Thresholding based segmentation technique is used by Mustaqeem et al [14]. Firstly the quality of the scanned images is enhanced and then morphological operations are applied to detect the tumor along with their proposed hybrid segmentation. The proposed system is easy to execute and thus can be managed easily.

Working Approach:

- Obtained MRI images are displayed in two-dimensional matrices having pixels as its elements.
- Gaussian low pass filter and averaging filters are used to remove salt and pepper noise from the image.
- Filters pixel's value is replaced with its neighborhood values.
- Gaussian high pass filter is used to enhance boundaries of the objects in the image.
- Threshold segmentation is used to convert the grayscale image into a binary image format.
- Watershed Segmentation is used to group pixels of an image on the basis of their intensities.
- Morphological operators are applied to the converted binary image to separate the tumor part of the image.

Furthermore, we extended our study into some more recent articles and a thorough description is given by-

Devkota et al. [15] established the whole segmentation process based on Mathematical Morphological Operations and spatial FCM algorithm which improves the computation time, but the proposed solution has not been tested up to the evaluation stage. It detects cancer with 92% accuracy and classifier has an accuracy of 86.6%. Yantao et al. [16] resembled Histogram based segmentation technique. The brain tumor segmentation task as a three-class (tumor including necrosis and tumor, edema and normal tissue) classification problem regarding two modalities FLAIR and T1. The abnormal regions were detected by using a region-based active contour model on FLAIR modality. The edema and tumor tissues were distinguished in the abnormal regions based on the contrast enhancement T1 modality by the k-means method and accomplished a Dice coefficient and sensitivity of 73.6% and 90.3% respectively.

Based on edge detection approaches, Badran et al. [17] adopted the canny edge detection model accumulated with Adaptive thresholding to extract the ROI. The dataset contained 102 images. Images were first preprocessed, then for the two sets of the neural network, for the first set canny edge detection was applied, And for the second set, adaptive thresholding was applied. The segmented image is then represented by a level number and characteristics features are extracted by the Harris method. Then two neural network is applied, first for the detection of healthy or tumor containing the brain and the second one is for detecting tumor type. Depicting the outcomes and comparing these two models, the Canny edge detection method showed better results in terms of accuracy. Pei et al. [18] proposed a technique which utilizes tumor growth patterns as novel features to improve texture based tumor segmentation in longitudinal MRI. Label maps are being used to obtain tumor growth modeling and predict cell density after extracting textures such as fractal, mBm etc. and intensity features. Performance of the model reflected as the Mean DSC with tumor cell density- LOO: 0.819302 and 3-Folder: 0.82122.

Dina et al. [19] introduced a model based on the Probabilistic Neural Network model related to Learning Vector Quantization. The model was evaluated on 64 MRI images, among which 18 MRI images were used as the test set, and the rest was used as a training set. The Gaussian filter smoothed the images. 79% of the processing time was reduced by the modified PNN method. A Probabilistic Neural Network based segmentation technique was implemented by Othman et al. [20]. Principal Component Analysis (PCA) was used for feature extraction and also to reduce the large dimensionality of the data [20]. The MRI images are converted into matrices, and then Probabilistic Neural Network is used for classification. Finally, performance analysis is done. The training dataset contained 20 subjects, and the test dataset included 15 subjects. Based on the spread value, accuracy ranged from 73% to 100%.

Concentrating on Region based Fuzzy Clustering and deformable model, Rajendran et al. [21] accomplished 95.3% and 82.1% of ASM and Jaccard Index based on Enhanced Probabilistic Fuzzy C-Means model with some morphological operations. Zahra et al. [22] applied LinkNet network for tumor segmentation. Initially, they used a single Linknet network and sent all training seven datasets to that network for segmentation. They did not consider the view angle of the images and introduced a method for CNN to automatically segment the most common types of a brain tumor which do not require pre-processing steps. Dice score of 0.73 is achieved for a single network, and 0.79 is obtained for multiple systems.

2.3 Summary

In this chapter, we reviewed a total number of 52 research articles and discussed about their working procedures. This literature study shows that a lot of research works had been introduced regarding brain tumor segmentation and detection. Some of the researchers applied traditional classifiers, while the others implemented deep learning methods. Some works achieved a significant result using traditional approaches while some works did not. But after studying these works, we can claim that deep learning performs better than traditional classifiers because of their learning mechanism and use of memory in the network.

Chapter 3

A Statistical Analysis on Brain Tumor Segmentation Techniques

3.1 Overview

From the inception of Image processing, Medical Image is undoubtedly one of the most decisive sectors and brain tumor detection is one of the most crucial task in this field. Many researchers have worked in this field. We have studied 52 research articles which have been done for brain tumor detection and segmentation. In this chapter, we will present a statistical analysis of these research articles based on years, citations, segmentation techniques etc.

3.2 Year and Citation wise Distribution

Working on a total of 52 research articles dated from 2007 to 2018 and based on the collected information, we break down the total process of segmentation along with respective figures. We try to select the articles based on various criteria such as- citation, year, dataset, etc. but mostly focused on segmentation techniques. Apart from the single and mixed segmentation technique, we further go through the articles which adopted the Neural Network.

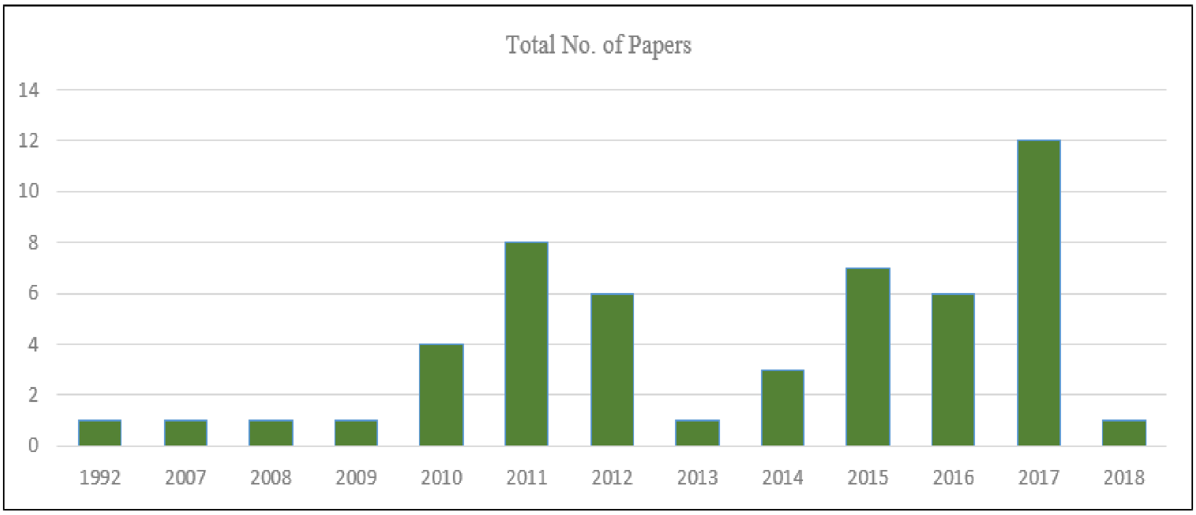


Figure 3.1: Year-wise distribution of the articles

In figure 3.1 and 3.2, statistics of the articles according to years and based on segmentation types is being represented with all the accessible information.
From figure 3.1, we can see that from year 2010 to 2013 and from year 2015 to 2017, maximum number of research works have been done. And from figure 3.2, we can depict that most of the researchers adopted Neural network based detection (supervised) and clustering based segmentation (unsupervised).

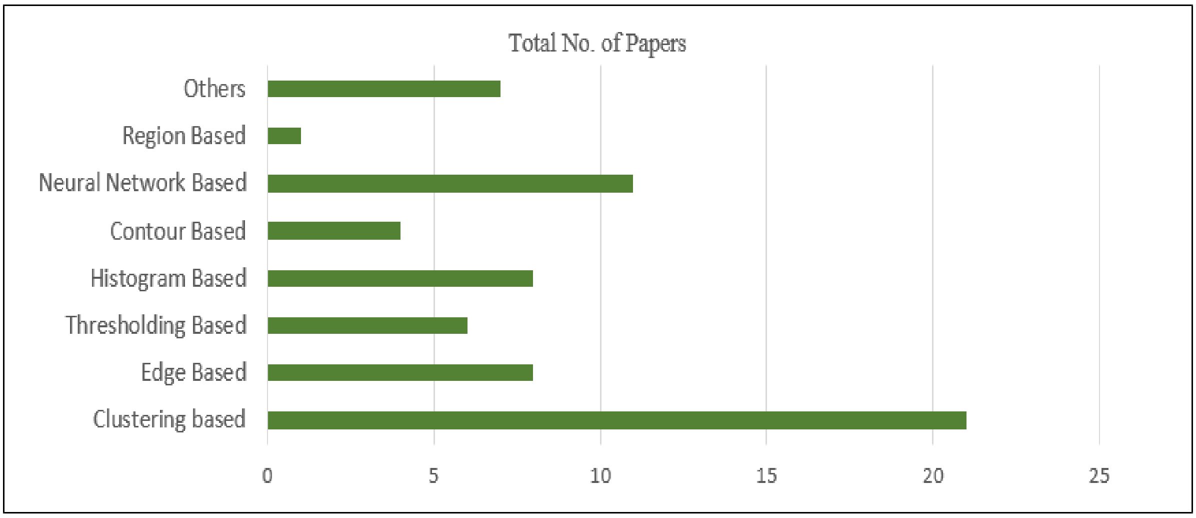


Figure 3.2: Distribution of segmentation techniques used in the articles

Further, in figure 3.3, categorizing the articles based on the number of citation is described, where a total of 42 papers where the segmentation technique belongs to the primitive image processing techniques.

In figure 3.4, the papers appertain to Neural Network based segmentation and the statistics represent the information about these papers citation.

From figure 3.3 and 3.4, we can see that the research articles which was done by Rajesh et al. [43] and Shenbagarajan et al. [65] has the highest number of citations.

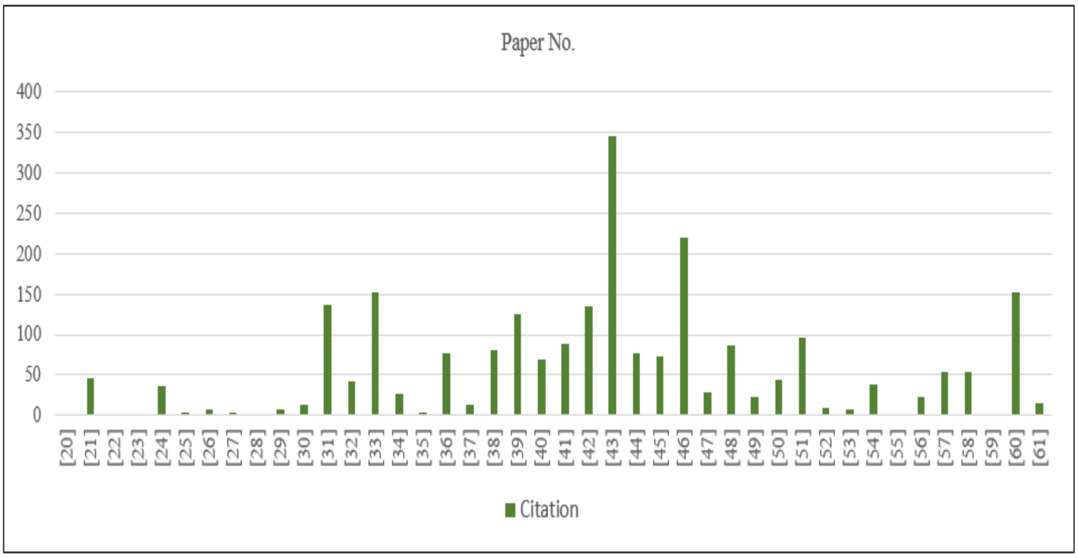


Figure 3.3: Citation-wise distribution of basic image processing techniques used in the articles

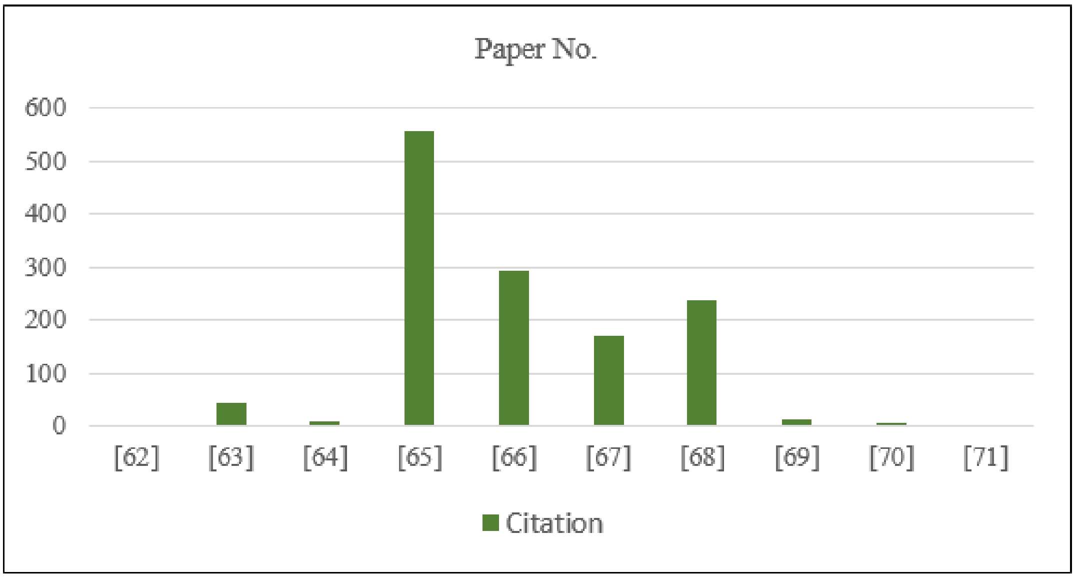


Figure 3.4: Citation-wise distribution of Neural Network based segmentation techniques used in the articles

3.3 Segmentation Technique wise Distribution

Segmentation subdivides an image into its constituent regions or objects based on some concurrent characteristics where the objects which are depicted are strongly related to the regions. The level of details to which the subdivision is carried depends on the problem being solved, that is, segmentation should stop when the objects or regions of interest in an application have been detected i.e. for the distinction of the tumor interests lied on separating the abnormal tissues.

Distribution of Skull Stripping techniques based on uses is represented through Figure 3.5. Summarizing all the information from figure 3.5, we can infer that skull stripping technique is gradually getting undone according to years.

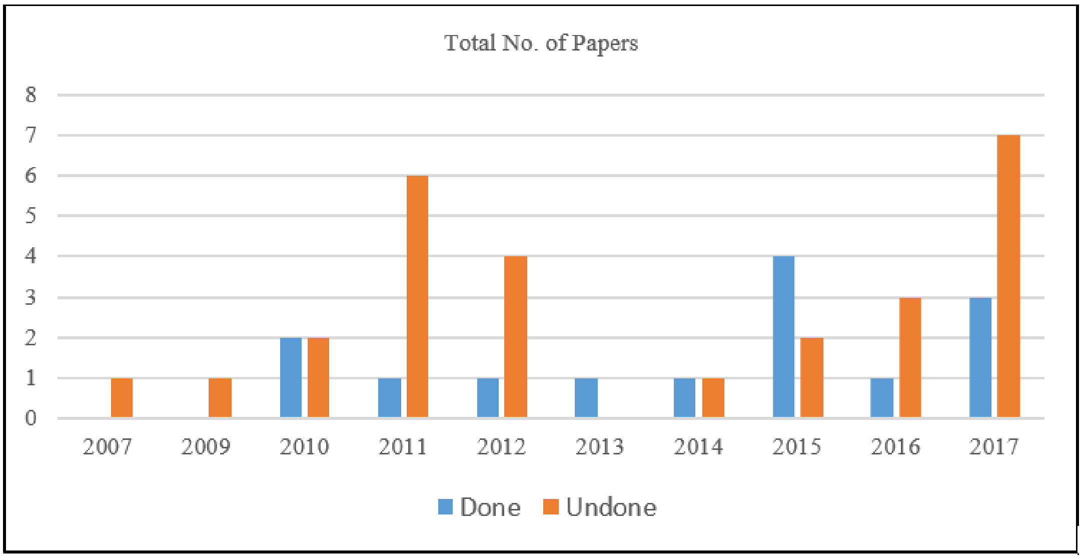


Figure 3.5: Distribution of usage of Skull stripping technique based on years

3.3.1 Layer Based Segmentation

Articulating an ROI and extracting decisive features, Layer Based segmentation techniques are the most indispensable one. In this segmentation process, three layers are generated from an image, entitled as mask, graphics and text layer. JBIG (Joint Bi-level Image Experts Group) algorithm is used to losslessly compress the mask layer, text layer is compressed using token based order [24], and graphics layer are compressed using the JPEG coder.

A layered model is used for object detection and image segmentation that composites the result of a bank of object detectors defining shape masks and explaining the appearance, depth order, and that evaluates both class and instance segmentation [25].

3.3.2 Region Based Segmentation

Deng et al. [26] proposed an adaptive region growing method based on the two preeminent subjects which cover variances and gradients along and inside of the boundary curve in order to overcome the difficulty of manual threshold selection. There are significant region-based segmentation methods which we can use to accomplish our goal, which are given below-

3.3.2.1 Region Growing

Region growing is the simplest and straightforward region based segmentation that groups' pixels or sub-regions into larger regions based on some pre-defined criteria. The common procedure is to differentiate one pixel with its neighbors [27]. There are two types of region growing method-

1. **Seeded Region Growing Method:** Along with the image, seeds are being taken as input and marking each of the objects that are to be segmented. The regions are iteratively grown through comparison of all unallocated neighboring pixels to the regions [28].

2. **Unseeded Region Growing Method:** It does not require the seed point and begins with a single region. At each iteration, it works through considering the adjacent pixels in the same way as seeded region growing.

3.3.2.2 Clustering

In region-based segmentation, Determination of the data set which belongs together is known as clustering. Clustering can be done in two ways- Partitioning (carve up the data set according to some notion of the association between items inside the set) and grouping (wish to collect sets of data items that are relevant to the respective model) [29]. Over the decades, multiple clustering based segmentation techniques have been developed and researchers administered these techniques in their model-

1. **K-Means Clustering:** When it comes to vector quantization and signal processing, K-Means clustering algorithm is one of the most dynamic and compelling algorithms. The edema and tumor tissues were distinguished in the abnormal regions based on the contrast enhancement T1 modality by the k-means method [11]. In several other respective research articles, K-means clustering and histogram clustering is applied after the initial image converted to color space and then CIELab color model [15]. Applying the K-Means segmentation technique after pre-processing and skull masking [30].

2. **Fuzzy C-Means Clustering (FCM):** One piece of data belongs to two or more clusters. Developed by Dunn in 1973 [31] and improved by Bezdek in 1981 [32], is frequently used in Image Segmentation. Use of FCM algorithm on intensity features where FCM segments image into a pre-specified number of clusters (K), FCM gives a fuzzy membership (U) to describe the degree of similarity of one pixel to each cluster [33].

3. **Spatial Fuzzy C-Means Clustering:** This algorithm utilizes the local spatial information which is convenient in reducing noise distortion and intensity inhomogeneity in the segmentation [34]. Segmentation was done by spatial FCM. An extension of Probabilistic Neural Network (PNN), called WPNN, which uses anisotropic Gaussians rather than the isotropic Gaussians used by PNN was used for classification [35].

   Figure-3.6 shows the depiction of the above information and figure-3.7 shows the information about the total number of papers which used clustering according to years.

   From figure 3.6, we see that K-means Clustering and Fuzzy C Means Clustering were mostly used. Furthermore, from figure 3.7, we can infer that, segmentation based on clustering techniques was mostly carried out in the year 2017 and clustering based segmentation techniques are still in practice.

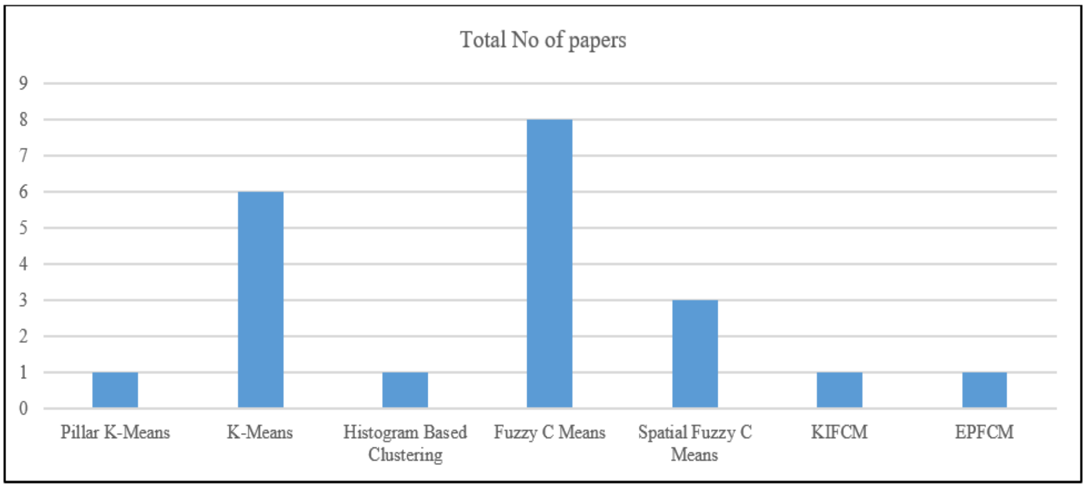


Figure 3.6: Number wise distribution of clustering techniques used in the articles

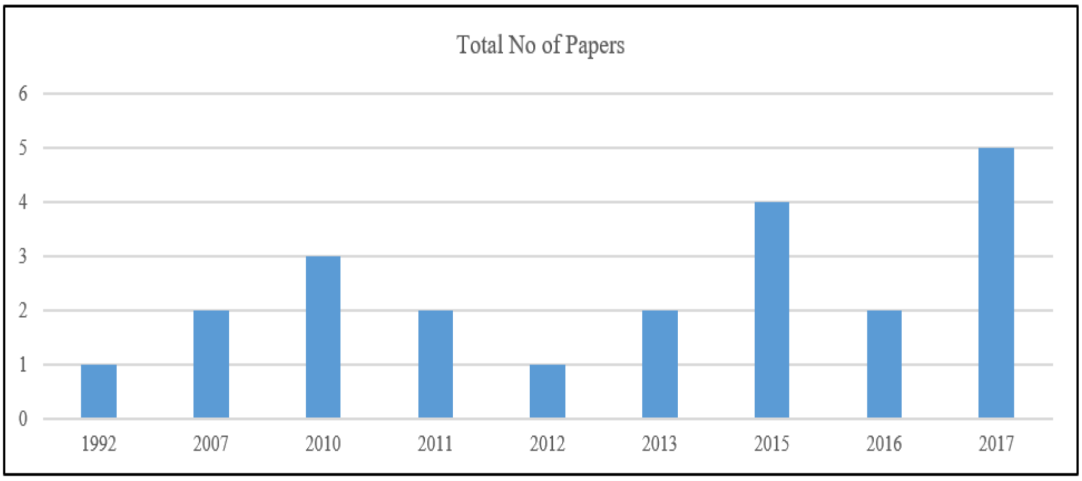


Figure 3.7: Year wise distribution of clustering techniques used according to years

3.3.3 Edge Based Segmentation

Edge based segmentation means segmentation of an image by identifying the edges of the Region of Interests. Edges can be connected and disconnected. According to the data or methodology, one needs region boundaries which are closed and the desired edges are the boundaries between such objects or spatial-taxon [36].

1. **Edge Detection:** Working with their developed thresholding and edge detection technique, Debnath et al. [37] identified the tumor successfully but the computation time, as well as the classifier and efficiency, is below average in contrast with the others' work.

