Segmentation of Glioma Tumor in Brain MRI Images

Author

Mohammad Kaab Zarrar

00000171322

MS-16-CE

Supervisor

Dr. Farhan Hussain

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

March, 2019

Segmentation of Glioma Tumor in Brain MRI Images

Author

MOHAMMAD KAAB ZARRAR

00000171322

MS-16-CE

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Computer Engineering

Thesis Supervisor:

DR. FARHAN HUSSAIN

Thesis Supervisor's Signature: _____

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

March, 2019

Declaration

I certify that this research work titled "*Segmentation of Glioma tumor in Brain MRI Images"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

MOHAMMAD KAAB ZARRAR

2016-NUST-Ms-Comp-00000171322

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student

MOHAMMAD KAAB ZARRAR

00000171322

Signature of Supervisor

Dr. Farhan Hussain

Copyright Statement

Acknowledgements

I am thankful to my Creator Allah Subhana-Watala to have guided me throughout this work at every step and for every new thought which You setup in my mind to improve it. Indeed, I could have done nothing without Your priceless help and guidance. Whosoever helped me throughout the course of my thesis, whether my parents or any other individual was Your will, so indeed none be worthy of praise but You.

I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout in every department of my life.

I would also like to express special thanks to my supervisor DR. FARHAN HUSSAIN for his help throughout my thesis and also for Machine Learning and Operating System courses which he has taught me. I can safely say that I haven't learned any other engineering subject in such depth than the ones which he has taught.

I would also like to pay special thanks to DR. USMAN AKRAM for his tremendous support and cooperation. Each time I got stuck in something, he came up with the solution. Without his help I wouldn't have been able to complete my thesis. I appreciate his patience and guidance throughout the whole thesis.

I would also like to thank DR. ARSALAN SHAUKAT and DR. WASI HAIDER BUTT for being on my thesis guidance and evaluation committee and express my special Thanks to Muhammad SADAM HUSSAIN for his help.

Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

Mohammad Kaab Zarrar

Dedicated to my exceptional parents and adored siblings whose tremendous support and cooperation led me to this wonderful accomplishment

ABSTRACT

A Brain tumor is an abnormal cell growth in the brain tissues, these tumors are difficult to treat and severely affect the patient's cognitive ability. Out of all brain tumors, gliomas are the deadliest with least survival rate. Glioma is one of the most common primary cancerous brain tumors. They are the most aggressive kind of cancer therefore; a better treatment and planning is crucial for the patient's overall survival. Before starting a treatment, it is essential to correctly differentiate healthy and cancerous tissues of the patient's brain. Both manual and automatic segmentation methods are utilized to segment the glioma brain tumors. With the advent of new approaches, automatic segmentation processes are becoming more effective and clinically accepted. The focus of automatic brain tumor segmentation task is to separate tumor tissue i.e. edema, tumor core from the healthy tissues i.e. white cells, Cerebrospinal Fluid and gray matter. We have developed a novel automatic segmentation framework consisting of ResNet architecture which is based on Deep Convolutional Neural Network (DCNN). Deep Convolutional Neural Networks (DCNN) consists of various layers i.e. convolution, pooling, activation, normalization and fully connected layers. The extra number of layers helps in learning more abstract features of the input. We utilized two-phase training in order to tackle the class imbalance problem in the dataset. Furthermore, we studied various loss function optimizer to fine tune our results. We tested our framework on a benchmark brats 2015 dataset where it achieved state-of-the-art performance and achieved better results on a Dice Score.

Key Words: *Brain tumor, Glioma, Tumor Segmentation, Convolutional Neural Network, ResNet*

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Description and Motivation

Segmentation and quantities assessment of tumors play a valuable role in medical imaging. It is crucial for the monitoring and planning of treatment strategies for the disease. These assessments can provide valuable knowledge about the spatial distribution of the lesions and different lesion types. Gliomas are very difficult to segment as they can be very defused with their surroundings They are poorly contrasted with abnormal and very irregular structures. Also, these tumors can exist in any part of the brain with almost any size and shape.

Early diagnosis of Glioma tumor can play a major role in better treatment and survival. However, automatic segmentation of tumor has been a challenging task till this date due to the variation in tumor characteristics among patients. Low intensity MRI images and irregular tumor shapes are some of the major reasons. The latest machine learning has shown some improvements in effective segmentation of these tumors. Advance convolutional neural network architectures i.e. ResNet [1], DenseNet [2], and Inception [3] architecture have seen to provide good classification results for various datasets.

Glioma is the most common primary tumor of the brain. Approximately 33% of all new brain tumors and 80% of all malignant brain tumors are glioma brain tumors. Glioma tumor occur in the glial cells that surround and support neurons in the brain. Although glioma is not the most common cancer type it has the highest mortality rate among various cancers. About 16,000 new cases of glioma cases are occurred in united states in 2018[4].

These neoplasms are classified as Low-Grade Gliomas (LGG) and High-Grade Gliomas (HGG). LGG consists of about 30% of all new gliomas and they are typically the early stages of the disease (stage 1 and stage 2). Advanced gliomas are categorized as High-Grade gliomas (stage 3 and stage 4) with the occurrence rate of about 70%. The one-year survival rate for glioma patient is 37.2%, the five-year rate is 5.1% and the ten-year rate is just 2.6% from the day of diagnosis, making it most lethal of all cancer types.

Gliomas are very difficult to segment as they can be very defused with their surroundings. Also, they are poorly contrasted with abnormal and very irregular structures. Also, these tumors can exist in any part of the brain with almost any size and shape. Early diagnosis of Glioma tumor can play a major role in better treatment and survival. Before starting a therapy, it is crucial for a medical practitioner to know the which cells are healthy and which ones are cancerous.

Healthy brain tissues are classified in three types as the white matter, gray matter and cerebrospinal fluid (CSF). Tumorous brain regions consist of edema, necrotic tissues and active tumorous tissues. Normally, glioma tissues are diffused into healthy tissues in a way that they are difficult to distinguish from healthy tissues. The goal for the glioma brain tumor segmentation problem is to correctly segment tumor infected tissues from the healthy tissues. Accurate segmentation of these regions is important in case of a glioma brain tumor because an estimation of the volumes of these sub-regions is crucial for planning and treatment follow up. Figure 1.1 Shows the tissue composition in human brain.

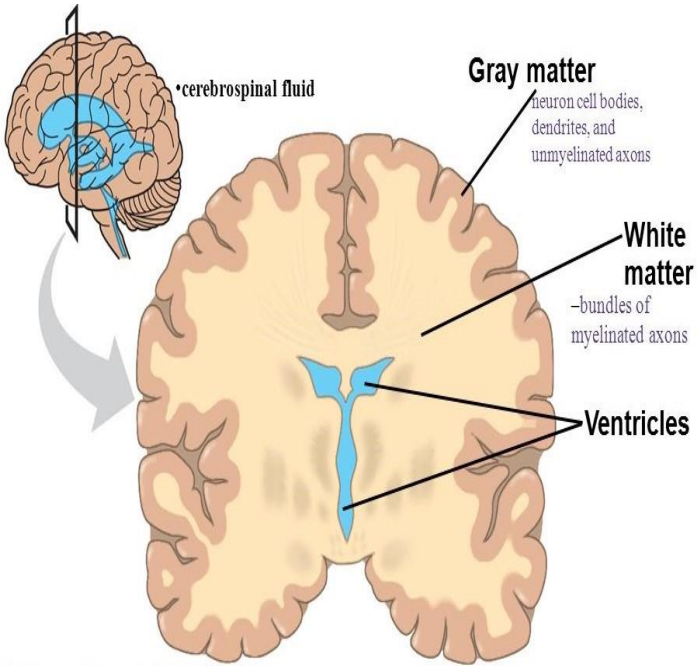

Figure 1. 1 Anatomy of Human Brain [5]

1.2 Brain Imaging Techniques

In the past two decades, there has been a huge advancement in medical imaging, new technologies for imaging have emerge i.e. X-ray, Ultrasonography, Computed Tomography (CT), Positron Emission Tomography (PET), Infrared thermography (IRT), Magnetic Resonance Spectroscopy (MRS) and Magnetic Resonance Imaging (MRI) [6]. MRI is being increasingly used now as days to diagnose cancer of various types including glioblastoma brain tumor tissues. MRI technique uses a magnetic field and radio waves to give a detailed pathology of the body.

Various MRI modalities are being used these days for tumors detection of which few of them are often used to access different brain regions. Each helps individually in providing different pathological information of the tissues inside the body. MRI modalities that are commonly used to diagnose brain tumors include T1(longitudinal relaxation time of tissue) commonly used for distinguishing healthy tissue, T1C (contrast adjustment of T1) can make tumors border appear brighter, T2(transverse relaxation time) can make edema region appear brighter and Flair (fluid attention recovery) pulse sequence can best distinguish edema from cerebrospinal fluid (CSF) [7]. Figure 1.2 shows MRI modalities in BRATS 2015 dataset.

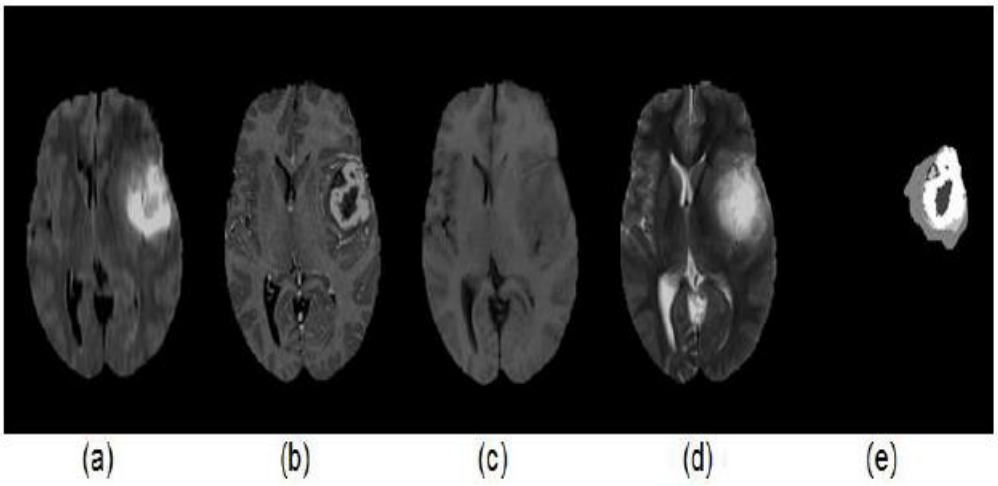

Figure 1. 2 MRI modalities (T1) (b) T1c, (c) T2, (d) T2flair and (e) the ground truth in BRATS 2015 dataset.