2. **Canny edge detection:** By optimizing the canny edge detection model and GA for canny edge detection. Malathi et al. [38] improved the efficiency used closed contour segmentation. Badran et al. [17] used two sets of a neural network in their work. For the first set, they used canny edge detection.

3. **Watershed Segmentation:** Mustaqeem et al. [16], applied thresholding segmentation and watershed segmentation, followed by morphological operations. Karthik et al. [39] used the watershed transform-based segmentation process to extract the necessary region of interest from the skull stripped MRI images.

Figure 3.8 shows the information about the number of papers which used edge detection techniques in their work. And from the figure, we see that most of the papers carried out watershed segmentation technique.

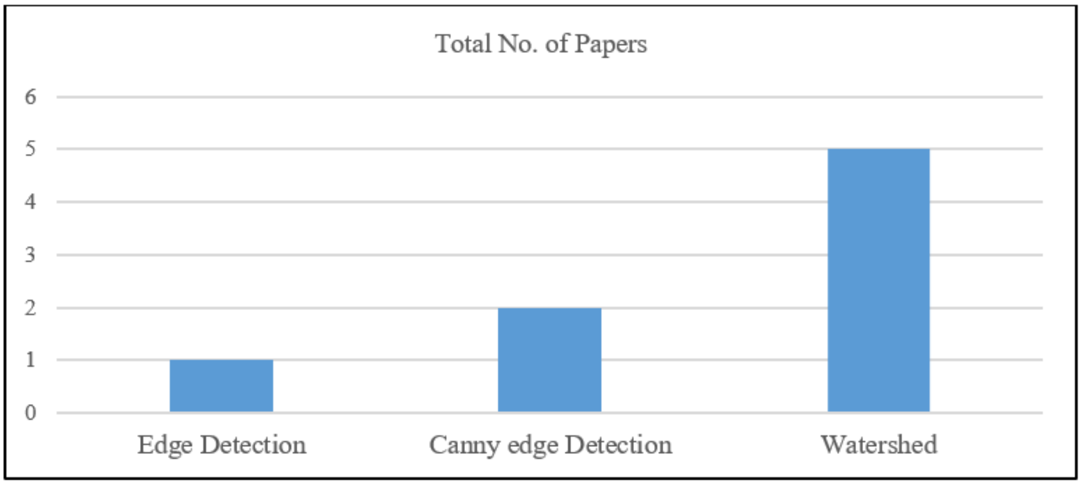


Figure 3.8: Number wise distribution of edge detection techniques used in articles

3.3.4 Thresholding Based Segmentation

Thresholding is based on a threshold-value or clip-level to convert a gray-scale image into a binary image and segments the region of interests. There are various types of thresholding methods which are depicted below:

1. **Binary Thresholding:** Debnath et al. [37] presented their algorithm which includes thresholding for tumor segmentation. Converting 24-Bit Color Images to 256 Gray Color Images and Calculating histograms the resulting images were converted to a binary thresholded image, histograms were calculated and at last edge detection algorithm was used [11].

2. **Adaptive or Dynamic Thresholding:** Different thresholds for different regions of the same image is calculated in this approach [40]. Badran et al. [17] tried two different segmentation techniques in their work and among them, one is adaptive thresholding.

3. **Otsu Thresholding:** This algorithm presumes that the image encompasses two classes of pixels following bi-modal histogram [41]. Mittal et al. [42] used Otsu Thresholding segmentation along with watershed technique, asserting that Otsu's thresholding chooses the threshold for minimizing the intra-class variance of the thresholded black and white pixels.

Figure 3.9 represents the number of papers where thresholding based approach was used for brain tumor segmentation and detection, where most of the papers adopted binary thresholding approach.

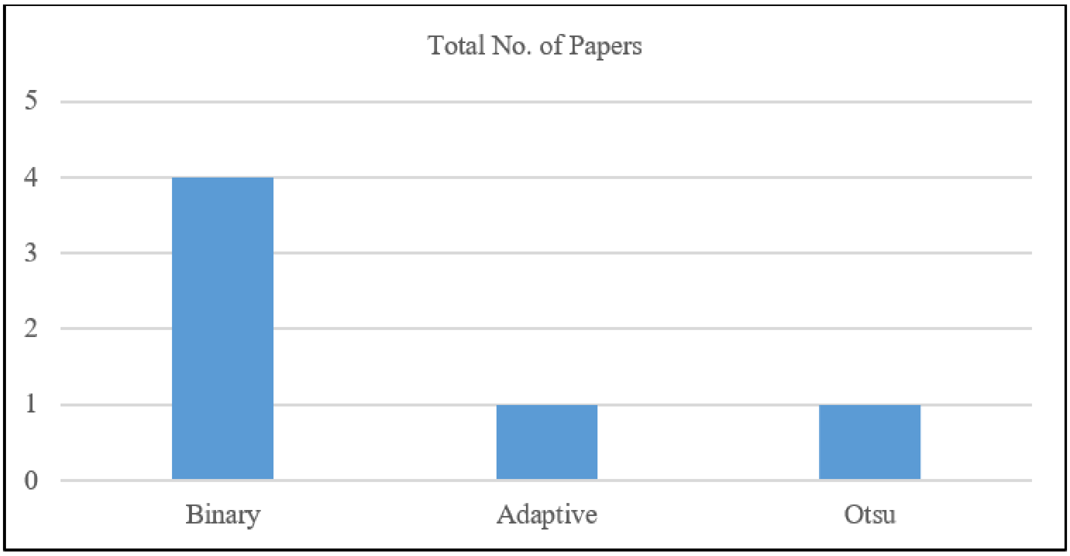


Figure 3.9: Type wise distribution of thresholding techniques used in the articles

3.3.5 Dataset wise Distribution

Several competition and resources from universities are open for all to work on the images of the brain tumor. For working on the segmentation of brain tumors in MRI scans, BRATS dataset is available for all which contains T1, T2 and FLAIR images.

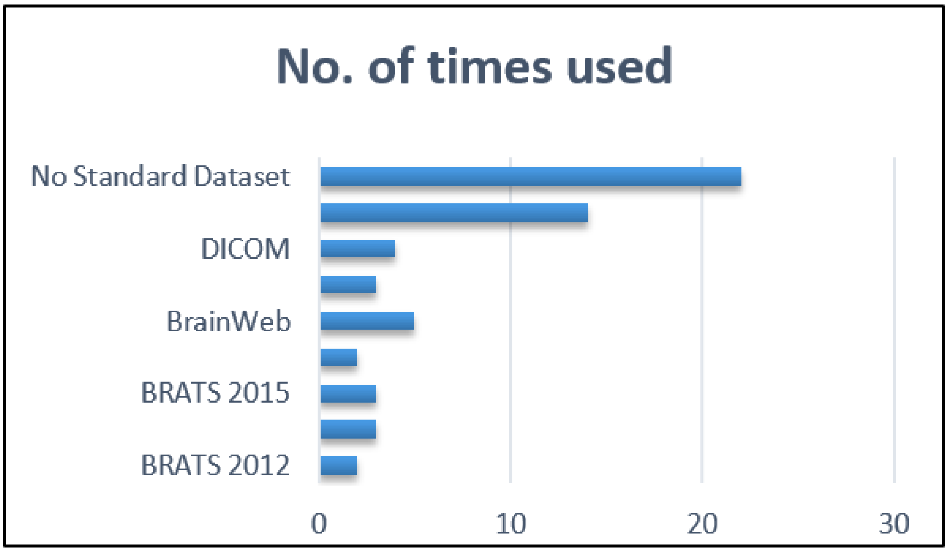


Figure 3.10: Distribution of the data-sets used in the articles

Several sources of databases were used in the papers. Figure 3.10 represents the names of the data-sets used by the articles and distribution of the dataset based on their usage. And from this figure, it can be deduced that most of the researchers did not use any standard dataset.

3.3.6 A Statistical Survey

We worked on a total of 52 papers, in which 10 of them used neural network based segmentation and the rest of them worked on various traditional segmentation techniques. Gleaning all the essential and statistical information, we compose a table which reflects the bottom line of the articles about their performance and relative work.

Table 3.1: Result-wise distribution of Clustering-based segmentation technique

Technique	Authors & Year	No of Images used	Citation	Result	Total
Pillar K-Means	(Rajesh et al., 2017)[43]	Unspecified	0	Computational Time: 0.7020 s (for k=3), 0.5304 s (for k=4)	1
K-Means	(Song et al., 2016)[16]	125	4	Dice: 80.7 ± 3.9, Sensitivity: 95.2 ± 1.2	7
	(Telrandhe et al., 2016)[30]	Unspecified	14	Momentum factor is 0.9 and total numbers of epochs are 500	
	(Wu et al., 2007)[44]	Unspecified	125	Separation of the lesion and detection of the tumor using the features derived from CIELab color model	
	(Selvakumar et al., 2012)[45]	Unspecified	135	Find the stage of the tumor by Area Calculation	
	(Liu et al., 2015)[46]	Unspecified	9	Jaccard Similarity Coefficient: 0.8702, 0.7619, 0.7300 for WM, GM, CSF.	
	(Vijay et al., 2013)[47]	100	54	Accuracy: 95%	
	(Rajesh et al.,2017)[43]	Unspecified	0	Computational Time: 1.2636 s (for k=3), 1.1232 s (for k=4)	

Table 3.1 continued from previous page

Technique	Authors & Year	No of Images used	Citation	Result	Total
Histogram Based Clustering	(Wu et al., 2007)[44]	Unspecified	125	Separation of the lesion and detection of the tumor using the features derived from CIELab color model	1
Fuzzy C Means	(Rajesh et al.,2017)[43]	Unspecified	0	Computational Time: 6.9732 s (for k=3), 11.8561 s (for k=4)	8
	(Rao et al., 2017)[48]	200	7	Dice Co-efficient: 79% (avg.), 88% (max.)	
	(Parveen et al., 2015)[49]	120	27	Accuracy: 91.66% for linear kernel; 83.33% for quadratic kernel, 87.50% for polynomial kernel	
	(Logeswari et al., 2010)[50]	Unspecified	76	Execution Time: 28.364 for 11×11 pixel window	
	(Nandha et al., 2010)[51]	120	87	Accuracy: 92.3%	
	(Gordillo et al., 2010)[52]	20	23	Jaccard Similarity Measure: 71% (lowest), 93% (highest)	
	(Sompong et al., 2013)[53]	Unspecified	7	Dice co-efficient: 84%	
	(Lawrence et al., 1992) [54]	12	338	False Negative: 20% for FCM/AFCM to 35% for FFCC.	

Table 3.1 continued from previous page

Technique	Authors & Year	No of Images used	Citation	Result	Total
Spatial fuzzy C means	(Devkota et al., 2017)[15]	19	2	Accuracy: 92%	3
	(Bing et al., 2011)[55]	3	345	Level set evolution stabilizes automatically once it approaches the genuine boundaries, suppressing boundary leakage and alleviates manual intervention	
	(Kanade et al., 2015)[56]	15	15	Low error rates	
KIFCM	(Abdel-Maksoud et al.,2015) [57]	204	136	Accuracy on Dataset1: 90.5%, Dataset 2: 100%, Dataset 3: 100%	1
Enhanced Possibilistic Fuzzy C-Means (EPFCM)	(Rajendran et al., 2011) [21]	15	46	Average similarity metrics: 95.3%, Jaccard index: 82.1%	1

Table 3.2: Result-wise distribution of Edge-based segmentation technique

Technique	Authors & Year	No of Images used	Citation	Result	Total
Edge Detection	(Debnath et al., 2011)[37]	12	36	Mean, Median, Std. Dev. And number of white pixels measured to detect the tumor	1
Canny Edge Detection	(Malathi, 2014)[38]	Unspecified	0	Dice: 90.13% - 93.26%	2

Table 3.2 continued from previous page

Technique	**Authors & Year**	**No of Images used**	**Citation**	**Result**	**Total**
	(Badran et al., 2017)[17]	102	70	False Positive: 18.75%	
Watershed	(Mustaqeem et al.,2012)[16]	60	152	Hybrid Segmentation technique is used	5
	(Karthik et al., 2015)[39]	Unspecified	14	Accuracy: 90%	
	(Viji et al., 2011)[58]	Unspecified	29	App Volume of 4075.65 mm and 1072.60 mm	
	(Mittal et al., 2017)[42]	Unspecified	0	Correctly locate the tumor based on intensity	
	(Dilber et al., 2016)[59]	2	152	Percentage of successfully pixel count: 86.25% - 93.21%	

Table 3.3: Result-wise distribution of Thresholding-based segmentation technique

Technique	**Authors & Year**	**No of Images used**	**Citation**	**Result**	**Total**
Binary Thresholding	(Debnath et al., 2011)[37]	12	36	Mean, Median, Standard Deviation And the number of white pixels measured to detect the tumor	4
	(Ilhan et al., 2017)[11]	100	4	Accuracy: 96%	
	(Mustaqeem et al., 2012)[16]	Unspecified	152	Hybrid Segmentation technique is used	
	(Akram et al., 2011)[35]	100	45	Accuracy: 97%	
Adaptive Thresholding	(Badran et al., 2017)[17]	102	70	False Positive: 18.75%	1
Otsu Thresholding	(Mittal et al., 2017)[42]	Unspecified	0	Correctly locate the tumor based on intensity	1

Table 3.4: Result-wise distribution of Histogram-based segmentation technique

Authors & Year	No of Images used	Citation	Result	Total
(Devkota et al., 2017)[15]	19	2	Accuracy 92%	8
(Debnath et al., 2011)[37]	12	36	Mean, Median, Std. Dev. And the number of white pixels measured to detect the tumor	
(Nabizadeh et al., 2017)[60]	Unspecified	8	Execution Time: 140 ms	
(Song et al., 2016)[16]	125	4	Dice: 80.7, Sensitivity: 95.2	
(Wu et al., 2007)[44]	Unspecified	125	Separation of the lesion and detection of the tumor using the features derived from CIELab color model	
(Logeswari et al., 2010)[50]	Unspecified	76	Execution Time: 28.364 for 11 × 11 pixel window	
(Bauer et al., 2011)[61]	10	220	DSC: 0.84 (intrapatient case), 0.77 (interpatient leave-one-out case)	
(Viji et al., 2011)[58]	Unspecified	29	App Volume of 4075.65 mm and 1072.60 mm	

Table 3.5: Result-wise distribution of Neural Network based segmentation technique

Authors & Year	No of Images used	Citation	Result	Total
(Sobhaninia et al., 2018) [22]	3064	0	Dice Score: 0.73 (Single Network), 0.79 (Multiple Networks)	
(Pereira et al., 2016) [62]	392	238	Accuracy: 70%	

10

Table 3.5 continued from previous page

Authors & Year	No of Images used	Citation	Result	Total
(Havaei et al., 2017) [63]	65	292	Dice Co-efficient varies from 0.80 - 0.88	
(Dina et al., 2012) [19]	82	44	Accuracy: 100%	
(Othman et al., 2011) [20]	35	11	Accuracy: 98%	
(Lawrence et al., 1992) [54]	12	338	False Negative: 20% for FCM/AFCM to 35% for FFCC.	
(Corso et al., 2008) [64]	20	172	Accuracy: 70%	
(Shenbagarajan, et al., 2016) [65]	80	13	Accuracy: 94%	
(Elisee et al., 2017) [66]	613	6	Accuracy: 93.20%	
(Shree et al., 2017) [36]	25	1	Accuracy: 95%	

Table 3.6: Result-wise distribution of Contour-based segmentation technique

Technique	Authors & Year	No of Images used	Citation	Result	Total
Parametric deformable active contour model model with gradient vector field(GVF)	(Rajendran et al., 2011)[21]	15	46	Average similarity metrics: 95.3%, average Jaccard index: 82.1%	1
Content Based Active Contour Model	(Jainy et al., 2012)[67]	428	96	Gives substantial results for homogeneous tumors against different and similar background	1
Region Based Active Contour	(Shenbagarajan, et al., 2016) [65]	80	13	Accuracy: 94%	1

Table 3.6 continued from previous page

Technique	Authors & Year	No of Images used	Citation	Result	Total
Localized Region based active contour	(Elisee et al., 2017) [66]	613	6	Accuracy: 93.20%	1

Table 3.7: Result-wise distribution of other segmentation techniques

Technique	Authors & Year	No of Images used	Citation	Result	Total
Tumor Growth Model, Lattice-Boltzmann Method	(Pei et al., 2017)[18]	100	1	Mean DSC with tumor cell density: 0.82122 (complete), 0.685811 (core), 0.812388 (enhancing)	1
Masking based on Symmetric Property	(Mariam et al.,2017)[68]	40	3	Accuracy: 95.5%	1
Fuzzy Logic Gaussian Mixture Model	(Ji et al, 2012)[69]	80	89	Jaccard Similarity: 0.8138 (for GM), 0.9339 (for WM)	1
Local Independent Projection	(Huang et al., 2014)[70]	120	73	Dice Similarity: 79.8 ± 17.0 for high grade tumor	1
Berkeley Wavelet Transformation	(Bahadure et al., 2017)[71]	201	39	Accuracy: 96.51%, Specificity: 94.2%, Sensitivity: 97.72%, Dice co-efficient: 0.82	1
Self-Organizing Map	(Demirhan et al., 2015)[72]	63	54	Average Dice for Tumor: 60.92%	1
Graph Cut	(Binaghi et al., 2014)[73]	Unspecified	2	Jaccard index: 0.867 (for inter-patient), 0.031 (for intra-patient)	1

3.4 Summary

In this chapter, we presented a statistical analysis of various brain tumor segmentation techniques such as- Layer based, Region based, Thresholding based, Edge detection based segmentation tecniques which was presented with necessary tables, figures and graphical representations along with their descriptions. Furthermore, we constituted a numerical representation of the research papers along with their performance measures where neural network based segmentation has been carried out.

Chapter 4

Background Study

4.1 Medical Image

Medical imaging means the visualization of body parts, tissues, or organs, for use in clinical diagnosis, treatment and disease monitoring. Imaging techniques encompass the fields of radiology, nuclear medicine and optical imaging and image-guided intervention.

4.2 Importance of Medical Image Analysis

Digital image processing is one of the modern and advance technology which process on the photos or videos. Nowadays x-ray is the main important application of Digital Image Processing. Before x-ray, it was very difficult to examine human bone of his body because the doctor has to cut the body skin and flesh to know about the bone of the body whether it is crack or damaged or not. In every perspective and every lane, image processing can be applied whether it is for security or for personal use.

From the discovery of X-ray by Roentgen in 1895 to the present day imaging tools like Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), the technology has progressed much. The advances in imaging technology will continue as time progresses. However, today the focus of systems is shifting from medical imaging focus from the generation and acquisition of images to post-processing and management of image data. This is stimulated by the need to make efficient use of the data that already exists. Recent progress in imaging research has shown the potential the technology can have to improve and transform many aspects of clinical medicine.

4.3 Machine Learning in Medical Image Analysis

Advances in both imaging and computers have synergistically led to a rapid rise in the potential use of artificial intelligence in various radiological imaging tasks, such as risk assessment, detection, diagnosis, prognosis, and therapy response, as well as in rapid disease discovery. Computer-extracted (radio-mic) features can serve as input to machine learning algorithms (i.e., computer algorithms that "learn" a specific task given specific input data). With such machine learning methods, multiple radio-mic features are merged into a single value, such as a tumor signature, which might be related to the likelihood of disease state [74]. Various machine learning techniques have been applied across the decades, for example, linear discriminant analysis, support vector machines, decision trees, and random forests, and neural networks. Reviews of machine learning have been written over the past many years including those that serve as tutorials to new investigators into the field [75] [76].

4.4 Image Processing for Brain Tumor Segmentation

One of the most challenging as well as demanding task is to segment the region of interest from an object and segmenting the tumor from an MRI Brain image is an ambitious one. Brain tumor segmentation from MRI is one of the most crucial tasks in medical image processing as it generally involves a considerable amount of data. Moreover, the tumors can be ill-defined with soft tissue boundaries. So it is a very extensive task to obtain the accurate segmentation of tumors from the human brain.

Image processing helps to improve the quality of the MRI images and subsequently to perform features extraction and classification. For brain tumor segmentation, image processing involves various steps such as skull stripping, pre-processing, tumor contouring etc.

4.4.1 Skull Stripping

Skull removing is one of the most important and essential steps for detecting the brain tumor and extracting the features. It is the process of eliminating all the non-brain tissues in the MRI images. By this method, it is possible to remove additional cerebral tissues such as fat, skin and skull in the brain images [77]. Some of the popular skull removing techniques are-image contour, skull stripping based on segmentation, morphological operation and based on histogram analysis and using threshold value.

4.4.2 Pre-processing

In medical image processing, medical images are corrupted by different type of noises. It is very important to obtain precise images to facilitate accurate observations for the given application [78]. Pre-processing steps include filtering, morphological operations etc.

- **Filtering:** Image filtering is useful for many applications, including smoothing, sharpening, removing noise, and edge detection. A filter is defined by a kernel, which is a small array applied to each pixel and its neighbors within an image. The process used to apply filters to an image is known as convolution, and may be applied in either the spatial or frequency domain. Spatial domain filtering can be classified into two types-smoothing and sharpening filters, according to their outputs.
  - **Smoothening Filter:** Smoothening spatial filters are used for noise reductions and blurring operations. It includes box filter, Gaussian blur filter, median blur filter, bilateral filter etc.
    1. **Box Filtering** is basically an average-of-surrounding-pixel kind of image filtering.
    2. **Gaussian blur filtering** is a 2D convolution operator that is used to 'blur' images and remove detail and noise. In this sense it is similar to the mean filter, but it uses a different kernel that represents the shape of a Gaussian ('bell-shaped') hump. The degree of smoothing is determined by the standard deviation of the Gaussian.
    3. **Median blur filter** is normally used to reduce noise in an image. Instead of simply replacing the pixel value with the mean of neighboring pixel values, it replaces it with the median of those values.
    4. **Bilateral filter** is a non-linear, edge-preserving, and noise-reducing smoothing filter for images. This preserves sharp edges.
  - **Sharpening Filter:** Sharpening spatial filters seek to highlight fine details. It removes blurring from images and highlights edges. Sharpening filters are based on spatial differentiation. It includes laplacian filter, sobel filter, difference filter etc.
    1. **Laplacian filters** are derivative filters used to find areas of rapid change (edges) in images.
    2. **Sobel filters** are typically used for edge detection.
    3. **Difference filters** enhance the details in the direction specific to the mask selected.

- **Morphological Operations:** Morphology means pixel shape based analysis. The objective of using morphological operations is to remove the imperfections in the structure of the image [79]. The most basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries. Morphological operations also include opening, closing, hit and miss transform etc.

  Morphological erosion removes islands and small objects so that only substantive objects remain.

  $$I \oplus H = \{(p+q) \mid for \ every \ p \in I, \ q \in H\} \tag{4.1}$$

  where I is the original image and H is the structuring element.

  Dilation is the opposite of Erosion. The value of the output pixel is the maximum value of all pixels in the neighborhood. In a binary image, a pixel is set to 1 if any of the neighboring pixels have the value 1. Morphological dilation makes objects more visible and fills in small holes in objects.

  $$I \ominus H = \{p \in Z^2 \ |(p+q) \in I, \ for \ every \ q \in H\} \tag{4.2}$$

  where I is the original image and H is the structuring element.

  In brain tumor segmentation, morphological operations are needed for removing soft tissue boundaries and for effective segmentation of the tumor portion.

- **Segmentation:** Segmentation means dividing the image into different regions and separating objects from background. Accurate segmentation of objects of interest in an image greatly facilitates further analysis of these objects. There are various types of segmentation algorithms such as edge based, thresholding based, region based, clustering based etc.
  Figure 4.1 represents the segmentation of an object from its background.