1.3 Glioma Tumor Segmentation Types

Glioma brain tumor segmentation can be performed using three methods i.e. manual, semi-automatic and fully automatic method. Manual glioma tumor segmentation is a difficult and time-consuming task as an expert need to evaluate a large amount of MRI images for a single patient. In manual segmentation, pathologist use previous knowledge gained through study and experience to segment tumorous and healthy tissues.

Semi-automatic method requires both human and computer assistance for segmentation. Manual and semi-automatic methods are time-consuming and prone to human errors as they involve human assistance.

Automatic-segmentation does not require human interaction and usually, computer algorithms draw the boundaries on the basis of knowledge which they have acquired during the learning process. Automatic segmentation method can perform segmentation task in less time without involving any human error Therefore, computer-aided medical image analysis is a good

solution. However, this is a challenging task due to the difference in shape, size and locality of these neoplasms. Figure 1.3 shown MRI modalities and segmentation results on BRATS 2015 dataset.

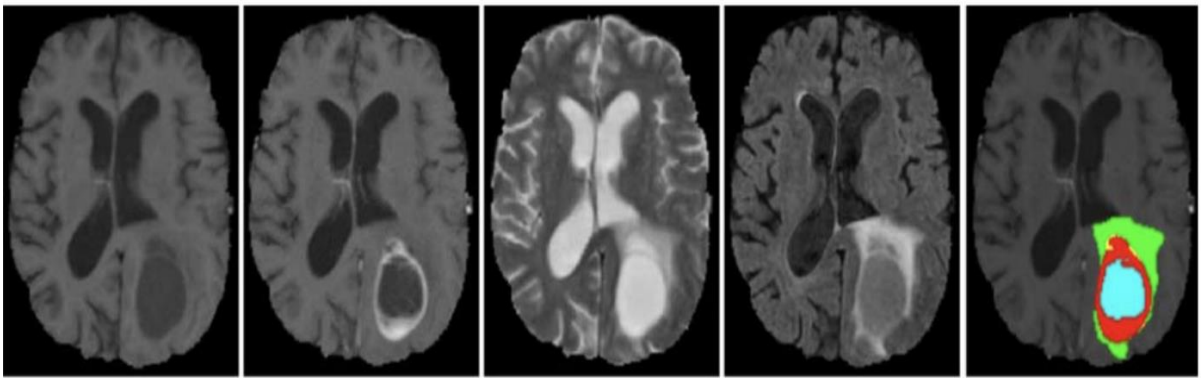

Figure 1. 3 MRI modalities from the BRATS dataset. From left to right: T1, T1C, T2, FLAIR and results after segmentation.

Recently, many automatic image processing techniques have been proposed by the researcher in order to segment brain tumor tissues from the healthy brain tissues This includes deep learning techniques which give optimal results on large datasets without having to specify meaningful features. Traditional image processing techniques can give good results on smaller datasets but the need meaningful features to train them.

Detecting glioma brain tumor can be very difficult as two pixels can have the same properties but different labels. This gives raise to probabilistic machine learning method which labels input image on the basis of their probability. Each image pixel is assigned the class with the highest probability among various classes. Figure 1.4 shows the simplest CNN architecture with two convolution layer, two pooling layer and one fully connected layer.

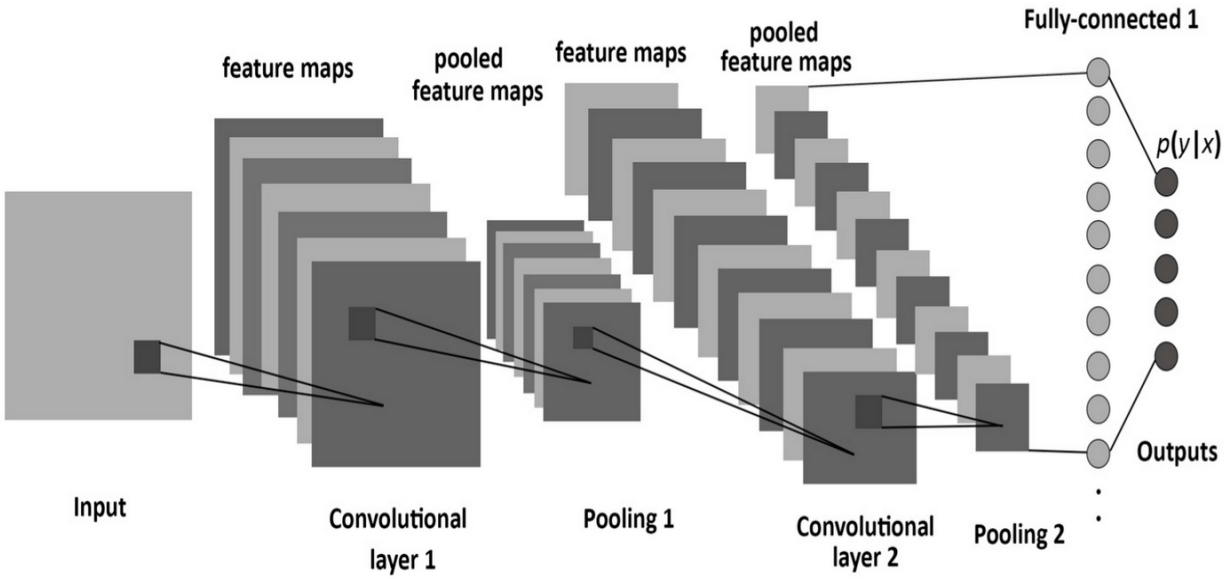

Figure 1. 4 Simplest CNN architecture design [8]

1.4 Research Challenge and Contribution

Automatic glioma brain tumor segmentation is a challenging task due to the difference in shape, size and locality of these neoplasms. These tumors have unclear boundaries with discontinuities which makes it difficult to segment using traditional edge-based methods. Glioma tumor MRI images obtained from clinics and synthetic databases [9] are inherently complex. MRI devices can create Noise problems due to motion and field inhomogeneity during image acquisition. These noises could alter intensity level across an image resulting in bad segmentation performance. To target the problem of glioma tumor segmentation, we implemented various Convolutional Neural Network Architecture. We performed experimentation utilized the ResNet architecture with some modification. The resulting ResNet architecture performs better in segmenting glioma tumors.

We added a dropout layer after every second convolution throughout the architecture. We also changed the loss function optimizer and used the optimizer having best results on our problem.

1.5 Objectives

The main objectives of this study are as follow:

- An automatic glioma brain tumor segmentation framework is proposed with some improvement in results.
- The proposed model uses complex Convolutional Neural Network Architecture to extract deep local and global contextual information from the data in order to segment different glioma tumor sub regions.

- The proposed model uses latest advances such as regularization, ResNet architecture and Non-linear activations.
- We also employed two-phase network training in order to tackle class imbalance problem which yield better performance.

1.6 Organization of Thesis

This research study contains 5 chapters. Chapter 1 explain the challenges in Glioma Brain tumor Segmentation, Imaging Modalities, Segmentation methods, and Research Objectives. In chapter 2 we present literature review of the field of Glioma Brain Tumor Segmentation. In in Chapter 3 we give detail insight of our proposed framework. In chapter 4 we discussed experimental setup and some challenges faced in Glioma Tumor Segmentation. In chapter 5 we give the summary of our proposed work and also discussion on feature work.

Chapter 2 Related Work

Brain tumor segmentation methods can be classified into three types (i.e. manual, semi-automatic and fully automatic segmentation methods) based on the level of human interaction required [10]. All these brain tumor methods are being used nowadays with varying level of performance.

In manual segmentation, radiologists use the information obtained through various multi-modalities images of a single patient with a background knowledge of human brain anatomy gained through experience and training [11]. To do that, radiologists manually draw the boundaries of a brain tumor and paint the different tumor regions with different colors. This is a time-consuming task as an expert has to go through each slice of the patient's MRI. Other than that performance totally depends on the expert's experience and care. Therefore, semi-automatic and fully automatic methods are being used nowadays in conjunction with manual methods for better performance.

In semi-automatic methods user initialize the process by inputting some parameters, waiting for results and to provide feedback response for the software computation. This whole process involves initialization, feedback response and evaluation [12]. Semi-automatic segmentation process may yield better results but results largely depend upon the expert and may vary for same experts on repetition. The semi-automatic process gives better results but it's still a time-consuming method.

Thomas et al. [13] proposed a method for semi-automated tumor segmentation based on region growing segmentation tool. A total of 320 segmentation of Flair and MPRage sequences were performed using a smart brush tool (a region growing based semi-automated tool). The algorithm starts with a region growing algorithm aiding in segmentation, after that a 2-D segmentation was manually performed which was 3-D interpolated after performing another perpendicular 2-D segmentation. Small changes were also made with the help of region growing tool or complete manually. The proposed methodology achieved a good performance but performance varied for each individual due to the presence of manual assistance in the methodology.

Nowadays most of the research is being carried on a fully automatic segmentation process. It does not involve any user interaction and is considerably less time-consuming. However, due to the irregular shape, varying size and location of tumors, this is one of the most challenging segmentation tasks. We divided fully automatic methods into four major types which are

Neural Network methods, Conventional Image processing methods, Clustering methods and Conventional Machine Learning method. Figure 2.1 shows a flowchart of a generic CAD system.