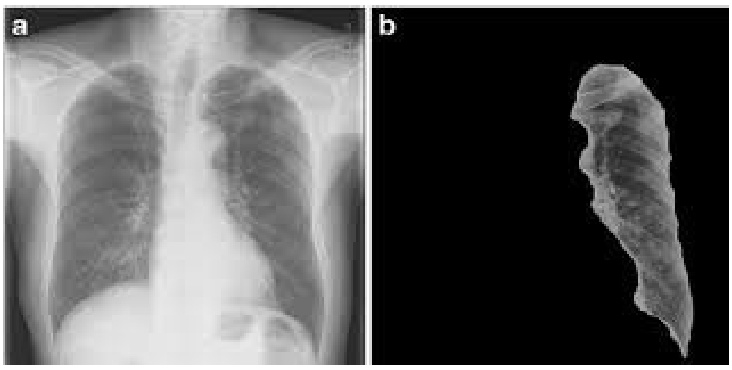


Figure 4.1: Segmentation of an object from background [80]

– **Fuzzy C-Means (FCM) Clustering:** Fuzzy C-Means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. One of the most widely used fuzzy clustering algorithms is the Fuzzy C-means clustering (FCM) Algorithm.

  Here, Clustering or cluster analysis involves assigning data points to clusters such that items in the same cluster are as similar as possible, while items belonging to different clusters are as dissimilar as possible. Clusters are identified via similarity measures. These similarity measures include distance, connectivity, and intensity [81].

  It is based on minimization of the following objective function : where m is any real number greater than 1, U_{ij} is the degree of membership of x_i in the cluster j, x_i is the i^{th} of d-dimensional measured data, c_j is the d-dimension center of the cluster is any norm expressing the similarity between any measured data and the center.
  Figure 4.2 represents a Fuzzy C-means segmented image.

Figure 4.2: Fuzzy C- Means Segmented Image [83]

– **K-Means Clustering:** K-Means clustering is a type of unsupervised learning. So, when we have unlabeled data, we can use K-Means Clustering. The K-means clustering algorithm is used to find groups which have not been explicitly labeled in the data. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K [82]. Based on the features those are provided, this algorithm iteratively assigns each data point to one of K groups. These are the results that K-Means Clustering algorithm gives: the centroids of the K clusters, which can be used to label new data and labels for the training data (each data point is assigned to a single cluster).

  A cluster refers to a collection of data points aggregated together because of certain similarities. A centroid is the imaginary or real location representing the center of the cluster. Each centroid of a cluster is a collection of feature values which define the resulting groups. Examining the centroid feature weights can be used to qualitatively interpret what kind of group each cluster represents. Figure 4.3 is the visual representation of the segmentation which was done by K-means clustering technique.

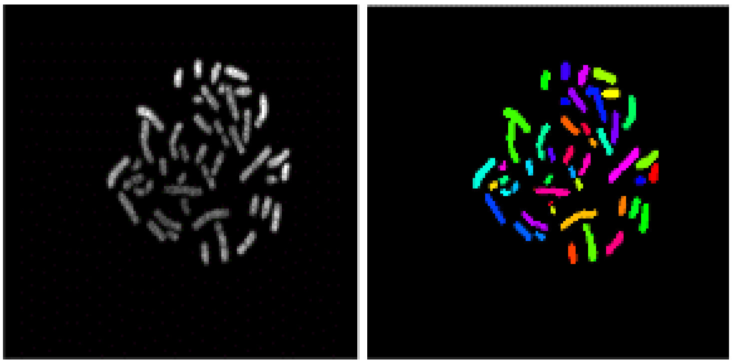


Figure 4.3: Segmentation Using K-Means Clustering [84]

– **Watershed Segmentation:** The watershed is a classical algorithm used for segmentation, that is, for separating different objects in an image. Watershed is a transformation on gray-scale images. The aim of this technique is to segment the image, typically when two regions-of-interest are close to each other i.e. their edges touch [14]. This technique of transformation treats the image as a topographic map, with the intensity of each pixel representing the height. For instance, dark areas can be intuitively considered to be 'lower' in height, and can represent troughs. On the other hand, bright areas can be considered to be 'higher', acting as hills or as a mountain ridge [14]. It is one of the best methods to group pixels of an image on the basis of their intensities. Pixels falling under similar intensities are grouped together. We can distinct some certain portion of the image for the perspective.
Figure 4.4 shows the result of watershed segmentation.

Figure 4.4: Segmentation Using Watershed Algorithm [85]

4.5 Traditional Machine Learning Classifiers

Evaluation of our proposed model is done by applying a classification algorithm. So building up a model which can detect or segment the tumor is half of the task that we have done. Justification of the model is done through model evaluation or assessment of the proposed model. We have done an approach which applied six traditional machine learning algorithm. To justify our model we have used these various types of classifiers.

We have used six traditional machine learning classifiers which are K-Nearest Neighbor, Logistic Regression, Multi-layer Perceptron, Naive Bayes, Random Forest, and Support Vector Machine to get the accuracy of the tumor detection of our proposed model.

4.5.1 K-Nearest Neighbor

KNN is one of the simplest of classification algorithms available for supervised learning. The idea is to search for the closest match of the test data in feature space. Typically the object is classified based on the labels of its k nearest neighbors by majority vote. If k=1, the object is simply classified as the class of the object nearest to it. When there are only two classes, k must be an odd integer. However, there can still be times when k is an odd integer when performing multi-class classification. After we convert each image to a vector of fixed-length with real numbers, we used the most common distance function for KNN which is Euclidean distance between two points x and y and it can be written as:

$$d(x,y) = |x-y| = \sqrt{(x-y).(x-y)} = (\sum_{i=1}^{m}(x_i - y_i)^2)^{\frac{1}{2}} \tag{4.3}$$

4.5.2 Logistic Regression

Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

Logistic regression uses an equation as the representation, very much like linear regression. Input values (x) are combined linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y). A key difference from linear regression is that the output value being modeled is a binary values (0 or 1) rather than a numeric value. Below is the logistic regression equation [93]:

$$y = \frac{(e^{b_0+b_1x})}{1+e^{(b_0+b_1x)}} \tag{4.4}$$

where y is the predicted output, b_0 is the bias or intercept term and b_1 is the coefficient for the single input value (x).

4.5.3 Multi-layer Perceptron

A multi layer perceptron (MLP) is a feed forward artificial neural network that generates a set of outputs from a set of inputs. A MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer [94]. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP uses back-propagation for training the network. Multi-layer are most of the neural networks expect deep learning. It uses one or two hidden layers.

4.5.4 Naive Bayes

Naive Bayes is a machine learning algorithm for classification problems. It is based on Bayes' probability theorem to predict the class of unknown data set. It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. The class with the highest probability is considered as the most likely class.

In a machine learning classification problem, there are multiple features and classes, say, $C_1, C_2, ..., C_k$. The main aim in the Naive Bayes algorithm is to calculate the conditional probability of an object with a feature vector $x_1, x_2, ..., x_n$ belongs to a particular class C_i. And the equation can be written as [95]:

$$P(C_i|x_1, x_2, ..., x_n) = \frac{P(x_1, x_2, ..., x_n|C_i) * P(C_i)}{P(x_1, x_2, ..., x_n)}, \ for \ \ 1 \leqslant i < k \tag{4.5}$$

or we can write,

$$P(C_i|x_1, x_2, ..., x_n) = (\prod_{j=1}^{j=n} P(x_j|C_i)) \ * \ \frac{P(C_i)}{P(x_1, x_2, ..., x_n)}, \ for \ \ 1 \leqslant j \leqslant n \tag{4.6}$$

4.5.5 Random Forest

Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where we join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision trees.

Random decision forests correct for decision trees' habit of over-fitting to their training set [96]. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [96].

4.5.6 Support Vector Machine (SVM)

Support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier [97]. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

In addition to performing linear classification, SVM can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

4.6 Deep Learning

Deep Learning is a subset of machine learning algorithms that is very good at recognizing patterns but typically requires a large number of data. Deep learning excels in recognizing objects in images as it is implemented using three or more layers of artificial neural networks where each layer is responsible for extracting one or more features of the image.

4.6.1 Neuron

The basic building unit of neural networks are artificial neurons, which imitate human brain neurons. These artificial neurons are powerful computational units that have weighted input signals and produce an output signal using an activation function. These neurons are spread across the several layers in a neural network.

Figure 4.5 depicts the information of a neuron:

A neuron has three parameters, namely:

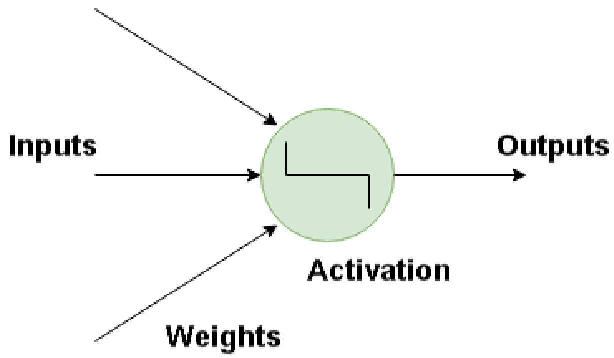


Figure 4.5: Basic structure of a neuron [86]

- **Weight:** When a signal (value) arrives, a neuron gets multiplied by a weight value. If a neuron has three inputs, it has three weight values which can be adjusted during training time [87].
- **Bias:** It is an extra input to neurons and it is always 1, and has its own connection weight. This makes sure that even when all the inputs are none (all 0's) there is going to be an activation in the neuron.
- **Activation Function:** Activation functions are also known as transfer functions. It helps in classification or partition. It uses a threshold, according to this threshold, we make the division into two or many classes.

4.6.2 Artificial Neural Network

Deep learning consists of artificial neural networks that are modelled on similar networks present in the human brain. These neural networks work in multiple layers so this kind of machine learning is called deep learning.

Artificial neural networks are a way of calculating an output from an input (a classification) using weighted connections ("synapses") that are calculated from repeated iterations through training data. Each pass through the training data alters the weights such that the neural network produces the output with greater "accuracy" (lower error rate). The combination of working memory and speed is crucial when we're doing hundreds of thousands of matrix multiplications. Figure 4.6 represents a simple artificial neural network:

An artificial neural network generally has three layers. Layers are made up of some nodes which are interconnected. The three layers of ANN are described in the following:

- **Input Layer:** This layer consists of neurons and they just receive the inputs and pass

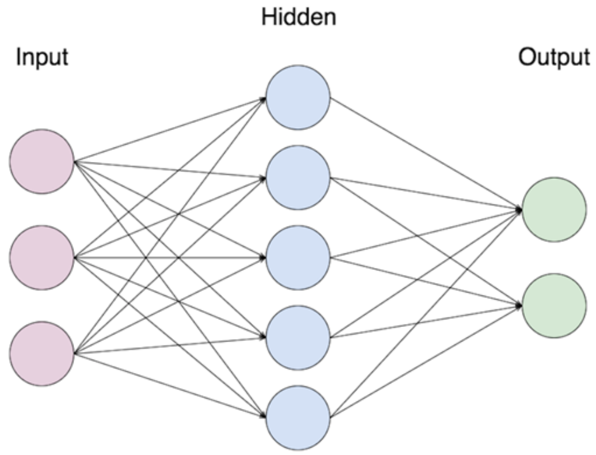


Figure 4.6: A simple Artificial Neural Network [86]

it to the next layer. The number of layers in the input layer should be equal to the attributes or features in the dataset.

- **Hidden Layer:** In between input and output layer there are hidden layers based on the type of model. Hidden layers contain vast number of neurons. The neurons in the hidden layer apply transformations to the inputs before passing them to the next layer. As the network is trained the weights get updated, to be more predictive. The actual processing of the data is done via a system of weighted connections in the hidden layer. The hidden layers are linked to the output layer.
- **Output Layer:** The output layer is the predicted feature, or class in a classification problem, it basically depends in the type of the built model. The output layer gives the output based on the information passed from the hidden layer.

The internal structure of ANN can be changed by itself based on the information passing through it. This is done by the adjustment of the weights. Every connection in the neural network generally has a weight that controls the signal between the two neurons. If the output is good, the adjustment is no longer needed, but if the output is poor, then the system adapts by changing the weights to improve the output. The performance evaluation of the output is done by the system by comparing the output with the original output given before in the training mode [6].

4.6.3 Using Neural Networks for Images

Neural network can be used to recognize or detect object category but it will require more works to uniquely identify an object. A classical neural network requires to input a set of features extracted from each of the image. Deep neural network (DNN) works with image pixels.

An image can be represented as a matrix, each element of the matrix containing color information for a pixel. The matrix is used as input data into the neural network. The small dimensions of the images, to easily and quickly help learning, establish the size of the vector and the number of input vectors. The transfer function used is also called an activation function. Processing of images with artificial neural network involves different processes, such as:

- **Image pre-processing:** It is an operation which shows a picture (contrast enhancement, noise reduction) with the same dimensions as the original image. The objective of images' pre-processing with ANN consists in improving, restoring or rebuilding images.
- **Data reduction or feature extraction:** This step involves extracting a number of features smaller than the number of pixels in the input window. The operation consists in compressing the image followed by extracting geometric characteristics such as edges, corners, joints, facial features, etc.
- **Segmentation:** Segmentation means the division of an image into different meaningful regions.
- **Recognition:** It involves the determination of objects in an image and their classification.

Processing images with artificial neural networks successfully resolve the problems of classification, identification, authentication, diagnostics, optimization and approximation [87].

4.6.4 Types of Neural Networks

There are several kinds of artificial neural networks. These type of networks are implemented based on the mathematical operations and a set of parameters required to determine the output. Some of the popular neural networks are [86]:

- **Feedforward Neural Network:** This neural network is one of the simplest forms of ANN, where the data or the input travels in one direction. The data passes through the input nodes and exit on the output nodes. This neural network may or may not have the hidden layers. In simple words, it has a front propagated wave and no back propagation by using a classifying activation function usually.
- **Radial Basis Function Neural Network:** Radial basis functions consider the distance of a point with respect to the center. RBF functions have two layers, first where the features are combined with the Radial Basis Function in the inner layer and then the

output of these features are taken into consideration while computing the same output in the next time-step which is basically a memory.

- **Kohonen Self Organizing Neural Network:** A self-organizing map (SOM) is a type of artificial neural network (ANN) that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map, and is therefore a method to do dimensionality reduction.

- **Recurrent Neural Network:** The Recurrent Neural Network works on the principle of saving the output of a layer and feeding this back to the input to help in predicting the outcome of the layer. This is also known as Long Short Term Memory.

- **Convolutional Neural Network:** Convolutional neural networks are similar to feed forward neural networks, where the neurons have learn-able weights and biases.

- **Modular Neural Network:** Modular Neural Networks have a collection of different networks working independently and contributing towards the output. Each neural network has a set of inputs which are unique compared to other networks constructing and performing sub-tasks. These networks do not interact or signal each other in accomplishing the tasks.

Among these mostly used neural networks, Convolutional Neural Networks are applied in techniques like signal processing and image classification techniques.

4.6.5 Convolutional Neural Network (CNN)

The basic idea of Convolutional Neural Network was introduced by Kunihiko Fukushima in 1980s [88]. Convolutional Neural Networks (ConvNets or CNNs) are a category of Neural Networks that have proven very effective in areas such as image recognition and classification. Computer vision techniques are dominated by convolutional neural networks because of their accuracy in image classification. CNN is a class of deep, feed-forward artificial neural networks (where connections between nodes do not form a cycle) & use a variation of multi-layer perceptrons designed to require minimal pre-processing.

ConvNet architectures make the explicit assumption that the inputs are images, which allows us to encode certain properties into the architecture. These then make the forward function more efficient to implement and vastly reduce the amount of parameters in the network.

ConvNets are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity.

The whole network still expresses a single differentiable score function: from the raw image pixels on one end to class scores at the other. And they have a loss function on the last layer.

(i) Why CNN is different than simple Neural Network: Convolutional Neural Networks have a different architecture than regular Neural Networks [86]. Regular Neural Networks transform an input by putting it through a series of hidden layers. Every layer is made up of a set of neurons, where each layer is fully connected to all neurons in the layer before and where neurons in a single layer function completely independently and do not share any connections between themselves. Finally, there is a last fully-connected layer, which is the output layer that represent the predictions. Regular Neural Networks do not scale well to full images.

Convolutional Neural Networks are a bit different. First of all, the layers are organized in three dimensions: width, height and depth. Further, the neurons in one layer do not connect to all the neurons in the next layer but only to a small region of it. Lastly, the final output will be reduced to a single vector of probability scores, organized along the depth dimension.

Moreover, CNNs perform convolution operation in case of matrix multiplication.

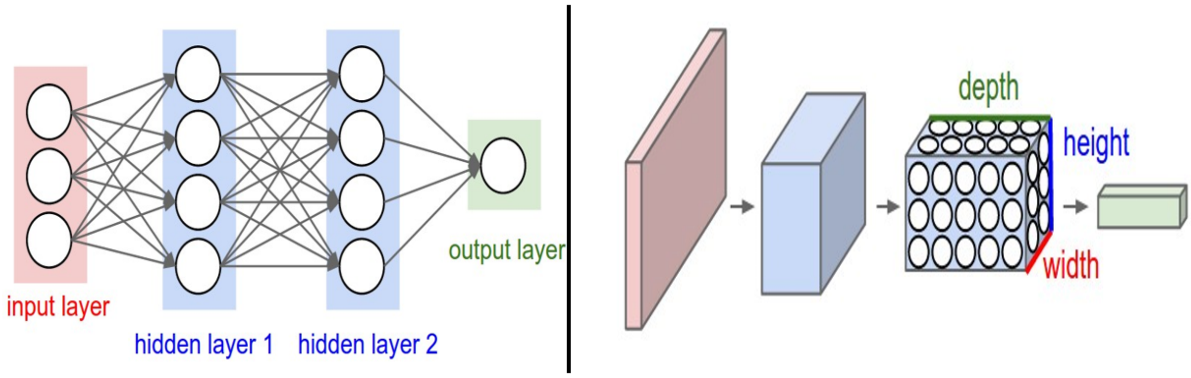


Figure 4.7: A simple neural network and A Convolutional Neural Network [86]

In the above, in figure 4.7, the left side represents a regular three Layer neural network. On the other hand, the right side of the figure represents a CNN which arranges its neurons in three dimensions (width, height, depth) [86].
Every layer of a CNN transforms the 3D input volume to a 3D output volume of neuron activation's. In this example, the red input layer holds the image, so its width and height would be the dimensions of the image, and the depth would be three (Red, Green, Blue channels).

(ii) The Convolution Operation: Convolutional Neural Networks perform a mathematical operation, known as convolution operation. Convolution is a mathematical operation on two functions (f and g) and it produces a third function. The convolution operation of f and g is denoted as f*g . It is defined as the integral of the product of the two functions after one is reversed and shifted. This operation is a particular kind of integral transform [89]:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau) \ d\tau \tag{4.7}$$

There are three elements that enter into the convolution operation [89]:

- **Input image:** It is the image that is given as an input.
- **Feature detector:** The feature detector is often referred to as a “kernel" or a “ filter". Sometimes a 5*5 or a 7*7 matrix is used as a feature detector.
- **Feature map:** The feature map is also known as an activation map. It is called feature map because it is also a mapping of where a certain kind of feature is found in the image.
  Figure 4.8 represents the convolution operation.

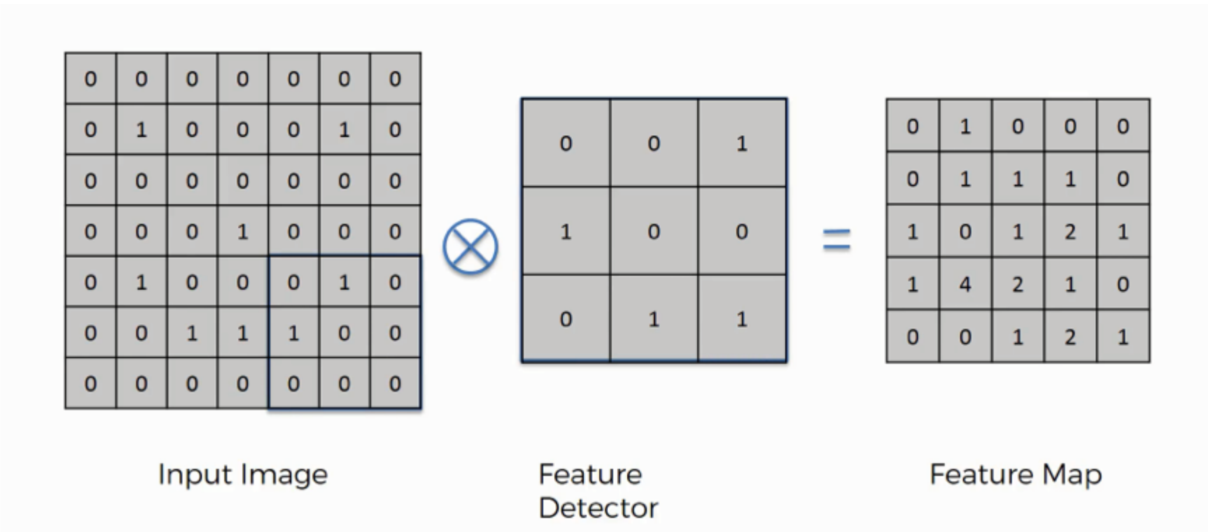


Figure 4.8: Convolution Operation [83]

(iii) The Convolution Arithmetic: Here the facts are shown how properties of an output image changes from input by some factors [90].

Let, Input Image Size = I, (means width I and Height I i.e. image size is I * I)
Filter Size = F, (Filter is F * F)
Number of Filters = K
Number of Strides = S
Amount of Zero Padding = P
Then the output image size will be:

$$O = \frac{I - F + 2P}{S} + 1 \tag{4.8}$$

- **Forward Pass for Convolutional Layer:** In forward pass of convolutional layer, a filter performs dot multiplication with all the parts of the input matrix one after another of filter size, then sum all the elements of the single dot product and adds the bias value with it and places the final value in the corresponding location of the output matrix for that dot product. This process continues across full input image with all filters and for each filter there comes an output.

- **Back-propagation for Convolutional Layer:** In back-propagation, the cost function is first found out, then this cost function measures the displacement with the output. After that, applying gradient descent on this function will update the filter value of the previous layer. This process continues until it reaches to the input layer.

The iteration of forward pass and back-propagation will continue until the network finds the desired output.

(iv) Layer's Used to a Build CNN Model: A simple CNN is a sequence of layers, and every layer of a CNN transforms one volume of activation's to another through a differentiable function. Three main types of layers used to build CNN architectures are:

- **Convolutional Layer:** The Convolutional layer is the core block of the Convolutional Neural Network. It has some special properties. It does most of the computational heavy lifting.The CONV layer's parameters consist of a set of learn-able filters. Every filter is small spatially (along width and height), but extends through the full depth of the input volume.

  For example, a typical 3X3 filter on a first layer of a ConvNet might have size 5*5*3 (i.e. 5 pixels width and height, and 3 because images have depth 3, the color channels). During the forward pass, each filter is convolved across the width and height of the input volume and dot products are computed between the entries of the filter and the input at any position. As the filters are slided over the width and height of the input volume, a 2-dimensional activation map will be produced that gives the responses of that filter at every spatial position. Intuitively, the network will learn filters that activate when they see some type of visual feature such as an edge of some orientation. These activation maps are stacked along the depth dimension and the output volume is produced.