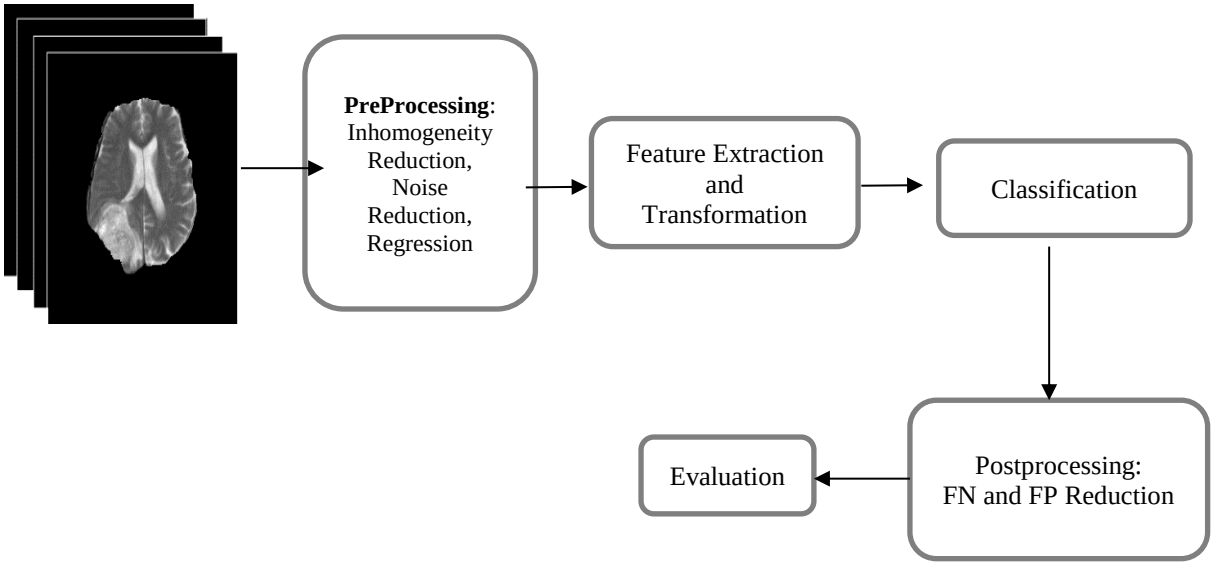

Figure 2. 1 Flowchart of a generic Computer Aided Diagnosis (CAD) System

2.1 Conventional Image Processing Methods

Most conventional image processing methods focus on the shape of an image. The value of pixels in an output image depends on the corresponding pixel in an input image with its surrounding. Region growing is the most common morphological operation algorithm and is used to extract connected pixel of the similar region from an image. Similarity criteria mostly depend on the range of pixel intensity values or other features. Region base method has a disadvantage known as partial volume effect. It occurs on the pixel that is actually the border of two tissue type causing the pixel to blur as the voxel represent more than one tissue type.

Sudharani et al. [14] proposed a morphological technique for brain tumor segmentation in which they first applied brightness adjustment technique followed by the resampling of an image. After that they converted grey images into color images followed by a histogram normalization technique. Then tumor area was calculated using geometrical operations which were followed by a threshold technique to highlight region of interest. Fast Fourier transform and lookup table conversion were applied before finally applying erosion and dilation. Their system achieved an accuracy of 89.2%. Ishmam et al. [15] proposed a method in which they applied a dynamic threshold to obtain a region of interest followed by k-mean clustering to separate out tumor regions. After this region growing technique was utilized with high

tolerance value of 12. Proposed method achieved efficiency of 0.85 on brats 2012 data set. Their method was able to calculate the area of the tumors with significant accuracy.

It has been noted that the BRATS 2012 dataset contain few numbers of images so they are not efficient on deep learning algorithms as these algorithms need a larger dataset to perform well. However conventional image processing-based algorithm can work well on smaller dataset as well. This is the big advantage of using conventional image processing technique over deep neural networks.

2.2 Conventional Machine Learning Methods

Conventional machine learning methods provide an efficient way to automate the analysis and diagnosis of medical images. These algorithms can effectively reduce the burden on radiologist *in the practice of radiology. These methods fall into supervised learning category which can learn features by their labels and make an accurate prediction. Some most commonly used methods include Random Forests, Support Vector Machine. Random Forest is a flexible, easy to use algorithm that can produce great classification result most of the time. They are most widely used algorithms based on their simplicity and ability to be utilized for both classification and regression.

Random forest RF is a supervised learning method, it creates random forests which is an ensemble of decision trees trained using the bagging method. Nicholas et al. [16] proposed a supervised learning approach in which they used multiple sets of features such as intensity, geometry and asymmetry to segment whole brain and tumor regions on the basis of random forests-derived probabilities.

They arranged a matrix of features with their label and gave the data to random forest training algorithm to predict the label. Each sample of an image passes through each tree of the ensemble where they were labeled according to their class. Each vote was converted to voxel-wise probabilities estimate for each class via some specified mechanism. They used two stages of RF training to perform the brain tumor segmentation task.

Mohammadreza et al. [17] proposed a novel segmentation approach based on superpixel and their classification. They extracted a number of unique features from each superpixel of the Flair MRI images. They trained these features on Extremely Randomized Trees (ERT) algorithm and separated turmeric and non-turmeric regions. Furthermore, they also trained these features on the support vector machine (SVM), in order to compare their results.

Support vector machine when given the labeled input training data outputs an optimal hyperplane which can categorize each new test example. Support vector machine, first

introduced by Vladimir Vapnik is highly useful real-time algorithm as it uses less computation power with significant accuracy. In medical imaging, such algorithms can be very handy where high processing power is difficult to achieve and time is critical.

Samya et al. [18] proposed a technique based on random forest and support vector machine. Their method first used RF to classify foreground input pixel and generate segmentation results which were then forwarded to SVM classifier. SVM classifier then performed segmentation on a large region of interest (ROI) which was missing in the first stage. Hence SVM targeted global features while RF targeted more local feature. These two steps were iterated until optimal results were achieved.

2.3 Clustering Methods

Clustering is a type of unsupervised machine learning algorithms mean there is no prior information about pixel labels. In supervised machine learning algorithms, each sample has two parts: one is an input feature and the other is its label. The purpose of supervised learning is to obtain a functional relationship from training data that also generalizes on testing data. Unsupervised learning algorithms can be very beneficial in the scenario where pixel labels are not given. One advantage that the clustering technique gets over the deep neural network is the execution time, as these techniques are much faster in execution compared to the deep neural networking techniques. The table 2.1 gives detail insight of advantages and disadvantages of various techniques.

Table 2. 1 Advantages and disadvantage of various image processing method.

No.	Method	Advantage	Disadvantage
1	Clustering	clustering is a fast and relatively easy algorithm for small data sets. Clustering techniques are much faster in execution compared to the deep neural networking techniques.	Performance is significantly less in comparison with Deep Convolutional Neural Network techniques.
2	SVM	SVM is highly useful real-time algorithm as it uses less computation power with significant accuracy. Also, SVM algorithm can perform well even when the training data is not very large. Once the boundaries are defined small changes in data would not affect the results hence avoids overfitting.	Again, performance is significantly less compared to Deep Neural Network Methods.
3	Random Forest	Random forests are a very simple algorithm that is able to be utilized for both classification and regression. Like SVM method random forests reduces variance and helps is avoiding overfitting of	Unable to learn nonlinear low-level representation as feed forward neural network does. Another disadvantage of random forest technique is its poor performance on unbalanced class data. (BRATS dataset is an unbalanced class problem).

		the data. Powerful method with good performance.	
4	Conventional Image Processing Methods	These methods i.e. geometrical operations need very less amount of training data. Takes less amount of time for classification.	Performance is less compared to Deep Neural Network DNN method.
5	DNN	Most classification algorithms need meaningful features as input while DNN can learn meaning features on its own from the training data. Deep learning methods Outclasses other classification problem by a significant margin in multiple domains.	High computational cost meaning high performance hardware need to run DNN methods for days and in some cases up top weeks. They also need more computational time compared with SVM and Random forest methods. They need large training data to achieve significant performance this is to say that if there are only thousands of examples then deep learning algorithm would not perform very good. These methods lack theoretical foundation and it is very difficult to understand what's happening at the backend. Hyperparameter adjustment is difficult with very little to no real theory behind them.

There are many systems which use k-mean clustering for detection of brain tumors. K-mean clustering is a fast and relatively easy algorithm for large data sets but it sometimes gives

incomplete detection of tumor which can be costly if the tumor is malignant. In K-mean clustering, each data point must exclusively belong to one cluster center. Some systems, on the other hand, uses fuzzy C-mean FCM clustering and they work significantly well in segmenting all parts of the tumor.

In fuzzy C-mean clustering, a point must belong to at least one or two clusters center. Eman et al. [19] introduced a hybrid clustering method called k-mean integrated with fuzzy C-mean clustering (KIFCM) method. They first de-noised the input images and fed them to KIFCM algorithm. The KIFCM divided the tumor region from the healthy regions. It was helpful in segmenting more scattered point and categorize them into one or more classes. They evaluate their results on three datasets i.e. DICOM dataset, Brain Web data set and BraTS data set.

James et al. [20] introduced another method based on Otsu and fuzzy C-mean clustering technique, preoperative tumor region of interests blobs for Otsu and Fuzzy processing were created with VelocityAI tool. These blobs were then given to the clustering algorithms for tumor segmentation. The clustering technique was applied on Otsu and Fuzzy C-mean clustering technique using 3 and 4 classes. Their Fuzzy3 (three classes) algorithm was the best algorithm to performed with dice score of 0.91. Table 1 gives comparison of various image processing methods used for glioma tumor segmentation.

2.4 Neural Network Methods

Neural Network is the most commonly used brain tumor classification algorithms and are very popular among researchers due to their high performance. Common neural networks can contain a variety of layers i.e. input layer, convolution layer, pooling layer, drop-out layer, fully connected layer and a final output layer. Convolutional neural networks are the most commonly used deep neural network method for image recognition and segmentation. Most classification algorithms need meaningful features as input while CNN can learn meaning features automatically from the training data.

CNN algorithms take input images in the form of patches and pass them to convolutional filters and pooling layers to extract complex local and global features. Various CNN models have been proposed by various researchers and the performance depends on how well a CNN architecture can extract these features from the input data. Figure 4 compare the top CNN architecture designs these days. The comparison is done in the basis of ImageNet competition also called ILSVRC. The graph in figure 2.2 shows the ResNet architecture has the minimum error rate of 3.57%.