  Figure 4.9 shows an example of convolution in the convolutional layer of CNN. Moreover, Convolutional layer has some basic features [90], such as:

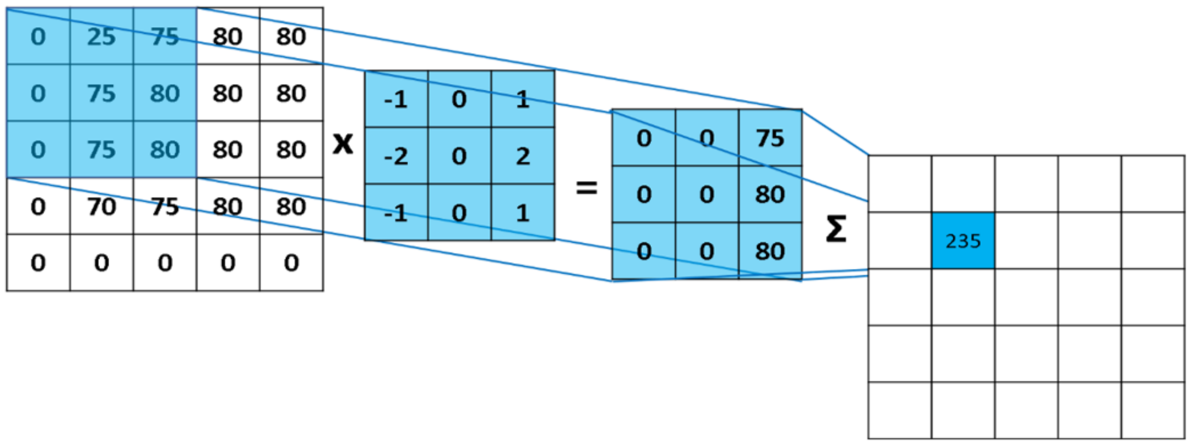


Figure 4.9: Convolution Operation of CNN [90]

- **Parameter Sharing:** Parameter sharing is sharing of weights by all neurons in a particular feature map.
- **Local Connectivity:** Local connectivity is the concept of each neural connected only to a subset of the input image (unlike a neural network where all the neurons are fully connected)

These features help to reduce the number of parameters in the whole system and makes the computation more efficient.

Three hyper-parameters control the size of the output volume of the convolutional layer. These parameters are:

- **Depth:** The depth of input volume in first layer is the number of color channels of that input image. If the input image is a color image then the depth is 3 that is the red channel, the green channel and the blue channel. If the image is black and white or gray-scale, then the depth is 1.

  The depth of the output volume is the number of filters that we use in the input.

  Figure 4.8 represents the depth which changes from 3 to 32 when we use 32 filters and the image size gets smaller.

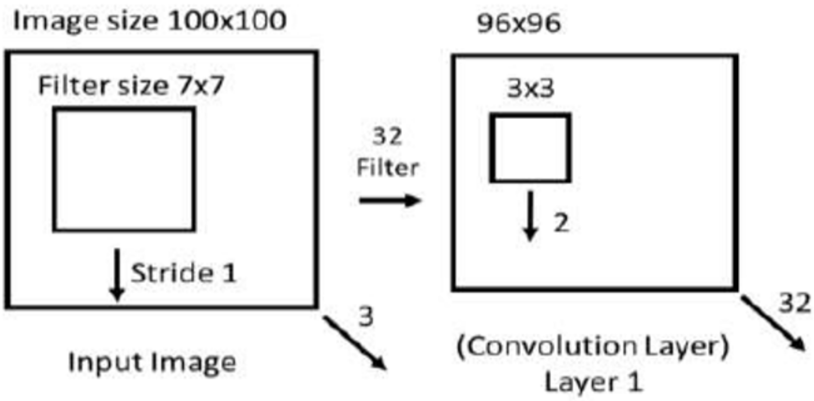


Figure 4.10: Depth changing from 3 to 32 using 32 filters [90]

- **Stride:** Stride is used to slide along width and height of the input image. When the stride is 1, then we move the filters one pixel at a time. When the stride is 2, then the filters jump 2 pixels at a time as we slide them around.

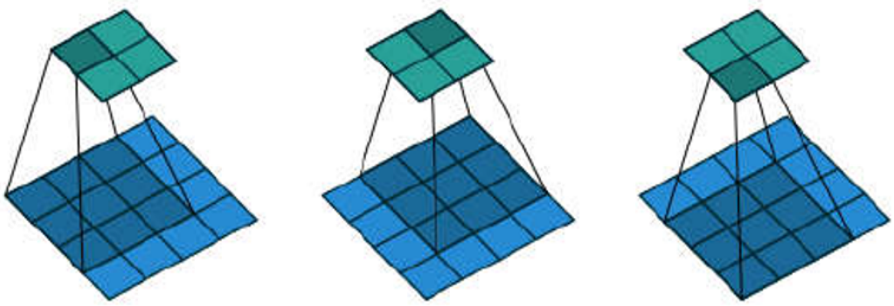


Figure 4.11: Sliding filter along an input image when the stride is 1 [90]

Figure 4.11 shows that a 2*2 filter moves along the input size of 4*4 through width and height when the stride is 1.

- **Zero Padding:** Sometimes in the input layer, we pad the input image with zero that is called zero-padding. Zero padding allows us to control the size of the input layer. If we don't use zero-padding, sometimes some property from the edges can be lost.
  Figure 4.12 illustrates the scenerio of zero-padding of an input.

- **Pooling Layer:** Pooling layer is another building block of CNN. Usually pooling layer is placed after the convolutional layer. Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network, and hence to also control over fitting. The Pooling Layer operates independently on every depth slice of the input and resizes it spatially. Pooling does not affect the depth dimension of the input volume.

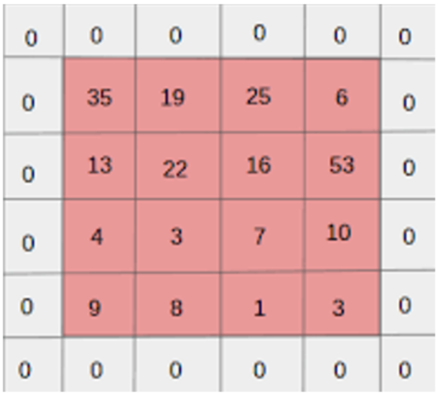


Figure 4.12: Zero-padding of an input (padding amount = 1) [90]

The task of pooling is done by is done by summarizing the sub-regions of the input using some methods like taking the average or the maximum or the minimum value of the sub-regions only. These methods are called pooling functions.

- **Different Kinds of Pooling Functions:**The pooling layer consists of some symmetric aggregation functions such as:
  * **Max Pooling:** It returns the maximum value from its rectangular neighborhood.
  * **Average Pooling:** It returns the maximum value from its rectangular neighborhood.
  * **Weighted Average Pooling:** It calculates its neighborhood weight based on distance from its center pixel.
  * **L2 Norm Pooling:** It returns the square root sum of its rectangular neighborhood.

  In most of the ConvNet architectures, Max Pooling is used to reduce the computational cost.

- **Pooling Layer Arithmetic:** Pooling layer works by sliding the window or filter across the input.
  Let, Spatial Extent = f
  Stride = s
  window size = w

  The equation [90] shows that the output size from the pooling layer will be,

$$O = \left\lfloor \frac{w - f}{s} \right\rfloor + 1 \tag{4.9}$$

- **Fully-Connected Layer:** In fully connected layer, every neuron is connected to its previous layer neuron like the neural network. Its activation is also computed by matrix multiplication with its weight followed by bias as like neural network. Usually, fully connected layer is a column vector.

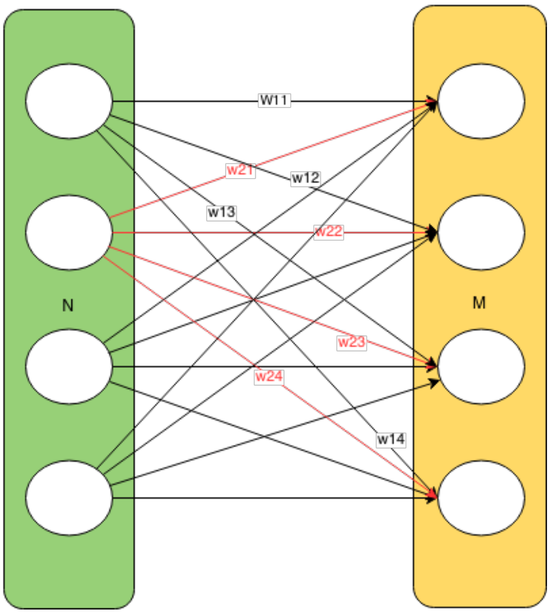


Figure 4.13: Fully Connected Layer of CNN [90]

Figure 4.13 shows the connection between two layers where the right one is the fully connected layer.

(iv) Activation Function: Activation functions are used to introduce non-linearity to neural networks. It squashes the values in a smaller range. For example, a sigmoid activation function squashes values between a ranges 0 to 1.

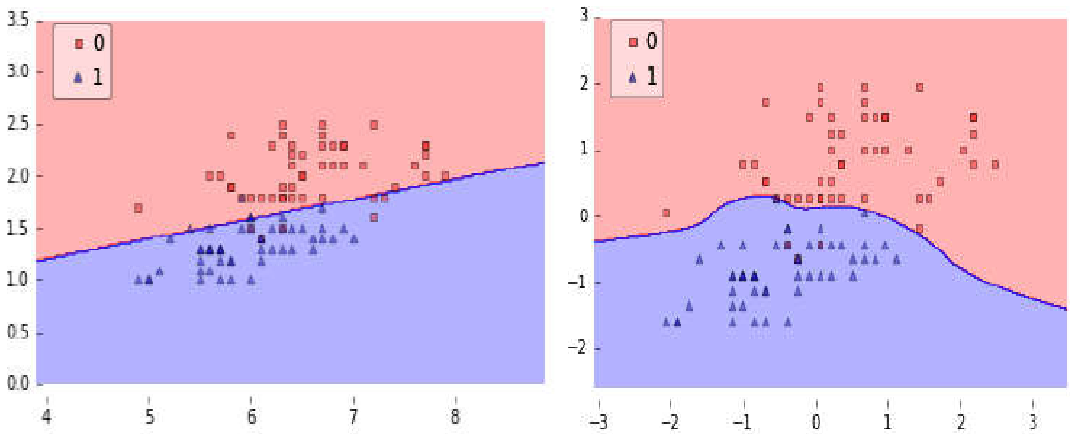


Figure 4.14: Before(left) and after(right) applying activation function [88]

Figure 4.14 demonstrates the importance of activation function. The figure on the left side is the scenario of applying logistic regression without the application of activation function and the figure on the right side is the same scenario with the application of activation function.

- **Commonly Used Activation Functions:** There are many activation functions used in deep learning industry. Here we will discuss about some activation functions in brief which are commonly in use.

  – **Sigmoid:** The Sigmoid function bounds the input value in between 0 to 1 range. For large positive number, it returns 1 and for large negative number, it returns 0. The mathematical representation of the sigmoid function is [91]:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{4.10}$$

Figure 4.15 represents the curve of the sigmoid function:

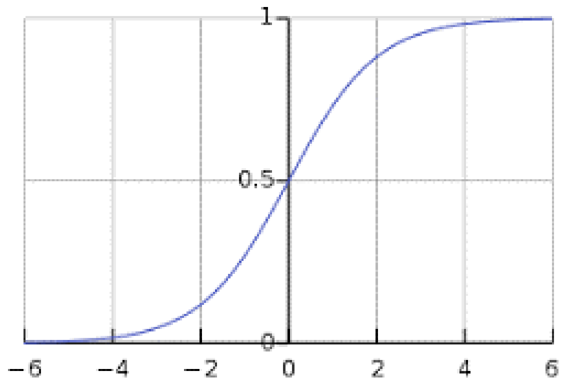


Figure 4.15: Curve of Sigmoid Function [91]

Advantages of Sigmoid Function:

1. It Provides smooth gradient and prevents any 'jump' in output values.
2. It normalizes the output of each neuron.
3. It enables clear predictions.

Disadvantages of Sigmoid Function:

1. It causes a 'vanishing gradient' problem. For very high or very low input values, there is almost no change to the prediction. This can result in the network refusing to learn further, or being too slow to reach an accurate prediction.
2. Outputs are not zero centered.

- **Tanh/Hyperbolic Tangent:** This function is like the sigmoid function. It bounds all real numbers to the range [-1, 1]. The tanh function is mainly used classification between two classes [91].

  Figure 4.16 represents the curve of the hyperbolic tangent function:

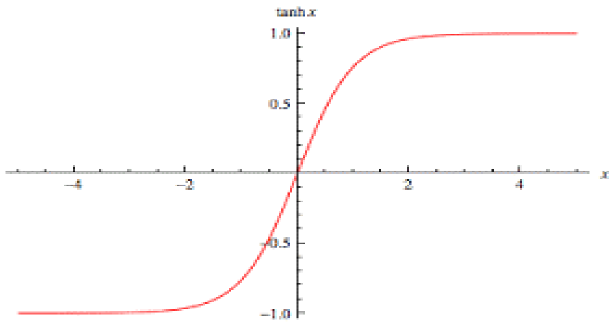


Figure 4.16: Curve of tanh Function [91]

Advantages of tanh Function:

1. This function is zero centered, that makes it easier to model inputs that have strongly negative, neutral, and strongly positive values.
2. Otherwise like the sigmoid function.

Disadvantages of tanh Function:
The disadvantages of tanh function is as like as the sigmoid function.

- **ReLU (Rectified Linear Unit):** ReLU refers to Rectified Linear Unit. It simply thresholds the input value to zero. For positive, it returns the number and for negative, it returns 0. In AlexNet architecture, after using ReLU as an activation function, it was 6 times faster than using the tanh function [91]. The formula of ReLu is as following [91]:

$$f(x) = max(0, x) \tag{4.11}$$

The following figure 4.17 represents the curve of ReLu:

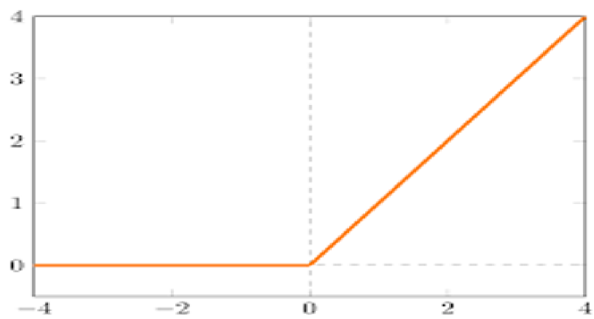


Figure 4.17: Curve of ReLU Function [91]

Advantages of ReLU Function:

1. It is computationally efficient and allows the network to converge very quickly.
2. Although it looks like a linear function, ReLU is non-linear.
3. It has a derivative function and allows for back-propagation.

Disadvantages of ReLU Function:

It introduces the Dying ReLU problem. When inputs approach zero, or are negative, the gradient of the function becomes zero, the network cannot perform back-propagation and cannot learn.

– **Leaky ReLU:** It is an improved version of ReLU. It solves the Dying problem of ReLU. For a positive number, it works like ReLU, and for a negative number, the number is multiplied by a very small number (i.e. 0.001).

The mathematical representation of this function is [91]:

$$f(x) = 1(x < 0)(ax) + 1(x \geq 0)(x) \tag{4.12}$$

The following figure-4.18 represents the curve of leaky ReLU:

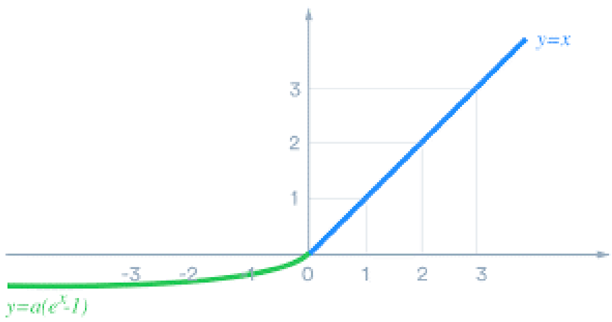


Figure 4.18: Curve of leaky ReLU Function [91]

Advantages of Leaky ReLU Function:

1. It prevents dying ReLU problem. This variation of ReLU has a small positive slope in the negative area, so it does enable back-propagation, even for negative input values.
2. Otherwise like ReLU.

Moreover, there are some other activation functions which are used in the field of deep learning.

(v) Regularization Function: In machine learning, regularization is a way to prevent over-fitting. Regularization reduces over-fitting by adding a penalty to the loss function. By adding this penalty, the model is trained such that it does not learn interdependent set of features weights.

If a simple model is not performing well at predicting due to poor generalization and a complex model may not perform well due to over-fitting. In this case, regularization helps to choose the preferred complexity for the model.

The two most important regularization techniques are Dropout and Batch Normalization.

- **Dropout:** Dropout is an approach to regularization in neural networks which helps reducing interdependent learning amongst the neurons. Dropout forces a neural network to learn more robust features that are useful in conjunction with many different random subsets of the other neurons.

  During the forward pass, Dropout temporarily removes some of the neurons. For example, if we fix dropout to 20%, then every 1 out of 5 neurons will be inactivated during the forward pass. During the backward pass, any weigh update will not be applied to these neurons.

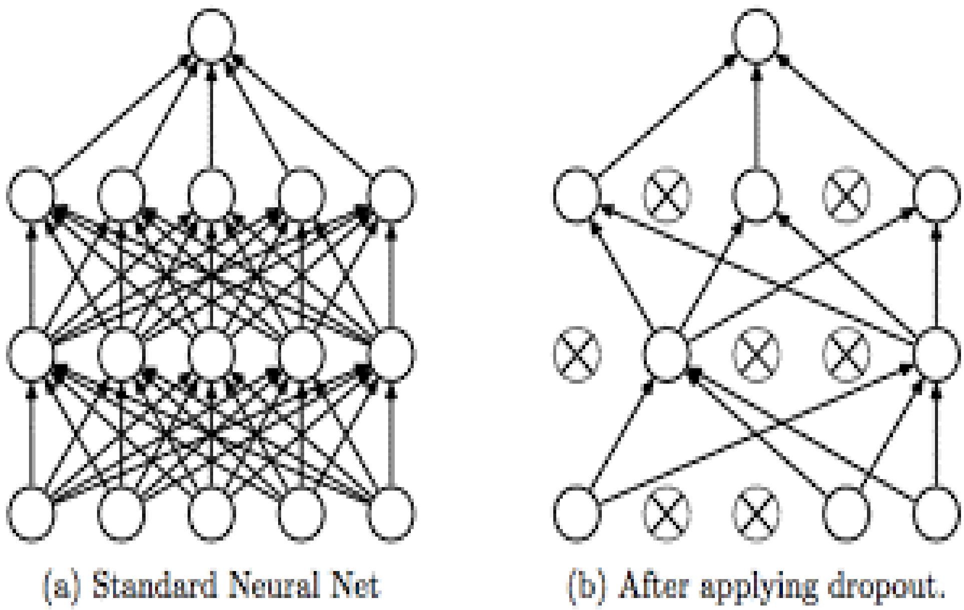


Figure 4.19: Network before and after Dropout [92]

  The above figure- 4.19 shows how dropout performs in the neural network. Cross neuron represents inactive neurons in this network.

- **Batch Normalization:** Batch Normalization is a method to reduce the internal co-variate shift in neural networks, leading to the possible usage of higher learning rates. In principle, the method adds an additional step between the layers, in which the output of the layer before is normalized. BN further prevents smaller changes to the parameters to amplify and thereby allows higher learning rates, making the network even faster.

  Batch normalization (BN) consists of two algorithms. The first algorithm is the transformation of the original input of a layer x to the shifted and normalized value y. The second algorithm is the overall training of a batch-normalized network.

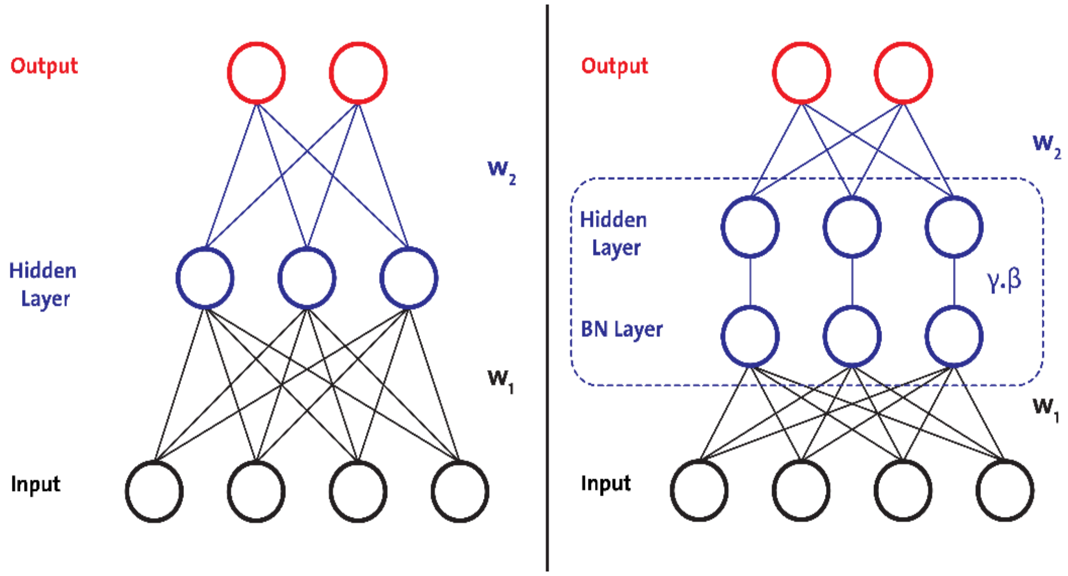


Figure 4.20: A Neural Network before and after Batch Normalization [92]

The above figure- 4.20 shows a standard neural network before and after having Batch Normalization.

(vi) Hyper-parameters of CNN architecture: In our proposed model, we used different types of hyper-parameters. Following is a brief discussion about the hyper-parameters.

- **Bias:** Bias is an error from erroneous assumptions in the learning algorithm. High bias can cause an algorithm to miss the relevant relations between features and target outputs (under-fitting).

- **glorot_uniform:** It is named as Xavier normal initializer. It draws samples from a uniform distribution within [-limit, limit].

- **Learning rate:** The learning rate or step size in machine learning is a hyper-parameter which determines to what extent newly acquired information overrides old information.

- **Beta_1 and Beta_2:** The algorithm calculates an exponential moving average of the gradient and the squared gradient, and the parameters beta1 and beta2 control the decay rates of these moving averages.
- **Epsilon:** It is a very small number to prevent any division by zero in the implementation.
- **Dropout or Decay weight:** A simple and powerful regularization technique for neural networks and deep learning models is dropout.
- **Amsgrad:** AMSGrad (in Keras, it is controlled by setting amsgrad = True for Adam optimizer), which uses the maximum of past squared gradients in order to allow the rarely-occurring minibatches with large and informative gradients to have greater influence on the overall direction, otherwise diminished by exponential averaging in plain Adam.
- **Epoch:** Epoch is defined as the number forward and backward pass of all training examples.
- **Batch Size:** The number of training examples in one forward/backward pass.

(vii) Limitations of CNN: CNNs only capture local 'spatial' patterns in data. If the data can't be made to look like an image, ConvNets are less useful. Moreover, CNNs are very bad at encoding different representations of pose and orientation within themselves. In CNNs, the orientations and their surroundings are not taken into account.

4.7 Summary

This chapter includes a brief explanation of all the topics which are related to our thesis work along with their working procedure, their advantages and disadvantages etc. Firstly, we have tried to describe the topics of basic image processing and then we moved on to the description of the basic CNN architecture along with its different hyper-parameters.

Chapter 5

Proposed Methodology

5.1 Overview

There are two proposed model by which we can detect the abnormal cells in brain MRI. We have tried to detect the tumor using traditional machine learning algorithms and also using a convolutional neural network. In classification using traditional machine learning step, we have tried to train the proposed model using six machine learning algorithms: KNN, Logistic Regression, Multi-layer perceptron, Naive Bayes, Random Forest, and SVM. At first, we will discuss the proposed segmentation technique to distinguish the tumor. After that, we will discuss the detection of the tumor using traditional machine learning algorithm. After that, we will introduce the proposed model and thoroughly describe the all the layers related to detect the brain tumor using CNN.