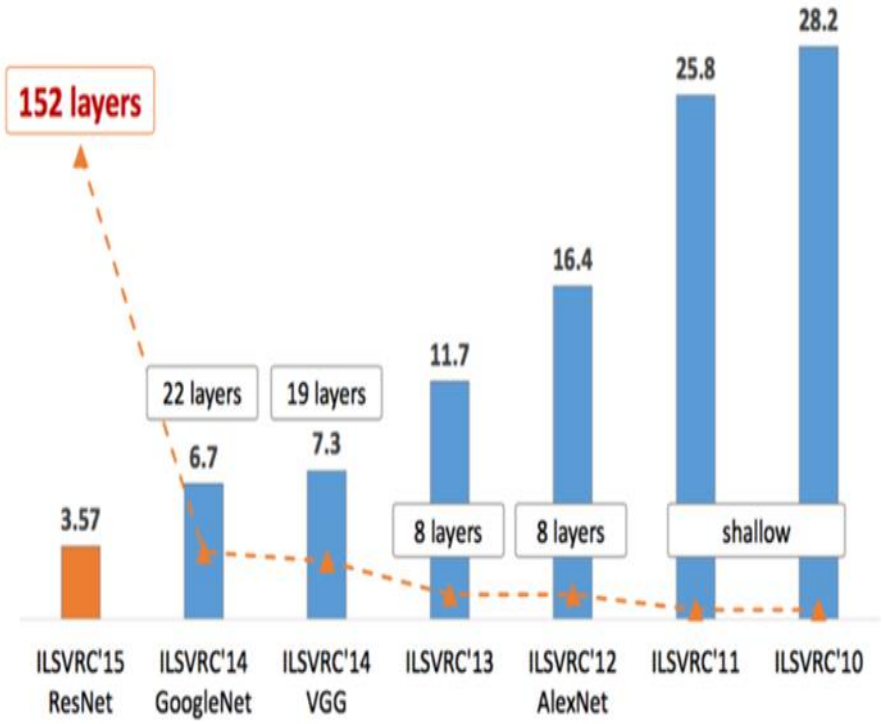

Figure 2. 2 Winner Architecture design of ILSVRC competition (yearly)[21]

Mohammad et al. [22] proposed an architecture in which they used small input patch as which were convolved with series of 3X3 kernels followed by a max-out and max-pooling layer. Max-out layer compares the input feature values and outputs the maximum value of a feature map for each spatial position. They also used two pathway architecture meaning the architecture contains two streams of input patches i.e. one with small 7X7 (local) receptive field and another with 13X13 large (global) receptive field. Both paths yield different feature map and they were concatenated to yield a single feature map with both local and global receptive field features. Their algorithms achieved dice score of 0.85 on Brats dataset. Table 2.2 gives comparison of ILSVRC winner network architectures.

Table 2. 2 Comparison of Top CNN Architectures

CNN Algorithm	Convolution Layers	MACCs (Millions)	Parameters (millions)	Activation (millions)	ImageNet Top-5 error
AlexNet	5	1140	62.4	2.4	16.4%
VGGNet-16	16	15470	138.3	29.0	8.1%
GoogleNet	22	1600	7.0	10.4	6.7%
ResNet-50	50	3870	25.6	46.9	3.5%

Saddam et al. [23] proposed an architecture design in which two CNN'S were used in such a way that output of the first network was concatenated with the input of the second network and thus named it a 'nexus architecture'. They proposed a variety of nexus architecture designs. Input patch of 33X33 size was extracted from BRATS dataset. Furthermore, they used the dropout and batch normalization layer to further optimize the architecture. They also applied N4ITK and intensity normalization technique as a pre-processing step. These complex nexus architectures were able to achieve good dice score on BraTs 2015 dataset.

Most of the brain segmentation techniques are on a 2-D plane while Konstantinos et al. [24] proposed a method in which they used a 11-layer deep 3D brain tumor segmentation architecture. They used smaller kernel for their architecture for more in-depth feature field. They also exploit parallel multiscale processing to get both local and contextual information for this they input of two pathways were centered at the same location, but the second segment which was also low resolution was down sampled by a factor of 3. Their architecture achieved 0.90 performance on the dice score. Figure 2.3 shows a comparison between 34-layer deep plain architecture vs 34-layer deep ResNet architecture.