5.2 Our Working Approach for Brain Tumor Segmentation

We have conducted our work for brain tumor segmentation and detection in two stages. The stages are:

- Stage-1: Brain Tumor Detection using Traditional Classifiers
- Stage-2: Brain Tumor Detection using Deep Learning

5.2.1 Stage-1: Brain Tumor Detection using Traditional Classifiers

In our first prospective model, brain tumor segmentation and detection using machine learning algorithm have been done, and a comparison of the classifiers for our model is illustrated.

Our proposed system of Brain tumor detection using traditional machine learning algorithm consists of seven stages:

- Step-1: Skull stripping
- Step-2: Filtering and enhancement
- Step-3: Segmentation by Fuzzy C-means Algorithm
- Step-4: Morphological operations
- Step-5: Tumor extraction & contouring
- Step-6: Classification by traditional classifiers

The results of our work accomplished satisfactory results. The main stages of our proposed model (Fig-5.1) will be illustrated in the following sections.

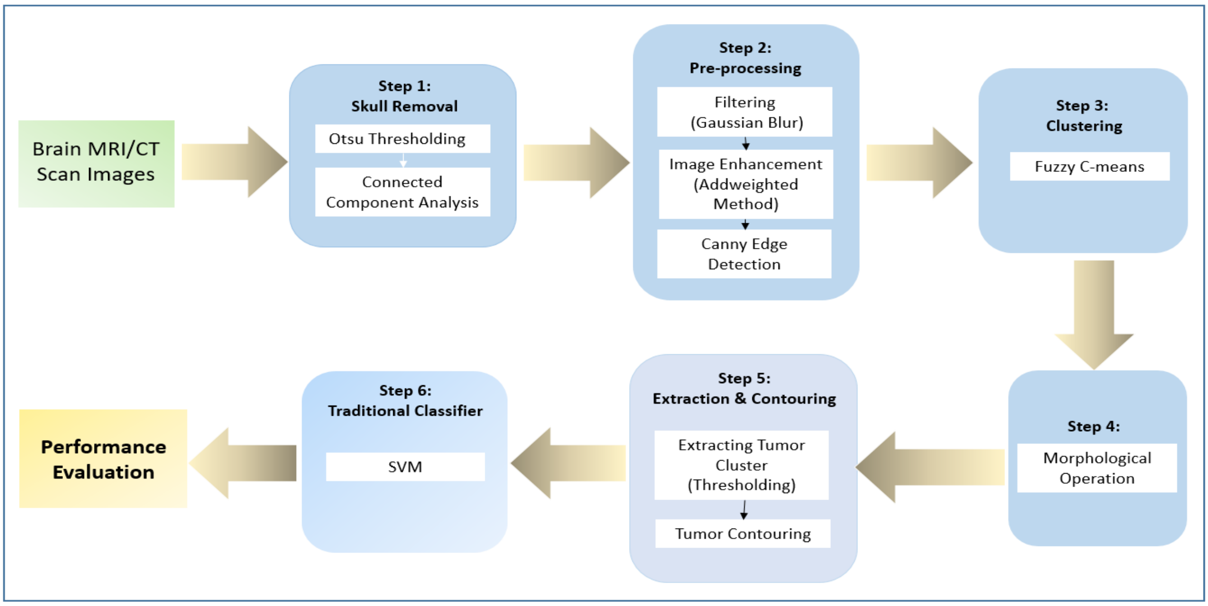


Figure 5.1: Proposed methodology for classification using Traditional Classifiers

We will comprehensively describe the proposed model of tumor segmentation and for classification using traditional Machine learning classifier, we used six classifiers and evaluated the performance.
In the later section, we will comprehensively describe the sections of the proposed methodology (fig-5.1)

5.2.1.1 Step-1: Skull Stripping

Skull stripping is a very important step in medical image processing as the background of the MRI image not containing any useful information, and it only increases the processing

time.
In our work, we come up with a hybrid model to remove the skull portion from the MRI images in three steps. These three steps can be elaborately explained in fig. 5.2.

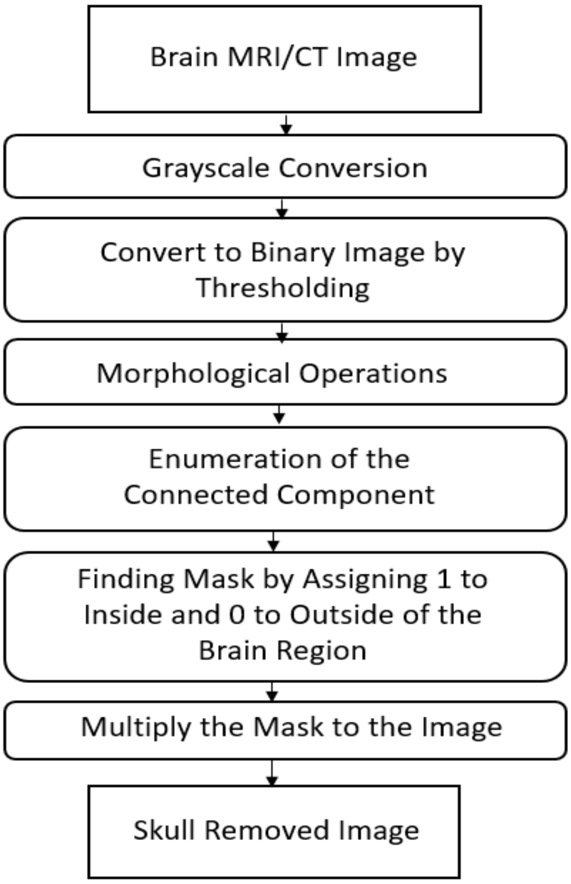


Figure 5.2: Skull stripping technique for brain tumor segmentation

- For skull removal, at first we used Otsu's Thresholding method which automatically calculates the threshold value and segments the image into background and foreground.
- After binarization of the MRI, erosion operation had been performed before applying connected component analysis.
- At the last stage of our skull stripping step, we used connected component analysis to extract only the brain region and as a consequence, we removed the skull from the MRI.
- We found the largest component which is the skull and then we found the mask by assigning 1 to inside and 0 to outside of the brain region.
- Multiplying the mask to T1, T2 and FLAIR images we got the skull removed MRI.

5.2.1.2 Step-2: Filtering and Enhancement

For better segmentation, we need to maximize the MRI image quality with minimized noise as brain MRI images are more sensitive to noise than any other medical image.

Gaussian blur filter was used in our work for Gaussian noise reduction existing in Brain MRI which prevailed the performance of the segmentation. Because we just need the abnormal tissues of the Brain MRI, we can avoid the subtle detail. In a MRI, there may be very small area of abnormal region or tumor tissues, but we need the maximal one. Also in MRI image the amount of Gaussian noise can be comparatively higher than any other noises. That is why Gaussian filter was used. Then we enhanced the image by using add-Weighted method (it provides a blending effect to the images). First, we blur the image where we know that by using smoothing filter to an image, we can suppress most of the high-frequency components. Then, we subtract this smoothed image from the original image (the resulting difference is known as a mask). Thus, the output image will have most of the high-frequency components that are blocked by the smoothing filter. Adding this mask back to the original will enhance the high-frequency components.

5.2.1.3 Step-3: Segmentation by Fuzzy C-means (FCM) Algorithm

Fuzzy C-Means clustering algorithm is used for segmentation, which allows one piece of data to belong to two or more clusters. We got the fuzzy clustered segmented image at this stage, which ensured a better segmentation. Unlike k-means where data point must exclusively belong to one cluster center here data point is assigned membership to each cluster center as a result of which data point may belong to more than one cluster center.

We also employed the Normalized cut algorithm, Thresholding technique, K-Means Clustering, and watershed segmentation. But considering the pros and cons described in background studies, we adopt FCM algorithm for segmentation.

5.2.1.4 Step-4: Morphological Operations

To segment the tumor, we only need the brain part rather than the skull part. For this, we applied morphological operations in our images.

At first, erosion was done to separate weakly connected regions of the MRI image. After erosion, we will get multiple disconnected regions in our images. Dilation is applied afterward. In case of noise removal, erosion is followed by dilation using the same structuring element which is called opening operation.

5.2.1.5 Step-5: Tumor extraction & contouring

We will find the edge of the abnormal tissues by which we can mark the perimeter. The maximum length between edges to another edge will be the perimeter of the Tumor tissues. Tumor cluster extraction was done by an intensity-based approach which is thresholding. The output of this image is the highlighted tumor area with a dark background.

After tumor Contouring, we can find the segmented tumor of the brain MRI. Further, we will classify about the tumor and find some imperative features of the tumor which are co-related with the brain MRI.

5.2.1.6 Step-6: Classification by traditional classifiers

We used six traditional machine learning classifiers which are K-Nearest Neighbor, Logistic Regression, Multi-layer Perceptron, Naive Bayes, Random Forest, and Support Vector Machine to get the accuracy of tumor detection of our proposed model. Among these six classifiers, SVM gave the best result. In this section, we will describe why SVM gives the most improved result.

A main advantage of the KNN algorithm is that it performs well with multi-modal classes because the basis of its decision is based on a small neighborhood of similar objects [100]. Therefore, even if the target class is multi-modal, the algorithm can still lead to good accuracy. However, a major disadvantage of the KNN algorithm is that it uses all the features equally in computing for similarities which can lead to classification errors, especially when there is only a small subset of features that are useful for classification [100].

Logistic regression is a better traditional learning algorithm because the output of logistic regression is more informative than other classification algorithms. But Logistic regression tends to underperform when there are multiple or non-linear decision boundaries and for MRI images it will not work well. They are not flexible enough to naturally capture more complex relationships. Also this classifier needs long training time sometimes.

Naive Bayes classifier does not work well with the brain MRI because there is a great chance of losing the accuracy. In brain MRI, dependencies between the abnormal and normal tissue play considerable importance.

Random forests algorithm has less variance than a single decision tree. It means that it works correctly for a large range of data items than single decision trees. But the BRATS dataset is not a large one to employ the random forest algorithm on it. The main disadvantage of

Random forests is their complexity. They are much harder and time-consuming to construct than decision trees. It also requires more computational resources and is also less intuitive.

Support Vector Machine (SVM) gives more improve result in terms of classifying an image. SVM works well with even unstructured and semi-structured data like text, images, and trees. With an appropriate kernel function, we can detect the tumor. We consider all the pixels of the MRI as a feature. Unlike in neural networks, SVM is not solved for local optima. It scales relatively well to high dimensional data.

We can say that for classifying tumor MRI, the outcomes of classification of an image using SVM gives better result with respect to other classifiers . That is why, We used Support Vector Machine algorithm for tumor detection in our proposed model. We got the best result using SVM over the other traditional classifiers.

5.2.2 Classification Using Convolutional Neural Network

Convolutional Neural Network is broadly used in the field of Medical image processing. Over the years lots of researchers tried to build a model which can detect the tumor more efficiently. It is a class of deep neural networks which is applied to interpreting visual imagery. A fully-connected neural network can detect the tumor, but because of parameter sharing and sparsity of connection, we adopted the Convolutional Neural Network (CNN) for our model.

A Five-Layer Convolutional Neural Network is introduced and implemented for tumor detection. The aggregated model consisting of seven stages including the hidden layers provides us with the better performing result for the apprehension of the tumor. One convolutional layer, One Pooling layer, One Flatten layer, and Two Fully connected layer work as the hidden layers along with the input and output layer.

We characterized the complete working flow into seven steps from which we can detect the brain tumor using CNN. In fig-5.3 the seven-step working flow diagram is shown.

Fig-5.4 is the proposed methodology for tumor detection using 5-Layer Convolutional Neural Network. In fig-5.4, five hidden layers are shown along with its dimension. We will discuss about the layer and its characteristics in the following section.

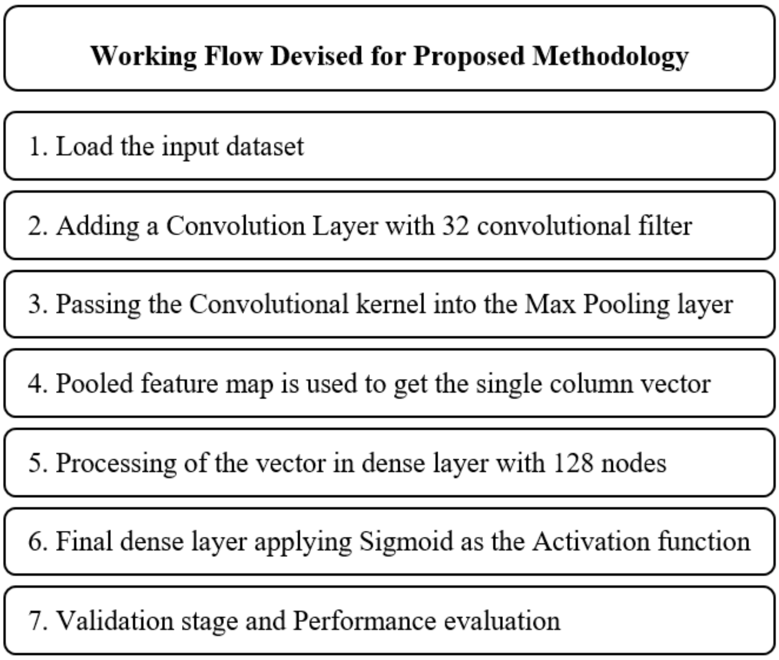


Figure 5.3: Working flow of the proposed five layer CNN Model.

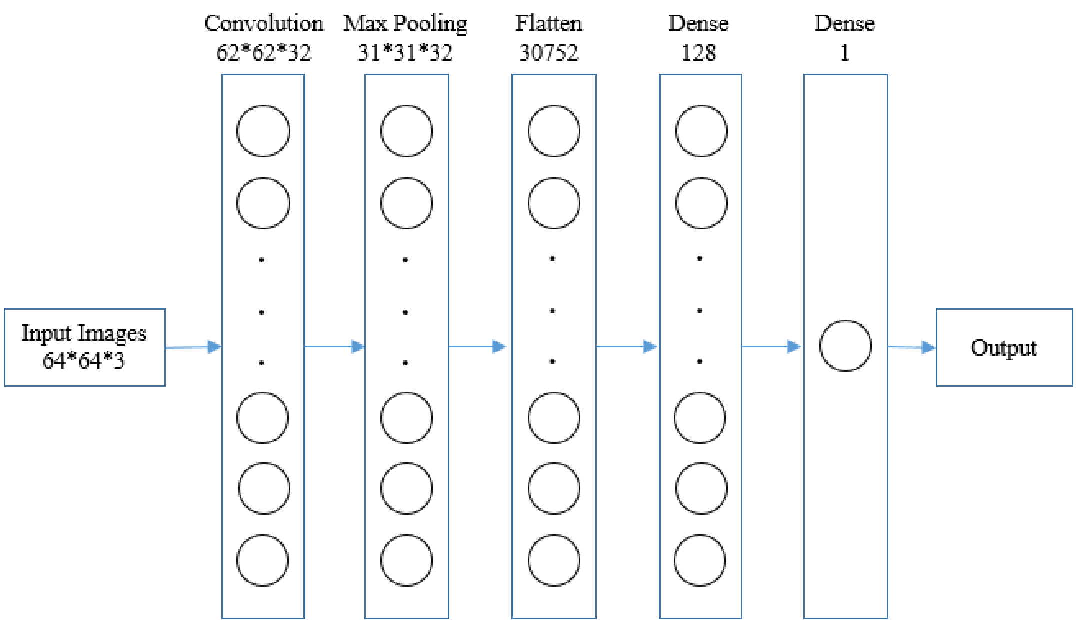


Figure 5.4: Proposed Methodology for Brain tumor detection using 5-Layer Convolutional Neural Network

5.2.2.1 Convolutional Layer

A Convolutional layer is the core building block of a CNN model. Using convolutional layer as the opening layer, an input shape of the MRI images is generated which is 64*64*3, converting all the images into a same dimension. After accumulating all the images in the same aspect, we created a convolutional kernel that is convoluted with the input layer - administering with 32 convolutional filters of size 3*3 each with the support of 3 channels tensors. Rectified Linear Unit (ReLU) is used as an activation function.

The input volume has size 64*64*3 which means 64 pixels width, 64 pixels height, depth of 3 and the filter size is 3*3. Then each neuron in the Convolutional Layer will have weights to a 3*3*3 region in the input volume, for a total of 3*3*3 = 27 weights and one will be added for bias parameter.

There are three hyper parameters which we will evaluate and those are- depth, stride and zero-padding. For our model, the input volume size is 64*64*3, filter size is 3*3 so the spatial extent or filter size is 3. We did not use padding in the border so that Zero Padding, and stride is 1. Using stride as 1 grant us to do all the spatial down-sampling in the pooling layers, with the CONV layers only transforming the input volume depth-wise. Now using the convolutional layer equation - (4.9) we find the dimension of convolutional layer which is 62, 62 and 32 respectively. After Convolutional layer, we introduced the max pooling layer as the second layer.

5.2.2.2 Max Pooling Layer

The main focus of pooling layer is to progressively reduce the spatial size of the representation in order to reduce the number of parameters and computational task in the network. It can control over-fitting because it can scale down the parameters. Using the max pooling layer, it can operate independently on every depth slice of the input and resizes it spatially.

Working on the Brain MRI image can also cost the contamination of the over-fitting and for this Max Pooling layer perfectly works for this perception. So, for our input image, we used MaxPooling2D for the proposed model. This convolutional layer runs on 31*31*32 dimension. Input size is 62*62*32 and after finding out the dimension of max pooling layer using the equation - (4.10), we got the output volume size of 31*31*32. Because of dividing the input images in both spatial dimensions, the pool size is (2, 2) which means a tuple of two integers by which to downscale by vertically and horizontally.

5.2.2.3 Flatten Layer

After the pooling layer, a pooled feature map is obtained. Flatten layer is one of the essential layers after the pooling because we have to transform the whole matrix representing the input images into a single column vector and its imperative for processing. It is then fed to the Neural Network for the processing. The dimension of this layer is 31*31*32 = 30752.

5.2.2.4 Fully Connected Layer

Two fully connected layers were employed- Dense-1 and Dense-2 represented the dense layer. The dense function is applied in Keras for the processing of the Neural Network, and the obtained vector works as an input for this layer.

There are 128 nodes in the hidden layer. Because the number of dimension or nodes proportional with the computing resources we need to fit our model we kept it as moderate as possible and for this perspective 128 nodes gives the most substantial result. ReLU is used as the activation function because of showing better convergence performance.

After the first dense layer, the second fully connected layer was used as the final layer of the model. In this layer, we used the sigmoid function as activation function where the total number of the node is one because we need to lower the uses of computing resources so that a more significant amount assuages the execution time. There is a chance of hampering the learning in deep networks for using of the sigmoid as the activation function. So we scale down the sigmoid function, and the number of the nodes is much lesser and easy to handle for this deep network.

In a nutshell, we have a five-layer CNN model by which we can detect a tumor from an MRI image. The entire working flow of the five-layer CNN model is illustrated in figure-3. First, we have to load the input dataset and all the images should be identical in size in the input image. After the input layer, we introduced a Convolution layer with 32 convolutional filters along ReLU as an activation function. Max pooling layer is employer after the ConvNet architecture. Pooled feature map is obtained from Max pooling layer which is used to get the single column vector and the mechanism is known as Flattening. Finally, we develop two fully connected layers which are named as a dense layer in our model with 128 nodes. The sigmoid function is used as the activation function in the final dense layer. Using Adam optimizer and binary cross-entropy as a loss function, we compiled the model and find the accuracy of detecting the tumor.

The performance of the CNN model is done by finding the accuracy of our proposed model.

We have developed an algorithm (Algorithm-1) to find the performance to detect the brain tumor. The following algorithm was used to evaluate the performance of our CNN model:

Algorithm 1: Accuracy of the proposed CNN model

```
loadImage();
dataAugmentation();
splitData();
laodData();
for each epoch in epochNumber do
    for each batch in batchSize do
        ŷ = model(features);
        loss = crossEntropy(y, ŷ);
        optimization(loss);
        Accuracy = (1 - loss) * 100%;
    end
end
```

5.3 Summary

In this chapter, the proposed methodology of segmentation and detection of brain tumor is described. Segmentation of the abnormal tissues along with the detection using two distinct process is thoroughly demonstrated with proper diagram and explanation.

Chapter 6

Experimental Results & Evaluation

6.1 Overview

In this section, we will comprehensively describe the outcomes of our proposed methodology. We carried out the tumor segmentation using Fuzzy C Means (FCM) algorithm and classify our model as tumor or non-tumor using two distinct mechanisms which are detection using traditional machine learning and detection using a convolutional neural network. Following we will do the performance evaluation process and compare the performance of the two models. Furthermore, we also analyze our model with the existing model in terms of segmentation and classification.

6.2 Experimental Setup

We used the Jupyter Notebook and various python packages such as Numpy, Pandas, OpenCV etc for image precessing. For the traditional classifiers, we used the Scikit-Learn. We used python version 3.6 with Anaconda. For training and testing our model through CNN, we used Tensorflow and keras framework. We used the dedicated GPU which is provided by Google Colab.

6.3 Dataset Acquisition

We used the benchmark dataset in the field of Brain Tumor Segmentation, and that is BRATS dataset [16]. BRATS is a Multimodal Brain Tumor Segmentation dataset named after MICCAI (The Medical Image Computing and Computer Assisted Intervention Society). The dataset is a labeled dataset differentiating tumor and non-tumor. The dataset breaks down into two segment- Training set and Testing set. All the images are from three categories- T1-weighted, T1-weighted and FLAIR (Fluid Attenuated Inversion Recovery). Both sets consist of two classes. One is for tumorous MRI (class-1) and the other one is for Non-tumorous MRI (class-0). In the training set, there are 187 tumor MRI images and 30 non-tumor MRI images. And there are 24 testing images for performance evaluation.

Furthermore, we also used a dataset developed by Jun Cheng [102] where a total of 3064 T1-weighted contrast-enhanced images from 233 patients with three kinds of brain tumor: meningioma (708 slices), glioma (1426 slices), and pituitary tumor (930 slices). It is only used for the evaluation of the CNN model.

6.4 Performance Measures

We have to discuss about the performance matrices in order to know in what extent our model works accurately. In this section we will discuss about the performance measure that we consider to evaluate our model. We need to familiarize with some terms regarding the performance measures.

6.4.1 Confusion Matrix

The Confusion matrix is one of the most intuitive metrics used to find the correctness and accuracy of the model. It is used for Classification problem where the output can be of two or more types of classes. The Confusion Matrix consists of four parameters which are described below-

- **True Positive (TP):** Number of tumor images that are correctly classified.
- **True Negative (TN):** Number of non-tumor images that are correctly classified.
- **False Positive (FP):** Number of non-tumor images that are misclassified as tumor.
- **False Negative (FN):** Number of tumor images that are misclassified as non-tumor.

6.4.2 Accuracy

It's the most popular performance matrix which measures how often the classifier produce the correct prediction. Mathematically Accuracy defined as the ratio of the number of correct predicted images and the total number of images and symbolically represented as-

$$Accuracy = \frac{Correct\ \ Predictions}{Total\ \ number\ \ of\ \ images} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6.1}$$

6.4.3 Precision

It is the retrieved information that are relevant to the model. Precision is the ratio of the number of tumor images that are correctly classified (TP) and the number of images classified or misclassified as tumor (TP + FP). The lower the FP the higher the Precision. The model is more effective in case of higher precision rate.

$$Precision = \frac{TP}{TP + FP} \tag{6.2}$$

6.4.4 Recall

It is the fraction of the images which are successfully retrieved. Recall is the ratio of the number of tumor images that are correctly classified and the number of images that are to be predicted. Sensitivity, Hit Rate, True Positive Rate are the other names of Recall. The lower the False negative the higher the recall because the number of tumor images that are classified as non-tumor is low.

$$Recall = \frac{TP}{TP + FN} \tag{6.3}$$

6.4.5 F-Score

It is the harmonic mean of Precision and Recall and is a measure of test accuracy. F-score reaches its best value at 1 (100% precision and recall) and worst value at 0. F-Score can be defined as-

$$F - Score = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \tag{6.4}$$

6.4.6 Specificity

Specificity is the True Negative Rate (TNR) of the model and the statistical measure of the binary classification test. As we are dealing with a binary classification (tumor or non-tumor) so we can use this as the performance evaluation. It is the ratio of the number of non-tumor images that are correctly classified (TN) and the number of images that are classified or misclassified as non-tumor (TN + FP). The lower the false positive (FP) the higher the specificity or selectivity.

$$Specificity = TNR = \frac{TN}{TN + FP} \tag{6.5}$$

6.5 Hyper-parameters setting for CNN Model

In this section, the hyper-parameter values in initialization and training stage that were used to implement our CNN model are mentioned. These information are depicted in the chart below:

Table 6.1: Hyper-parameter value of the proposed CNN Model

Stage	**Hyper-parameter**	**Value**
Initialization	Bias	Zeros
	Weights	glorot_uniform
Training	Learning rate	0.001
	Beta_1	0.9
	Beta_2	0.999
	Epsilon	None
	Decay	0.0
	Amsgrad	False
	Epoch	10
	Batch_size	32
	steps_per_epoch	80

6.6 Experimental Results

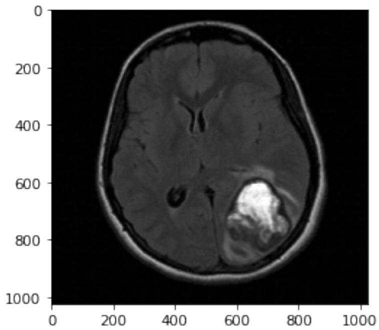
In this section, we will discuss the experimental result of the proposed methodology. We divide this section into three subsection which we will describe thoroughly.

6.6.1 Segmentation of the Brain Tumor using FCM

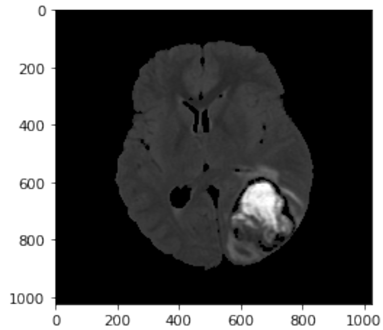
In this section, we will illustrate all the output images after each step of segmentation based on the proposed methodology.

6.6.1.1 Skull Stripping

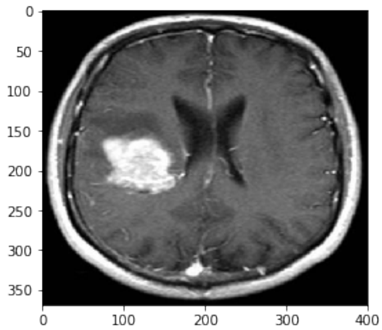
In Brain MRI image, Skull plays a less important role when it comes to segment the tumor. So we can say the skull is unnecessary or redundant part of the MRI. In six steps we remove the skull from the MRI. These six steps are: Gray-scale conversion, thresholding, Morphological operations, enumeration of the connected component, finding the mask and lastly multiply the mask to the MRI image. Fig-6.1 is the depiction of the skull removed image from input Brain MRI image.



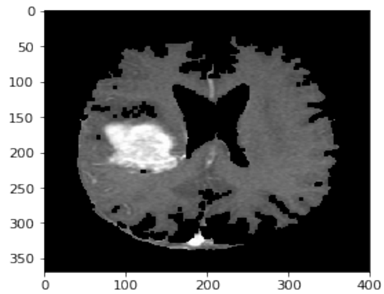
(a) Input Image - 1



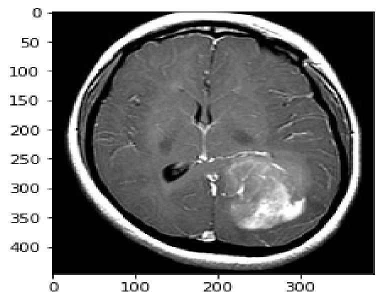
(b) Skull Removed Image - 1



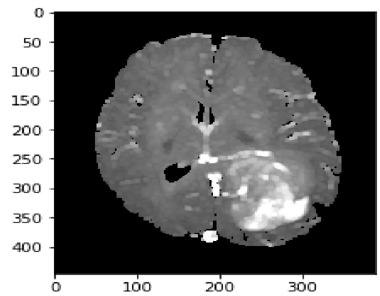
(c) Input Image - 2



(d) Skull Removed Image - 2



(e) Input Image - 3

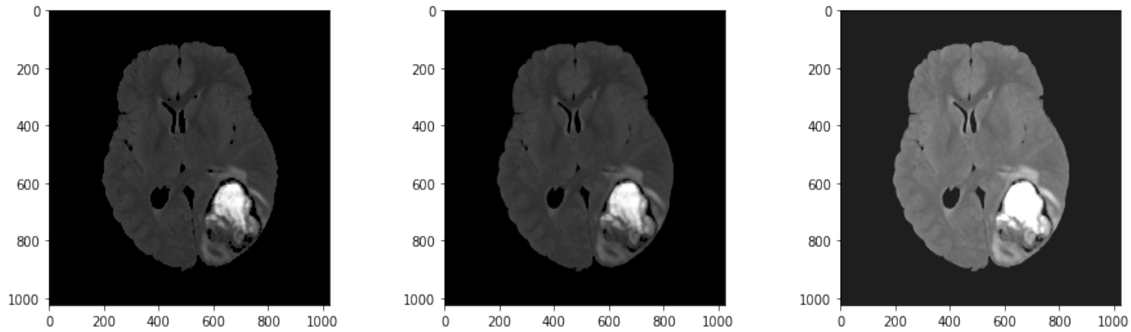


(f) Skull Removed Image - 3

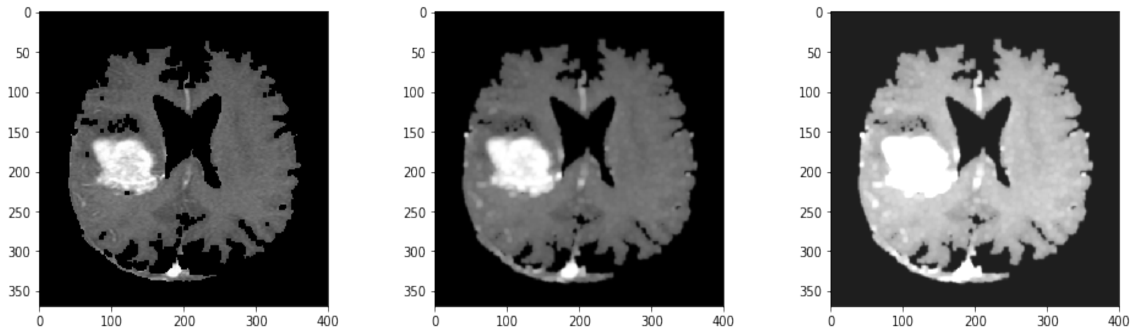
Figure 6.1: Skull Removal of the Brain MRI Image

6.6.1.2 Filtering & Enhancement

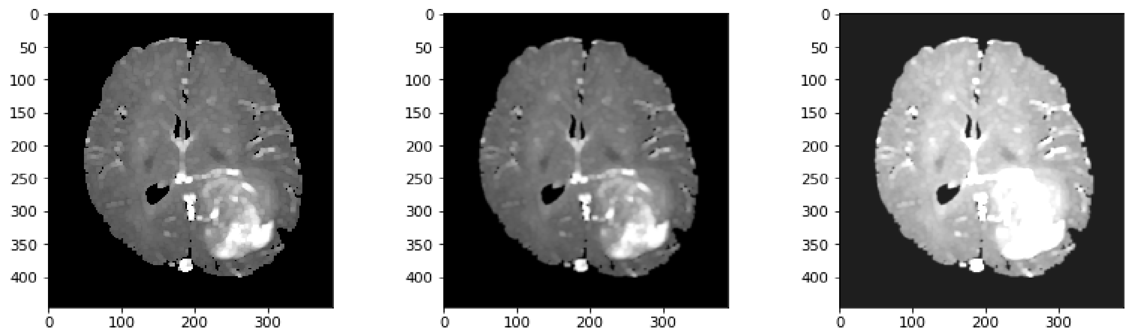
In this section, we enhanced the MRI image. Using the Gaussian blur filter, we remove the noise and enhancement is done using the Addweighted method. After that, we used our segmentation algorithm. We have shown the steps of this step in Figure-6.2 where the output of step-1 (skull removed MRI) is going as input in this section and output is the enhanced MRI. We also have shown the filtered image which is done using Gaussian blur filter, in between the skull removed MRI and Enhanced MRI.



(a) Skull Removed Image - 1 (b) Filtered Image - 1 (c) Enhanced Image - 1



(d) Skull Removed Image - 2 (e) Filtered Image - 2 (f) Enhanced Image - 2

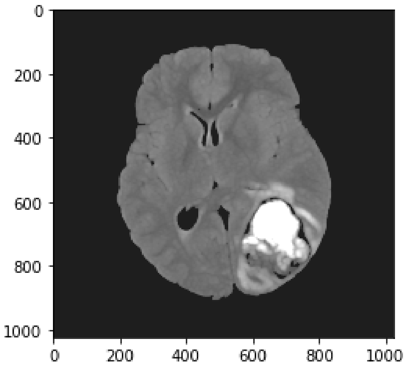


(g) Skull Removed Image - 3 (h) Filtered Image - 3 (i) Enhanced Image - 3

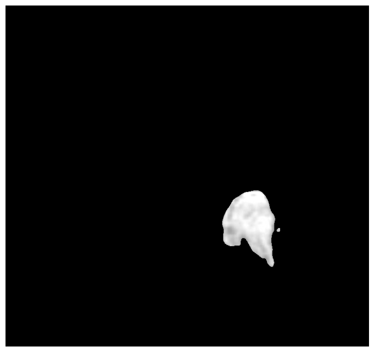
Figure 6.2: Filtering & Enhancement of the skull removed image

6.6.1.3 Segmentation using FCM

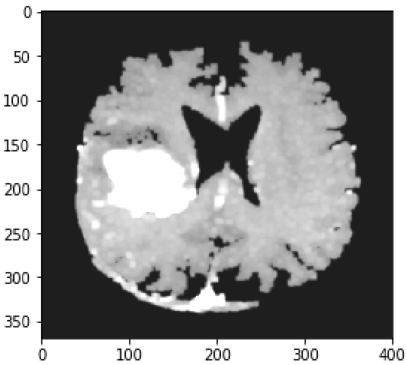
After Image Enhancement, we segment the abnormal tissues which are the brain tumor from the MRI. Using Fuzzy C Means segmentation algorithm we distinguish between the normal cells and abnormal cells. We used FCM, K-Means clustering, Normalized cut, Watershed segmentation, and thresholding method. But FCM gives us the improved result as it ensures that one piece of data to belong to two or more clusters which are important for Brain MRI. In Figure-6.3 the segmented brain tumor is shown.



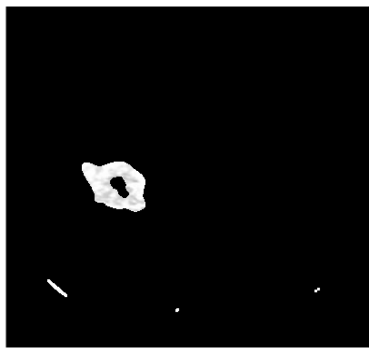
(a) Enhanced Image - 1



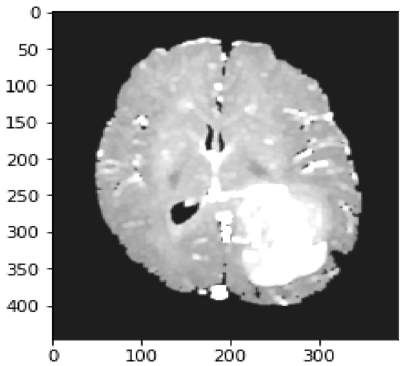
(b) Segmented Image - 1



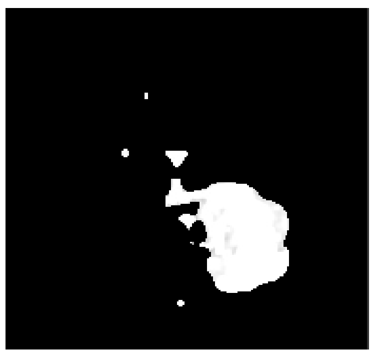
(c) Enhanced Image - 2



(d) Segmented Image - 2



(e) Enhanced Image - 3

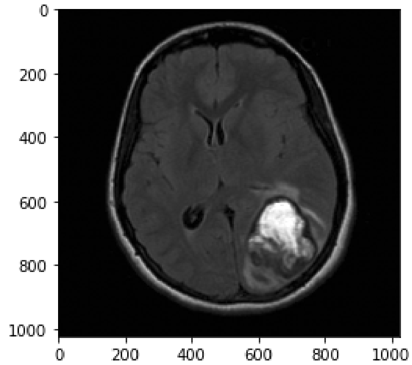


(f) Segmented Image - 3

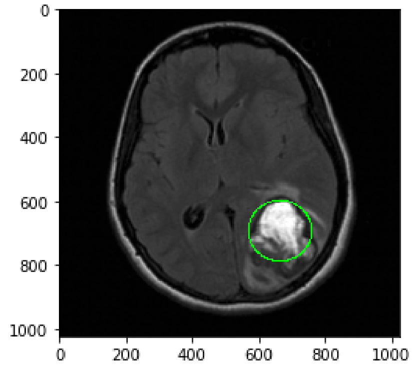
Figure 6.3: Segmentation using FCM of the enhance brain MRI image

6.6.1.4 Tumor Contouring

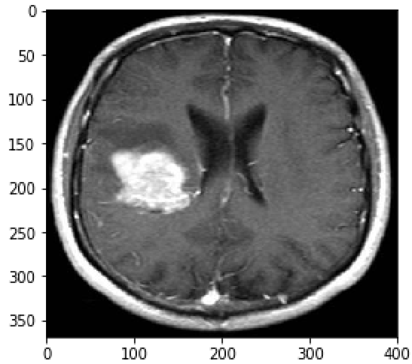
Finally, the last step is tumor contouring. In this step, we showed the contoured brain tumor MRI image. It is circled in green color indicates the location of the brain tumor. Tumor cluster extraction is done by an intensity-based approach named as thresholding. Figure-6.4 illustrated the contoured brain tumor MRI images.



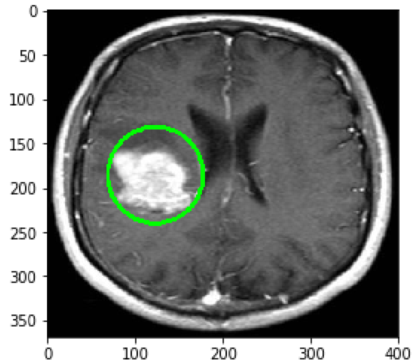
(a) Input Image - 1



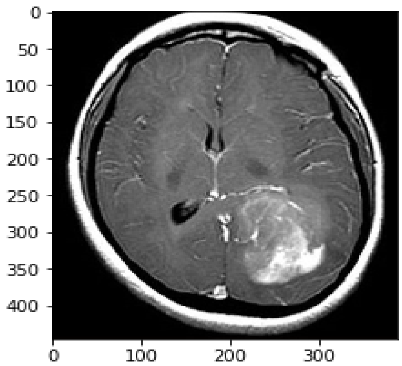
(b) Contoured Image Image - 1



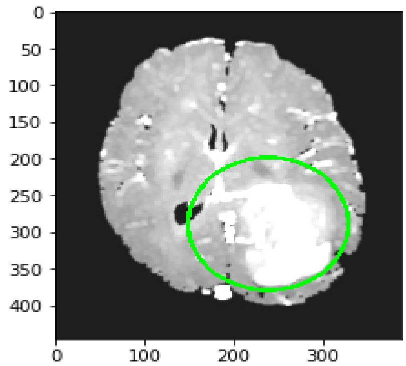
(c) Input Image - 2



(d) Contoured Image - 2



(e) Input Image - 3



(f) Contoured Image Image - 3

Figure 6.4: Tumor Contouring of the input MRI image

6.6.2 Classification using Traditional Machine Learning Algorithm

We employed Six traditional Machine Learning Algorithms to classify our model. The Six traditional Machine Learning Algorithms are K-Nearest Neighbor, Logistic Regression, Multi-layer Perceptron, Naive Bayes, Random Forest, and Support Vector Machine.

Based on splitting ratio we have shown the performance of the proposed model into two experiments which are described in the following section.

6.6.2.1 Experiment - I

In the first experiment, we split the dataset into **70 by 30** ratio which means 70% of the data goes for training and 30% of the data goes for testing. In Table - 6.2, we represented the performance of our proposed model based on this ratio. We can see that, SVM gives the best result in terms of Accuracy, Recall, Dice score and Jaccard Index.

Table 6.2: Performance Matrices using ML Classifiers (based on 70:30 splitting)

Classifiers	Accuracy	Recall	Specificity	Precision	Dice Score	Jaccard Index
K-Nearest Neighbor	0.8939	0.949	0.4280	0.9330	0.9410	0.889
Logistic Regression	0.8788	0.949	0.286	0.918	0.933	0.875
Multi-layer Perception	0.8939	1.000	0.167	0.894	0.944	0.894
Naive Bayes	0.7879	0.797	**0.714**	**0.959**	0.870	0.770
Random Forest	0.8939	**0.983**	0.167	0.903	0.943	0.892
Support Vector Machine	**0.9242**	**0.983**	0.428	0.935	**0.959**	**0.921**

6.6.2.2 Experiment - II

In the second experiment, we split the dataset into **80 by 20** ratio which means 80% of the data goes for training and 20% of the data goes for testing. In Table - 6.3, we represented the performance of our proposed model based on this ratio. Accuracy of 88.63% is achieved using SVM as the classifier.

Table 6.3: Performance Matrices using ML Classifiers (based on 80:20 splitting)

Classifiers	Accuracy	Recall	Specificity	Precision	Dice Score	Jaccard Index
K-Nearest Neighbor	0.8409	0.40	0.8810	0.48	**0.465**	0.412
Logistic Regression	**0.8863**	**0.50**	0.9050	0.20	0.285	0.167
Multi-layer Perception	**0.8863**	0.41	0.8864	0.48	0.442	**0.465**

Table 6.3 continued from previous page

Classifiers	Accuracy	Recall	Specificity	Precision	Dice Score	Jaccard Index
Naive Bayes	0.7727	0.22	**0.9143**	**0.40**	0.285	0.167
Random Forest	**0.8863**	**0.50**	0.905	0.20	0.285	0.167
Support Vector Machine	**0.8863**	**0.50**	0.905	0.20	0.285	0.167

6.6.3 Classification using Convolutional Neural Network

We trained our model based on two datasets. We have operated a different number of layers with a different value of hyper-parameters and we got the best result using the five-layer CNN model with respected splitting ratio and other criteria. Based on the number of layers, we break down the performance of the proposed CNN model into several sections. We will show the experimental results along with other characteristics and evaluation in this following sub-section. In Experiment-I and Experiment-II, we have shown the accuracy while tuning the learning rate and epoch using the five-layer model based on 70:30 and 80:20 splitting ratio respectively. Later, we have shown the complete outcomes of using five, six and seven layers CNN model and compare the results.

6.6.3.1 Experiment-I

In this section, the proposed five-layer CNN model is trained by splitting the dataset into **70 by 30 ratio**. Table - 6.4 represents the relation between learning rate, epochs, time to train and accuracy of the model based on this splitting ratio. When the learning rate is **0.001**, epoch is **50**, training time is **500 sec** then we get the best accuracy: **96.55%** in this splitting ratio.

Table 6.4: Training Time and Accuracy of the proposed CNN model (splitting ratio 70:30)

Learning Rate	Epochs	Time to train(sec)	Accuracy (%)
0.001	10	180	92.88
	20	231	93.23
	50	**500**	**96.55**
	100	1227	96.01
0.005	10	198	93.00
	20	**240**	**93.13**
	50	555	90.68
	100	1133	91.22

Table 6.4 continued from previous page

Learning Rate	Epochs	Time to train(sec)	Accuracy (%)
0.01	10	191	88.76
	20	235	90.92
	50	634	91.25
	100	**1347**	**93.45**

6.6.3.2 Experiment - II

In this section, the proposed five-layer CNN model is trained by splitting the dataset into **80 by 20 ratio**. In Table-6.5, a comparison of training time and accuracy is shown which is based on learning rate and epochs. When the learning rate is **0.001**, epoch is **10** then we get the best accuracy: **96.55%** and the training time is **175 sec** in this splitting ratio.

Table 6.5: Training Time and Accuracy of the proposed CNN model (splitting ratio 80:20)

Learning Rate	Epochs	Time to train(sec)	Accuracy(%)
0.001	**10**	**175**	**97.87**
	20	**233**	**97.87**
	50	527	95.74
	100	1200	95.69
0.005	10	177	96.03
	20	**203**	**97.62**
	50	488	95.55
	100	1027	95.55
0.01	10	178	92.09
	20	200	93.04
	50	599	**93.77**
	100	966	92.00

6.6.3.3 Experiment-III (Five-layer Architecture)

Furthermore, for the justification of the proposed five-layer CNN model, we have shown several experimental results based on different hyper-parameters and splitting ratio. In table-6.6, we showed the performance of five-layer CNN model based on 80:20 and 70:30 splitting ratio. From the table-6.6 we can see that when the splitting ratio is **80:20**, the batch size is **64** and epoch is **10** the model provides the best result of **97.87%** as accuracy. After this point, the model overfits the data. So, for five layer CNN model, we get the accuracy of **97.87%**.

Table 6.6: Performance of the proposed five-layer CNN model

Convolution Layer	Max Pooling	Split Ratio	Batch Size	Epoch	Accuracy (%)
62*62*32	31*31*32	80:20	32	8	**92.72**
				9	85.82
				10	86.85
				11	87.88
			64	8	93.67
				9	94.98
				10	**97.87**
				11	94.89
		70:30	32	8	81.35
				9	83.71
				10	87.87
				11	**89.13**
			64	8	88.07
				9	88.76
				10	91.23
				11	**94.90**

6.6.3.4 Experiment-IV (Six-layer Architecture)

In this section, we add one more convolutional layer of size **62*62*32**. We calculated the size of the convolutional layer using equation-4.8. Changing the dimension could cause the model with poor accuracy. Using two convolutional layers, one max pooling layer, one flatten layer and two fully connected layers, we get maximum accuracy of **94.21%**. In table-6.6, we got the maximum accuracy when the splitting ratio is 80:20, the batch size is 64 and epoch is 11. After this point, the model accuracy started decreasing gradually. So adding more convolutional layer will not increase the accuracy.

Table 6.7: Performance of Six-layer CNN model

Convolution Layer	Max Pooling	Split Ratio	Batch Size	Epoch	Accuracy (%)
62*62*32 62*62*32 (2 layer)	31*31*32	80:20	32	8	89.29
				9	92.83
				10	**93.76**
				11	93.62
			64	8	94.01
				9	94.08
				10	**94.39**

Table 6.7 continued from previous page

Convolution Layer	Max Pooling	Split Ratio	Batch Size	Epoch	Accuracy (%)
				11	94.21
		70:30	32	8	82.37
				9	83.71
				10	**84.24**
				11	86.17
			64	8	**88.27**
				9	82.23
				10	81.69
				11	80.07

6.6.3.5 Experiment-V (Seven Layer Architecture)

In the fifth phase of our experiment, we employed seven-layer (two convolutional layers, two max pooling layers, one flatten layer and two fully connected layers) of CNN model to measure the performance for the detection of brain tumor. As we seen in the earlier experiment that, adding more convolutional layer does not improve the performance of the model, we increase the Max Pooling layer. From table-6.8, we get the maximum accuracy of **95.67%** when the splitting ratio is **80:20**, batch size is **64** and epoch is **9**.

Table 6.8: Performance of Seven-layer CNN model

Convolution Layer	Max Pooling	Split Ratio	Batch Size	Epoch	Accuracy (%)
62*62*32 31*31*128 (3 layer)	31*31*32 15*15*128 (2 layer)	80:20	32	8	88.17
				9	**90.16**
				10	89.02
			64	8	95.33
				9	**95.67**
				10	95.21
				11	94.99
		70:30	32	8	79.73
				9	84.17
				10	**84.38**
				11	84.31
			64	8	87.07
				9	87.12
				10	**87.19**
				11	87.18

6.6.3.6 Model Validation

Two graphical representation is depicted below in the two figures based on the splitting ratio 70:30 of the proposed CNN model using the BRATS dataset. In fig-6.5 and fig-6.6, the splitting ratio is 70:30. In fig-6.6 which is the loss curve where the training and validation curve intersected in the earlier cycle of the epoch. So, the validation rate is not proportional with the model accuracy from the start which is not good for the model.

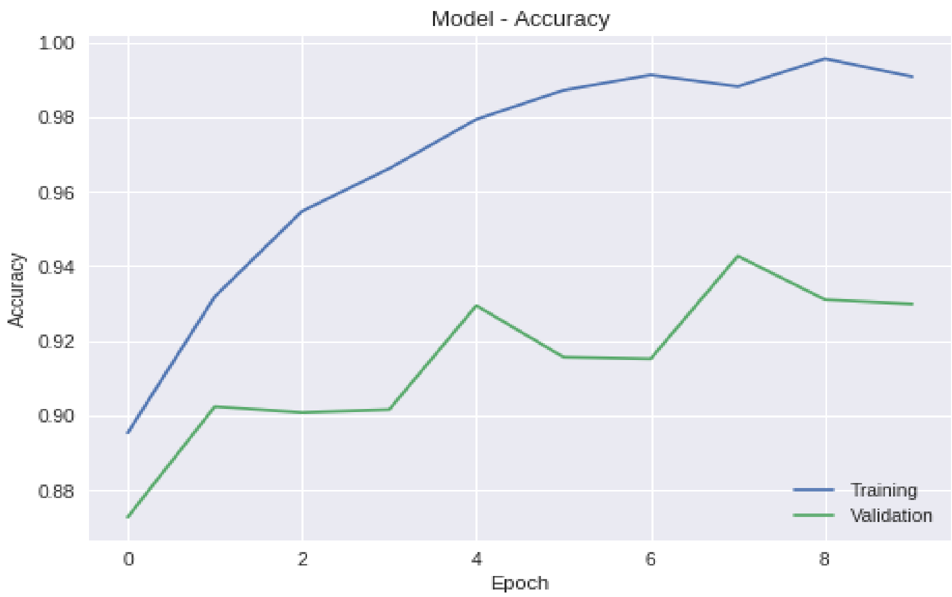


Figure 6.5: Model Accuracy curve based on 70:30 split ratio (BRATS Dataset)

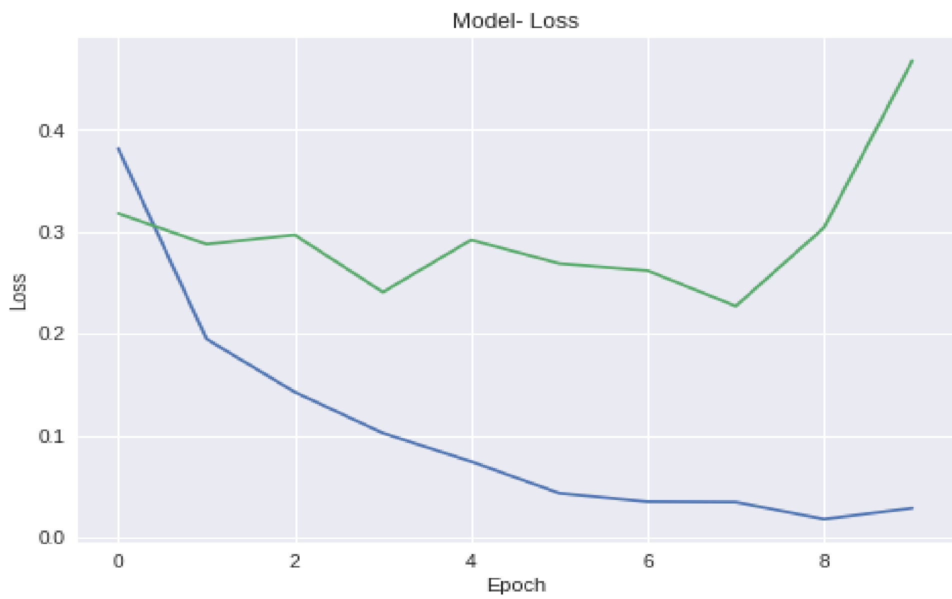


Figure 6.6: Model Loss Curve based on 70:30 split ratio (BRATS Dataset)

In fig-6.7, a representation of accuracy curve of the model is depicted. We can see from the loss curve (fig-6.8) that at epoch six the training and validation curve met. Hence we can say that for splitting ratio 80:20 the model gives the best result.

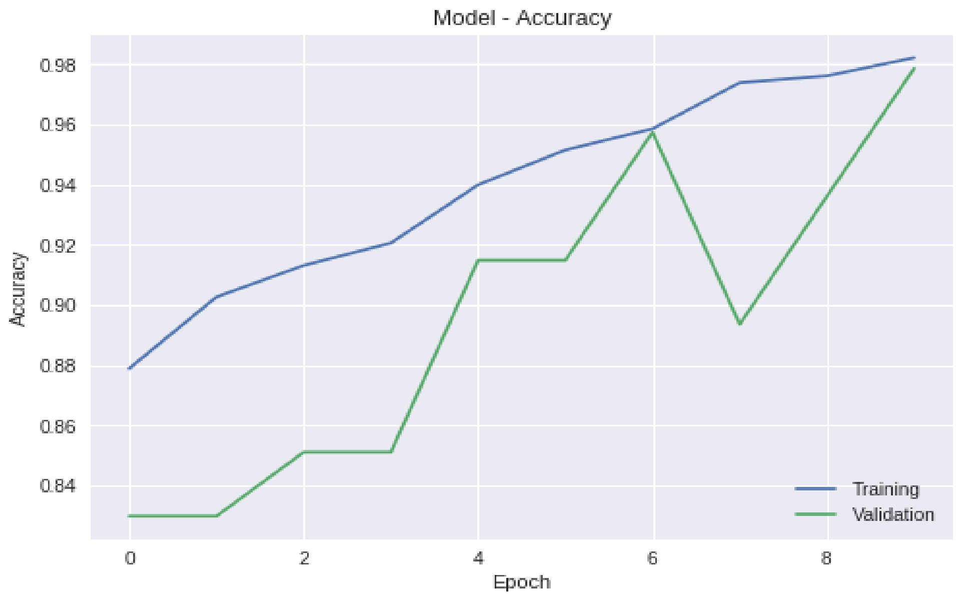


Figure 6.7: Model Accuracy Curve based on 80:20 split ratio (BRATS Dataset)

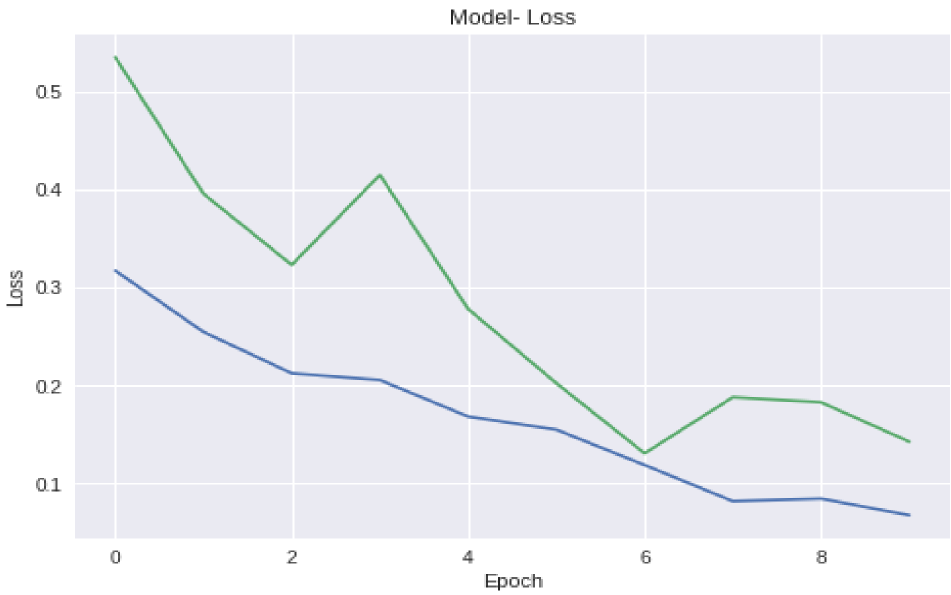


Figure 6.8: Model Loss Curve based on 80:20 split ratio (BRATS Dataset)

Fig-6.9 is the depiction of the of the Accuracy of the six algorithms based on 70:30 distribution ratio. Here, SVM gives the best result.

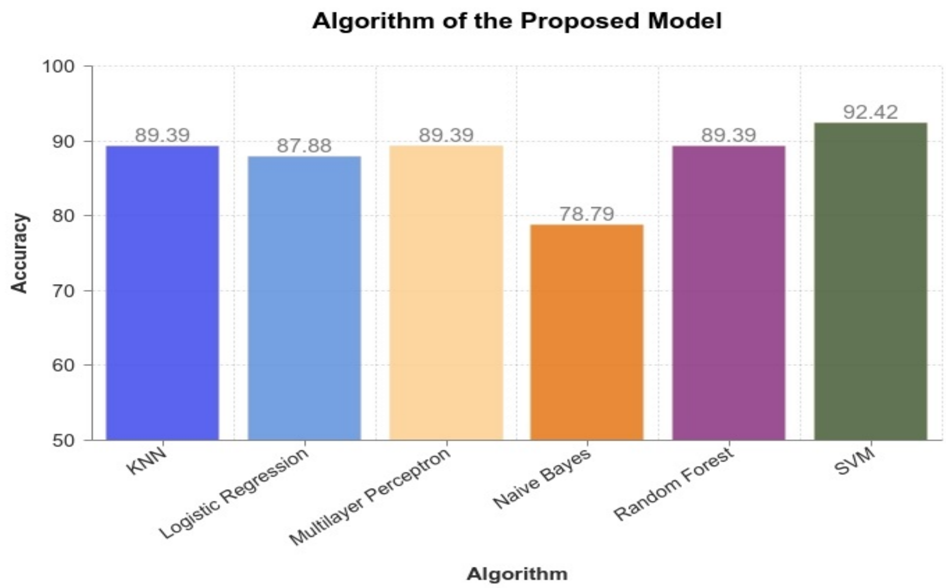


Figure 6.9: Accuracy of the Proposed Model using ML classifiers based on 70:30 split ratio

In fig-6.10, a representation of the of the Accuracy of the six algorithms based on 80:20 distribution ratio. Here, SVM gives the most improved result.

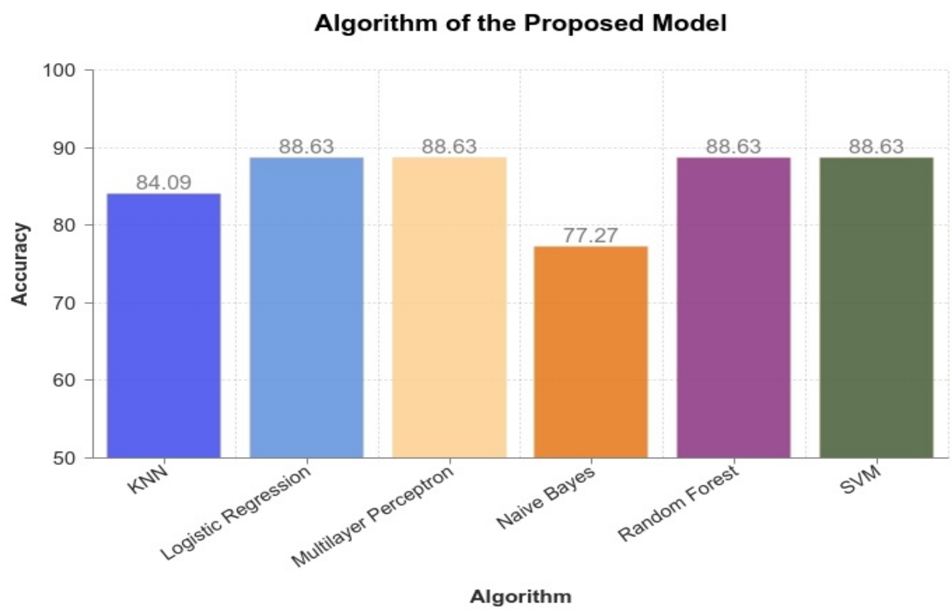


Figure 6.10: Accuracy of the Proposed Model using ML classifiers based on 80:20 split ratio

Fig 6.11 illustrated a bar chart of learning rate vs training time. We selected three values of learning rate such as 0.001, 0.005 and 0.01 based on this splitting ratio. The training time is increasing with the learning rate and at **0.001** learning rate and **10** epochs we get the minimal amount of training time.

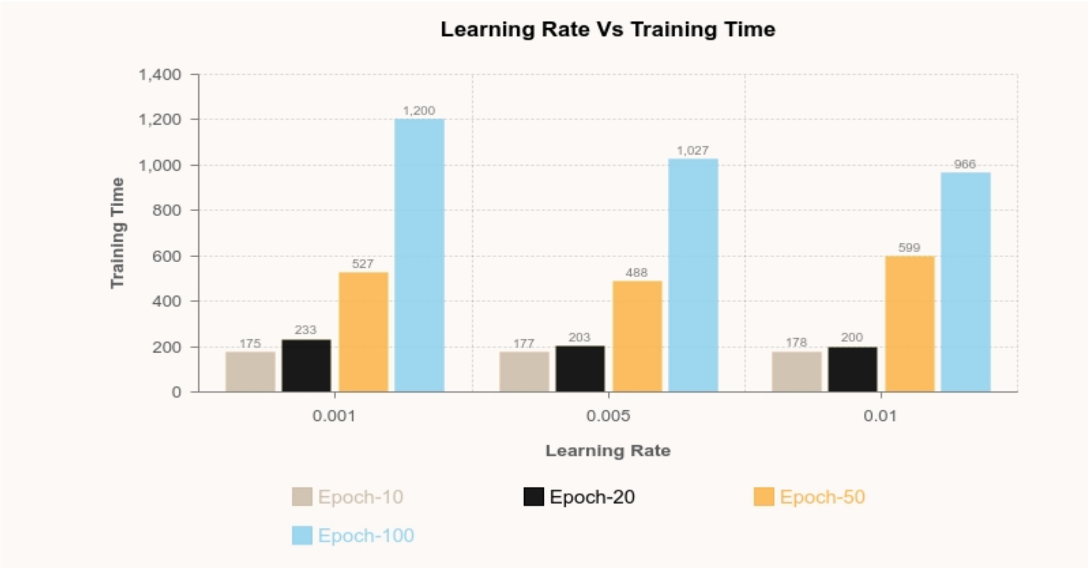


Figure 6.11: Learning Rate Vs Training Time of proposed CNN model (based on 80:20 splitting)

A bar chart of learning rate vs Accuracy is illustrated in fig. 6.12. We selected four values of epoch such as 10, 20, 50 and 100 based on this splitting ratio. The accuracy of the model is decreasing as the learning rate increases gradually. We get the best accuracy of **97.87%** at **0.001** learning rate.

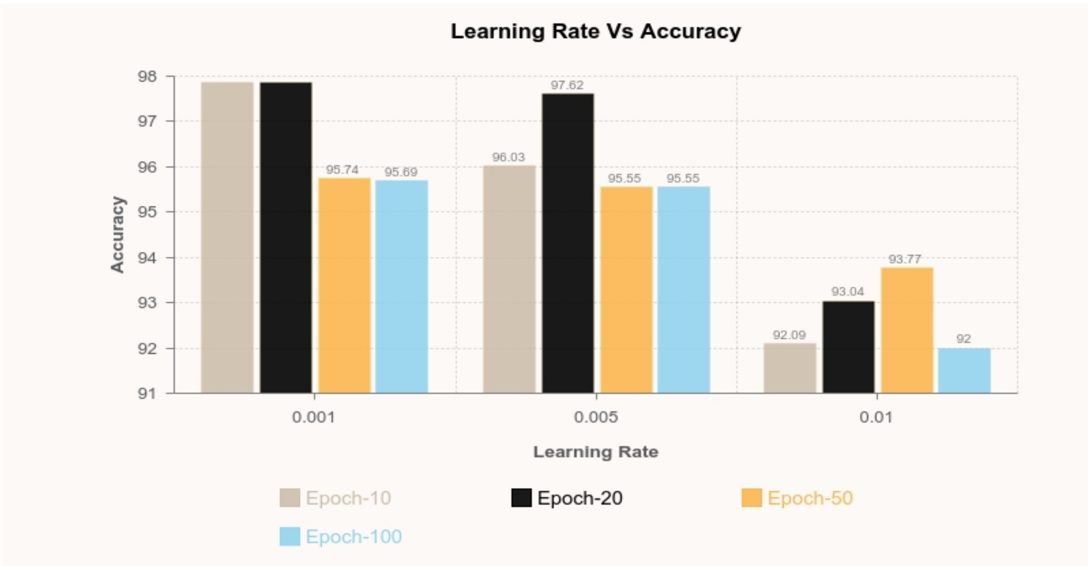


Figure 6.12: Learning Rate Vs Accuracy of proposed CNN model (based on 80:20 split ratio)

Figure 6.13 illustrated a bar chart of learning rate vs training time. We selected three values of learning rate such as 0.001, 0.005 and 0.01 based on this splitting ratio. The training time is increasing with the learning rate and at **0.001** learning rate and **10** epochs we get the minimal amount of training time: **180 seconds**.

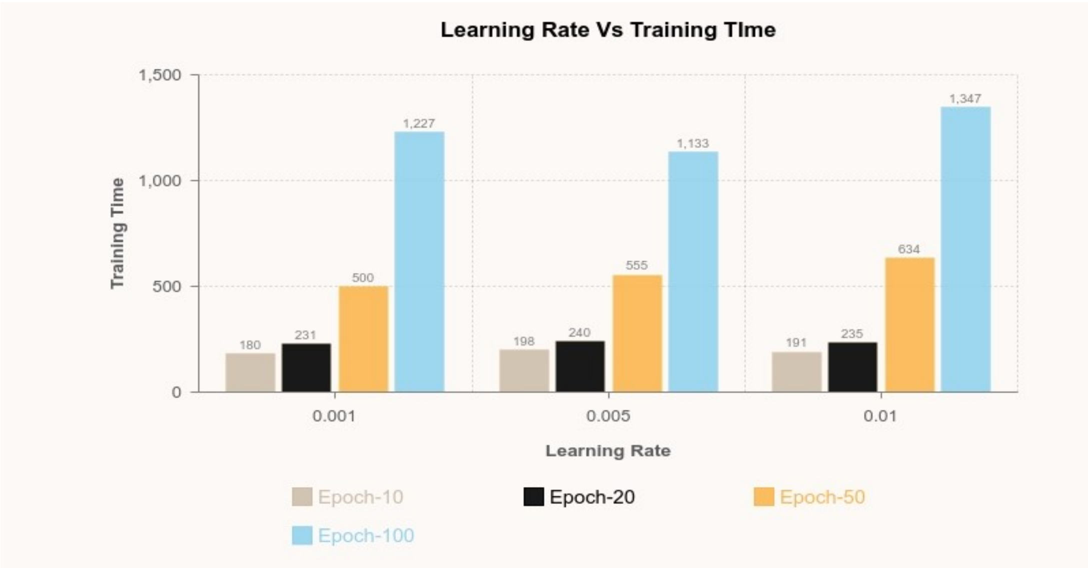


Figure 6.13: Learning Rate Vs Training Time of proposed CNN model (based on 70:30 split ratio)

In fig. 6.14, a bar chart of learning rate vs Accuracy is shown. We selected four values of epoch such as 10, 20, 50 and 100 based on this splitting ratio. We get the best accuracy of **92.88%** at **0.001** learning rate.

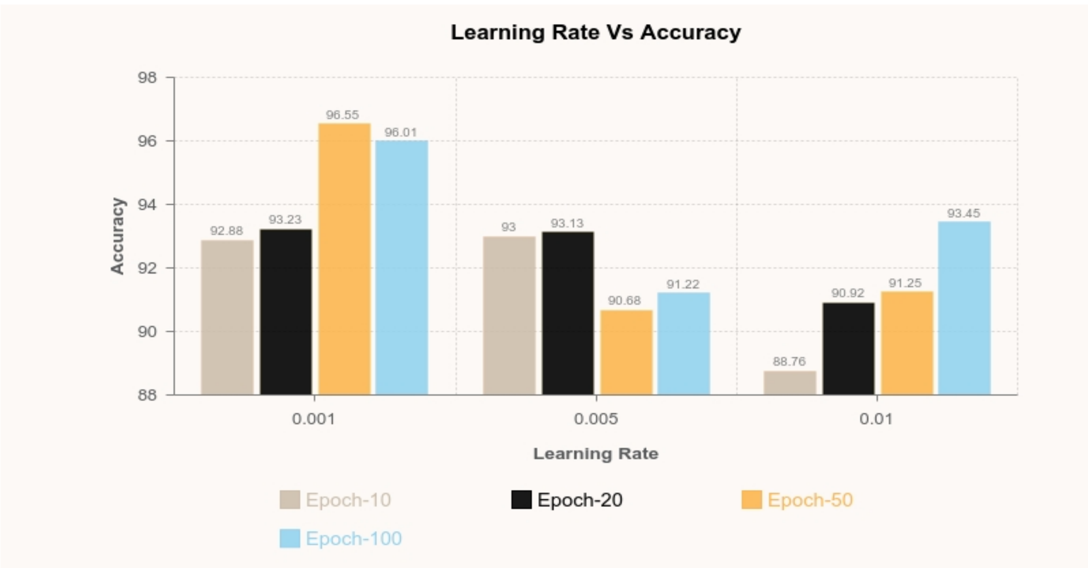


Figure 6.14: Learning Rate Vs Accuracy of proposed CNN model (based on 70:30 split ratio)

6.7 Performance Comparison

In this section, we compared the performance of the existing models to the proposed models. Though we used two dataset to measure the performance of the proposed model, we will take into consideration of the results of the BRATS dataset.

6.7.1 Comparison Between Traditional Machine Learning and CNN

We compared the proposed model of classification using Traditional Machine Learning algorithm and the CNN model depending on different splitting ratio. In traditional machine learning model, we evaluate the model based on two splitting techniques. Among them 70:30 splitting ratio gives us the best result of **92.42%** for traditional machine learning.

For the proposed CNN model, three types of CNN model is used and we got **97.87%** as the best accuracy for CNN models for five-layer CNN model. Table-6.9 showed the comparison of the CNN model based on the experiment - III, IV, V.

Table 6.9: Comparison of CNN models

No. of Layers	Convolution Layer	Max Pooling	Split Ratio	Batch Size	Epoch	Accuracy (%)
Five	**62*62*32**	**31*31*32**	**80:20**	**64**	**10**	**97.87**
Six	62*62*32 62*62*32	31*31*32	80:20	64	10	94.39
Seven	62*62*32 31*31*128	31*31*32 15*15*128	80:20	64	9	95.67

Table-6.10 illustrated the comparison of these two models. Comparing the two best results from proposed traditional machine learning and CNN model, CNN outperforms ML in terms of Accuracy. From this observation we can assert that the five-layer CNN model provides us the most better result among the other distribution and the learning rate is **0.001**, epoch **10** and the training time is **175 secs**.

Table 6.10: Performance Comparison Between the proposed ML and CNN model

Model	Splitting Ratio	Accuracy (%)
Traditional Machine Learning	70:30	92.42
	80:20	88.63
CNN	70:30	96.55
	80:20	**97.87**

6.8 Performance comparison Between the existing model the proposed CNN model

In this section, we will do a comparison among the existing model and our proposed methodology of five layer CNN model. Table-6.11 illustrated the comparison between the existing models and our models. We took seven related research articles based on the same dataset from different methods and compare their outcomes with the proposed CNN model.

Table 6.11: Performance comparison with the existing models and the proposed CNN model

No	Paper Name	Year	Method	Accuracy
1	Brain tumor segmentation based on a new threshold approach[11]	2017	Pixel Subtraction + Thresholding	96%
2	Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM [71]	2017	Berkeley Wavelet Transform + SVM	96.51%
3	Tumor Diagnosis in MRI Brain Image using ACM Segmentation and ANN-LM Classification Techniques [65]	2016	FCM + Artificial Neural Network	93.74% (Jaccard Index)
4	Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network [36]	2017	Probabilistic Neural Network	95%
5	**Proposed Model**	**2019**	**Five Layer CNN Model**	**97.87%**

6.9 Summary

We have shown the performance matrices with a short description and the value of the hyper-parameters for which we got the best result for the proposed CNN model. After that, the output of each step in the segmentation of the tumor is depicted. Later, we have shown all the experimental and empirical results with proper justification. Finally, we have done a comparison between our proposed models and also with the existing models.

Chapter 7

Conclusion & Future Works

7.1 Conclusion

Performance analysis of automated brain tumor detection from MR imaging and CT scan using basic image processing techniques based on various hard and soft computing has been performed in our work. Moreover, we applied six traditional classifiers to detect brain tumor in the images. Then we applied CNN for brain tumor detection to include deep learning method in our work. We compared the result of the traditional one having the best accuracy (SVM) with the result of CNN. Furthermore, our work presents a generic method of tumor detection and extraction of its various features.

In the context of the full dataset, it is necessary to parallelize and utilize high-performance computing platform for maximum efficiency. We tried our best to detect the tumors accurately but, nevertheless we faced some problems in our work where tumor could not be detected or falsely detected. So, we will try to work on those images and on the complete dataset. Hence, we will try to apply other deep learning methods in the future so that we can get a more precise and better result.

7.2 Limitations

There are some limitations of our thesis work that we have listed in this section which we are leaving to improve in our future works.

- The BRATS dataset has only 241 images
- Worked only on 2D images.
- We could have tried more traditional classifiers to increase the accuracy.

- Types of the tumor could not be classified

7.3 Future Works

There are more opportunities for improvement or research on our work in the future.

- Firstly, the number of images can be increased. The bigger the number of the images is, the better the model is trained.
- Secondly, we want to work on 3D images in future.
- Thirdly, more traditional classifiers can be applied to get more increased accuracy.
- Fourthly, we will try to classify the tumor if its benign or malignant after the detection of the tumor.
- Last but not the least, more variations of deep learning methods can be tested in future.

References

[1] McKinney PA Brain tumours: incidence, survival, and aetiology Journal of Neurology, Neurosurgery & Psychiatry 2004;75:ii12-ii17.

[2] "General Information About Adult Brain Tumors". NCI. Last Accessed Date: May 10, 2019.

[3] Nunn, Laura Silverstein; Silverstein, Alvin; Silverstein, Virginia B. (2006). "Cancer. Brookfield, Conn: Twenty-First Century Books". pp. 11-12. ISBN 0-7613-2833-5.

[4] David Lowell Strayer; Raphael Rubin; Rubin, Emanuel (2008). "Rubin's pathology: clinicopathologic foundations of medicine". Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins. pp. 138-139. ISBN 0-7817-9516-8.

[5] https://www.researchgate.net/figure/Example-of-type-III-tumor-involvement-A-49-year-old-man-with-lung-cancer-A-A-selective-fig3-38042515, Last Accessed Date: May 10, 2019.

[6] https://www.cancerresearchuk.org/sites/default/files/cstream-node/surv 5yr age brain 0.pdf, Last Accessed Date: March 05, 2019.

[7] https://www.wcrf.org/int/cancer-facts-figures/worldwide-data,World Cancer Research Fund, Worldwide cancer data, Last Accessed Date: May 10, 2019.

[8] Khurram Shehzad, Imran Siddique, Obed Ullah Memon, "Efficient Brain Tumor Detection Using Image Processing Techniques", International Journal of Scientific & Engineering Research, December 2016.

[9] Mrs.A.Sankari, Drs.S.Vigneshwari, "Automatic Tumor Segmentation Using Convolutional Neural Networks", *2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM)* .

[10] Prof. Vrushali Borase., Prof. Gayatri Naik. and Prof. Vaishali Londhe, "Brain MR Image Segmentation for Tumor Detection using Artificial Neural", International Journal Of Engineering And Computer Science, 2017.

[11] Umit Ilhan, Ahmet Ilhan, "Brain tumor segmentation based on a new threshold approach," Procedia Computer Science, ISSN: 1877-0509, Vol: 120, Page: 580-587, Publica-

tion Year: 2017.

[12] Rohini Paul Joseph, C. Senthil Singh and M.Manikandan, "Brain MRI Image Segmentation and Detection in Image Processing", International Journal of Research in Engineering and Technology, 2014.

[13] Miss. Shrutika Santosh Hunnur, Prof. Akshata Raut, Prof. Swati Kulkarni, "Implementation of Image Processing for Detection of Brain Tumors", IEEE, 2017.

[14] Anam Mustaqeem, Ali Javed, Tehseen Fatima, "An Efficient Brain Tumor Detection Algorithm Using Watershed & Thresholding Based Segmentation", *I.J. Image, Graphics and Signal Processing*, 2012,10,34-39.

[15] B. Devkota, Abeer Alsadoon, P.W.C. Prasad, A. K. Singh, A. Elchouemi, "Image Segmentation for Early Stage Brain Tumor Detection using Mathematical Morphological Reconstruction," *6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8 December 2017, Kurukshetra, India.*

[16] Song, Yantao & Ji, Zexuan & Sun, Quansen & Yuhui, Zheng. (2016). "A Novel Brain Tumor Segmentation from Multi-Modality MRI via A Level-Set-Based Model". Journal of Signal Processing Systems. 87. 10.1007/s11265-016-1188-4.

[17] Ehab F. Badran, Esraa Galal Mahmoud, Nadder Hamdy, "An Algorithm for Detecting Brain Tumors in MRI Images", *7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, 2017.*

[18] Pei L, Reza SMS, Li W, Davatzikos C, Iftekharuddin KM. "Improved brain tumor segmentation by utilizing tumor growth model in longitudinal brain MRI". Proc SPIE Int Soc Opt Eng. 2017.

[19] Dina Aboul Dahab, Samy S. A. Ghoniemy, Gamal M. Selim, "Automated Brain Tumor Detection and Identification using Image Processing and Probabilistic Neural Network Techniques", IJIPVC, Vol. 1, No. 2, pp. 1-8, 2012.

[20] Mohd Fauzi Othman, Mohd Ariffanan and Mohd Basri, "Probabilistic Neural Network for Brain Tumor Classification", *2nd International Conference on Intelligent Systems, Modelling and Simulation, 2011.*

[21] A. Rajendran, R. Dhanasekaran, "Fuzzy Clustering and Deformable Model for Tumor Segmentation on MRI Brain Image: A Combined Approach," *International Conference on Communication Technology and System Design 2011.*

[22] Sobhaninia, Zahra & Rezaei, Safiyeh & Noroozi, Alireza & Ahmadi, Mehdi & Zarrabi, Hamidreza & Karimi, Nader & Emami, Ali & Samavi, Shadrokh. (2018). "Brain Tumor Segmentation Using Deep Learning by Type Specific Sorting of Images".

[23] Kaatsch P (June 2010). "Epidemiology of childhood cancer". Cancer Treatment Reviews. 36 (4): 277-85.

[24] Maheswari, D. and Radha V. (2010) "Enhanced Layer Based Compound Image, Compression Proceedings of the First Amrita ACM-W celebration of Women in Computing (A2CWIC 2010)", Pp. 209- 216,ISBN:978-1-4503-0194-7.

[25] Yi Yang, Sam Hallman, Deva Ramanan, Charless C. Fowlkes. (2009-2010). "Layered object Models for Image Segmentation".

[26] Wankai Deng , Wei Xiao, He Deng, Jianguo Liu, "MRI Brain Tumor Segmentation With Region Growing Method Based On The Gradients And Variances Along And Inside Of The Boundary Curve", *3rd International Conference on Biomedical Engineering and Informatics, 2010.*

[27] Anam Mustaqeem, Ali Javed, Tehseen Fatima, "An Efficient Brain Tumor Detection Algorithm Using Watershed & Thresholding Based Segmentation", *I.J. Image, Graphics and Signal Processing*, 2012, 10, 34-39.

[28] Jianping Fan, D. K. Y. Yau, A. K. Elmagarmid and W. G. Aref, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," in IEEE Transactions on Image Processing, vol. 10, no. 10, pp. 1454-1466, Oct. 2001.

[29] MacKay, David (2003). "Chapter 20. An Example Inference Task: Clustering" (PDF). Information Theory, Inference and Learning Algorithms. Cambridge University Press. pp. 284&ndash, 292. ISBN 0-521-64298-1. MR 2012999.

[30] S. R. Telrandhe, A. Pimpalkar and A. Kendhe, "Detection of brain tumor from MRI images by using segmentation & SVM," 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), Coimbatore, 2016, pp. 1-6.

[31] J. C. Dunn (1973): "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics 3: 32-57

[32] J. C. Bezdek (1981): "Pattern Recognition with Fuzzy Objective Function Algoritms", Plenum Press, New York

[33] Wankai Deng , Wei Xiao, He Deng, Jianguo Liu, "MRI Brain Tumor Segmentation With Region Growing Method Based On The Gradients And Variances Along And Inside Of The Boundary Curve", *3rd International Conference on Biomedical Engineering and Informatics , 2010.*

[34] S. K. Adhikari, J. K. Sing, D. K. Basu and M. Nasipuri, "A spatial fuzzy C-means algorithm with application to MRI image segmentation," *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR) , Kolkata, 2015, pp. 1-6.*

[35] M. U. Akram and A. Usman, "Computer aided system for brain tumor detection and segmentation," *International Conference on Computer Networks and Information Technology , Abbottabad, 2011, pp. 299-302.*

[36] Shree, N. Varuna and T. N. R. Kumar. "Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network." Brain Informatics (2017).

[37] Debnath Bhattacharyya and Tai-hoon Kim, "Brain Tumor Detection Using MRI Image Analysis," UCMA 2011, Part II, CCIS 151, pp. 307-314, 2011., Springer-Verlag Berlin Heidelberg 2011.

[38] Malathi, R. and Dr. Nadirabanu Kamal, A.R., "Evaluating the Hausdorff Distance for Contour Segmentation of Brain Images," Australian Journal of Basic and Applied Sciences, 8(16) October 2014, Pages: 100-111.

[39] R. Karthik, R. Menaka, C. Chellamuthu, "A comprehensive framework for classification of brain tumor images using SVM and Curvelet transform", International Journal of Biomedical Engineering and Technology, Vol. 17, No. 2, pp. 168-177, 2015.

[40] Gonzalez, Rafael C., and Richard E. Woods. Digital Image Processing. Addison-Wesley, 1993.

[41] M. Sezgin & B. Sankur (2004). "Survey over image thresholding techniques and quantitative performance evaluation". Journal of Electronic Imaging. 13 (1): 146-165

[42] Kirti Mittal, Abhishek Shekhar, Prashant Singh, Manish Kumar, "Brain Tumour Extraction using Otsu Based Threshold Segmentation", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 4, April 2017.

[43] K. Rajesh Babu, P.V. Naganjaneyulu and K. Satya Prasad, "Performance Analysis for Efficient Brain Tumor Segmentation by using Clustering Algorithm," Indian Journal of Science and Technology, Vol 10(11), March 2017.

[44] Ming-Ni Wu, Chia-Chen Lin, Chin-Chen Chang, "Brain Tumor Detection Using Color-Based K-Means Clustering Segmentation", *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing , 2007.*

[45] J.selvakumar, A.Lakshmi, T.Arivoli, "Brain Tumor Segmentation and Its Area Calculation in Brain MR Images using K-Mean Clustering and Fuzzy C-Mean Algorithm", IEEE-International Conference On *Advances In Engineering, Science And Management (ICAESM -2012)* March 30, 31, 2012.

[46] Jianwei Liu, Lei Guo, "A New Brain MRI Image Segmentation Strategy Based on K-means Clustering and SVM", *7th International Conference on Intelligent Human-Machine Systems and Cybernetics , 2015.*

[47] J. Vijay and J. Subhashini, "An efficient brain tumor detection methodology using K-means clustering algorithm," *2013 International Conference on Communication and Signal Processing, Melmaruvathur, 2013, pp. 653-657.*

[48] C. H. Rao, P. V. Naganjaneyulu and K. S. Prasad, "Brain Tumor Detection and Segmentation Using Conditional Random Field," *2017 IEEE 7th International Advance Computing Conference (IACC)* , Hyderabad, 2017, pp. 807-810.

[49] Parveen, Amritpal singh, "Detection of Brain Tumor in MRI Images, using Combination of Fuzzy C-Means and SVM", *2nd International Conference on Signal Processing and Integrated Networks (SPIN) , 2015.*

[50] T.Logeswari and M.Karnan "An Improved Implementation of Brain Tumor Detection Using Segmentation Based on Hierarchical Self Organization Map" International Journal of Computer Theory and Engineering, Vol. 2, No. 4, August, 2010 1793-201.

[51] N. N. Gopal and M. Karnan, "Diagnose brain tumor through MRI using image processing clustering algorithms such as Fuzzy C Means along with intelligent optimization techniques," *2010 IEEE International Conference on Computational Intelligence and Computing Research , Coimbatore, 2010, pp. 1-4.*

[52] N. Gordillo, E. Montseny and P. Sobrevilla, "A new fuzzy approach to brain tumor segmentation," *International Conference on Fuzzy Systems, Barcelona, 2010, pp. 1-8.*

[53] Chaiyanan Sompong, Sartra Wongthanavasu, "Brain Tumor Segmentation Using Cellular Automata-based Fuzzy C-Means", *13th International Joint Conference on Computer Science and Software Engineering (JCSSE) , 2013.*

[54] Lawrene O. Hall, Amine M. Bensaid, Laurene P. Clarke, Robert P. Velthuizen, Martin S. Silbiger, and James C. Bezdek, "A Comparison of Neural Network and Fuzzy Clustering Techniques in Segmenting Magnetic Resonance Images of the Brain" IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL 3, NO 5, SEPTEMBER 1992.

[55] Bing Nan Li, Chee Kong Chui, Stephen Chang, S.H. Ong, "Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation" Computers in Biology and Medicine 41(2011) 1-10.

[56] Pranita Balaji Kanade, Prof. P.P. Gumaste, "Brain Tumor detection Using MRI images", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering Vol. 3, Issue 2, February 2015.

[57] Eman Abdel-Maksoud, Mohammed Elmogy, Rashid Al-Awadi, "Brain tumor segmentation based on a hybrid clustering technique", Elsevier. Volume 16, Issue 1, March 2015.

[58] K. S. Angel Viji and J. Jayakumari, "Automatic detection of brain tumor based on magnetic resonance image using CAD System with watershed segmentation," *2011 International*

Conference on Signal Processing, Communication, Computing and Networking Technologies , Thuckafay, 2011, pp. 145-150.

[59] Dawood Dilber, Jasleen, "Brain Tumor Detection using Watershed Algorithm", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 3, March 2016.

[60] Nooshin Nabizadeh, Miroslav Kubat, "Automatic tumor segmentation in single-spectral MRI using a texture-based and contour-based algorithm", Expert Systems with Applications, 77, pp. 1-10, 2017.

[61] Bauer S., Nolte LP., Reyes M. (2011) "Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization". In: Fichtinger G., Martel A., Peters T. (eds) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2011. MICCAI 2011.

[62] S. Pereira, A. Pinto, V. Alves and C. A. Silva, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," in IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1240-1251, May 2016.

[63] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, Hugo Larochelle, "Brain tumor segmentation with Deep Neural Networks", Medical Image Analysis, Volume 35, 2017, Pages 18-31, ISSN 1361-8415

[64] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha and A. Yuille, "Efficient Multilevel Brain Tumor Segmentation With Integrated Bayesian Model Classification," in IEEE Transactions on Medical Imaging, vol. 27, no. 5, pp. 629-640, May 2008.

[65] Shenbagarajan, A., Ramalingam, V., Balasubramanian, C., & Palanivel, S. (2016). "Tumor Diagnosis in MRI Brain Image using ACM Segmentation and ANN-LM Classification Techniques." Indian Journal Of Science And Technology, 9(1).

[66] Elisee Ilunga-Mbuyamba, Juan Gabriel Avina-Cervantes, Jonathan Cepeda-Negrete, Mario Alberto Ibarra-Manzano, Claire Chalopin, "Automatic selection of localized region-based active contour models using image content analysis applied to brain tumor segmentation" Computers in Biology and Medicine, Volume 91, 2017, Pages 69-79

[67] Jainy Sachdeva, Vinod Kumar, Indra Gupta, Niranjan Khandelwal, Chirag Kamal Ahuja, "A novel content-based active contour model for brain tumor segmentation", Magnetic Resonance Imaging, Volume 30, Issue 5, 2012, Pages 694-715, ISSN 0730-725X.

[68] Mariam Saii, Zaid Kraitem, "Automatic Brain tumor detection in MRI using image processing techniques", Biomedical Statistics and Informatics, Vol. 2, No. 2, pp. 73-76, 2017.

[69] Zexuan Ji, Yong Xia, , Quansen Sun, Qiang Chen, Deshen Xia, David Dagan Feng,

"Fuzzy Local Gaussian Mixture Model for Brain MR Image Segmentation", IEEE Transactions on Information Technology in Biomedicine, Vol. 16, No. 3, May 2012.

[70] M. Huang, W. Yang, Y. Wu, J. Jiang, W. Chen and Q. Feng, "Brain Tumor Segmentation Based on Local Independent Projection-Based Classification," in IEEE Transactions on Biomedical Engineering, vol. 61, no. 10, pp. 2633-2645, Oct. 2014.

[71] 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 Volume 2017.

[72] A. Demirhan, M. TÃűrÃij and Äř. GÃijler, "Segmentation of Tumor and Edema Along With Healthy Tissues of Brain Using Wavelets and Neural Networks," IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4, pp. 1451-1458, July 2015.

[73] Binaghi, Elisabetta & Omodei, Massimo & Pedoia, Valentina & Balbi, Sergio & Lattanzi, Desiree & Monti, Emanuele. (2014). "Automatic Segmentation of MR Brain Tumor Images using Support Vector Machine in Combination with Graph Cut". NCTA 2014 - Proceedings of the International Conference on Neural Computation Theory and Applications. 152-157. 10.5220/0005068501520157.

[74] A. Rajendran, R. Dhanasekaran, "Fuzzy Clustering and Deformable Model for Tumor Segmentation on MRI Brain Image: A Combined Approach," *International Conference on Communication Technology and System Design, 2011*.

[75] Avanzo M, Stancanello J, El Naqa I. "Beyond imaging: the promise of radiomics." Phys Med 2017;38:122-39.

[76] Erickson BJ, Korfiatis P, Akkus Z, Kline TL. "Machine learning for medical imaging." Radiographics 2017;37:505-15.

[77] https://www.mayoclinic.org/tests-procedures/needle-biopsy/about/pac-20394749, Mayo Clinic, Needle biopsy, Last Accessed Date: Accessed: April 24, 2019.

[78] https://www.researchgate.net/post/What is the difference between MRI scans, Accessed: April 24, 2019.

[79] Ravi Srisha, Am Khan, "Morphological Operations for Image Processing : Understanding and its Applications", *National Conference on VLSI, Signal processing & Communications , December 2013.*

[80] https://stackoverflow.com/questions/44179793/segmenting-an-object-from-background-using-matlab-based-on-feature-points/44180904#44180904, Last Accessed Date: April 26, 2019.

[81] https://home.deib.polimi.it/matteucc/Clustering/tutorial html/cmeans.html Cluster-

ing - K-means, Last Accessed Date: April 28, 2019.

[82] Rohini Paul Joseph, C. Senthil Singh and M.Manikandan, "Brain MRI Image Segmentation and Detection in Image Processing", International Journal of Research in Engineering and Technology, 2014.

[83] https://www.mathworks.com/matlabcentral/fileexchange/25532-fuzzy-c-means-segmentation, Last Accessed Date: April 30, 2019.

[84] Soumya.D.S, Arya.V, "Chromosome Segmentation Using K-Means Clustering", International Journal of Scientific & Engineering Research, Volume 4, Issue 9, September-2013 937 ISSN 2229-5518.

[85] https://www.pyimagesearch.com/2015/11/02/watershed-opencv/, Last Accessed Date: May 04, 2019.

[86] https://medium.com/machine-learning-for-humans/neural-networks-deep-learning-cdad8aeae49b, Last Accessed Date: May 10, 2019.

[87] https://hackernoon.com/everything-you-need-to-know-about-neural-networks-8988c3ee4491, Last Accessed Date: May 10, 2019.

[88] Mehrotra, Kishan & Mohan, Chilukuri & Ranka Preface, "Elements of Artificial Neural Nets", 1997.

[89] https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-1-convolution-operation, Last Accessed Date: June 10, 2019.

[90] LeCun, Yann & Bengio, Y & Hinton, Geoffrey. "Deep Learning", Nature, 521. 436-44. 10.1038/nature14539, 2015.

[91] http://prog3.com/sbdm/blog/cyh_24/article/details/50593400, Last Accessed Date: June 10, 2019.

[92] https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5, Last Accessed Date: June 10, 2019.

[93] https://www.medcalc.org/manual/logistic_regression.php , Last Accessed Date: June 15, 2019.

[94] https://medium.com/@AI_with_Kain/understanding-of-multilayer-perceptron-mlp-8f179c4a135f, Last Accessed Date: June 5, 2019.

[95] Sona Taheri, Musa Mammadov, "Learning the Naive Bayes Classifier with Optimization Models", Int. J. Appl. Math. Computer Sci., 2013, Vol. 23, No. 4, 787-795.

[96] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2008). The Elements of Statis-

tical Learning (2nd ed.). Springer. ISBN 0-387-95284-5.

[97] Ho TK (1998). "The Random Subspace Method for Constructing Decision Forests", IEEE Transactions on Pattern Analysis and Machine Intelligence. 20 (8): 832-844.

[98] https://www.sciencedirect.com/topics/neuroscience/support-vector-machines, Last Accessed Date: June 15, 2019.

[99] https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47, Last Accessed Date: June 15, 2019.

[100] Kim, J. I. N. H. O., B. S. Kim, and Silvio Savarese. "Comparing image classification methods: K-nearest-neighbor and support-vector-machines." Proceedings of the *6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics. Vol. 1001. 2012.*

[101] Sujan, Md. Hayder Khan et al. "A Segmentation based Automated System for Brain Tumor Detection." (2016).

[102] Cheng, Jun (2017): brain tumor dataset. figshare. Dataset.