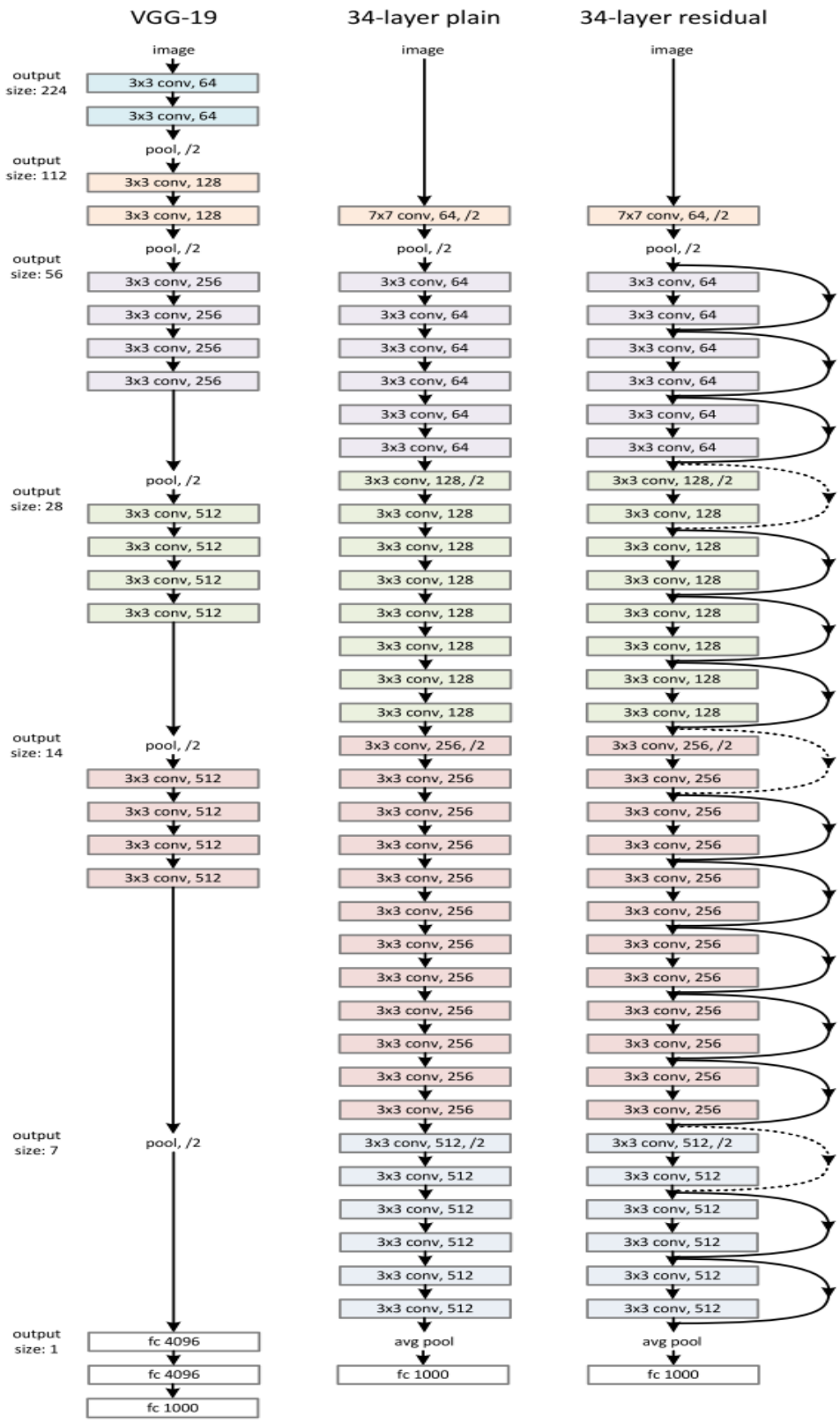

Figure 2. 3 ResNet Architecture design

Liu et al. [25] proposed a different architecture design in which they introduced a Dilated Convolution Refined (DCR) network consisting of 5 parts and ResNet50 architecture as a

backbone. The goal of this architecture was to extract most of the local as well as global features. Their network was divided into five slices. The first slice contained both input and output. On the right side of each slide, the deconvolution process was being carried out. While on the left side of top to bottom of each slide reset50 architecture was used. Each ResNet50 connect into DCR and result of each DCR were added with the deconvolution process of the lower layer thus making a complex architecture. Their proposed algorithm performed well on dice score.

Sérgio et al. [26] also applied N4ITK, and intensity normalization technique as a pre-processing step. They also included a post-processing step in which they remove smaller clusters with a size smaller than the threshold level. Their algorithm was the runner of BRATS 2013 competition. It has been observed that most brain tumor algorithm does not have a post-processing in their technique, that can be helpful in increasing the performance of the system. Therefore, a researcher should also emphasize on post-processing of their algorithm. Table 3 Compares the results of top most Glioma tumor segmentation technique on the basis of dice score.

2.5 Datasets

Several Datasets have been presented by different committees to encourage the researchers to take part in brain segmentation. Some of the most used publicly available dataset includes Brain Web [27], Internet Brain Segmentation Repository (ISBR) [28] and BRATS [29] for tumor segmentation, Isles [30] for evaluating stroke, MSSEG [31] for lesion segmentation and detection on MS data and NeoBrainS12 [32], MRBrainS [33] for segmentation of various brain regions. Most automatic brain tumor segmentation methods are performed on the datasets listed below as they allow to reproduce and compare the results between different studies.

2.5.1 BRATS

In order to measure and compare the various brain tumor segmentation methods, the Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) challenge was introduced in 2012. It consists of a large data set of brain tumor MRI images with five classes: Healthy tissue, edema, non-enhanced, necrosis, and enhanced regions of tumors. The training dataset size has grown over the years. The dataset contains both low grade and high-grade cases. It contains four type of imaging modalities i.e. T1 MRI, T1 contrast-enhanced MRI (T1C), T2 MRI, and T2 FLAIR MRI. All images in the BraTs datasets have the same 1 mm voxel resolution. BRATS dataset is considered as a benchmark dataset for comparing results between

various brain tumor segmentation techniques. Figure 2.4 shows detailed modality information for BRATS 2015 dataset.

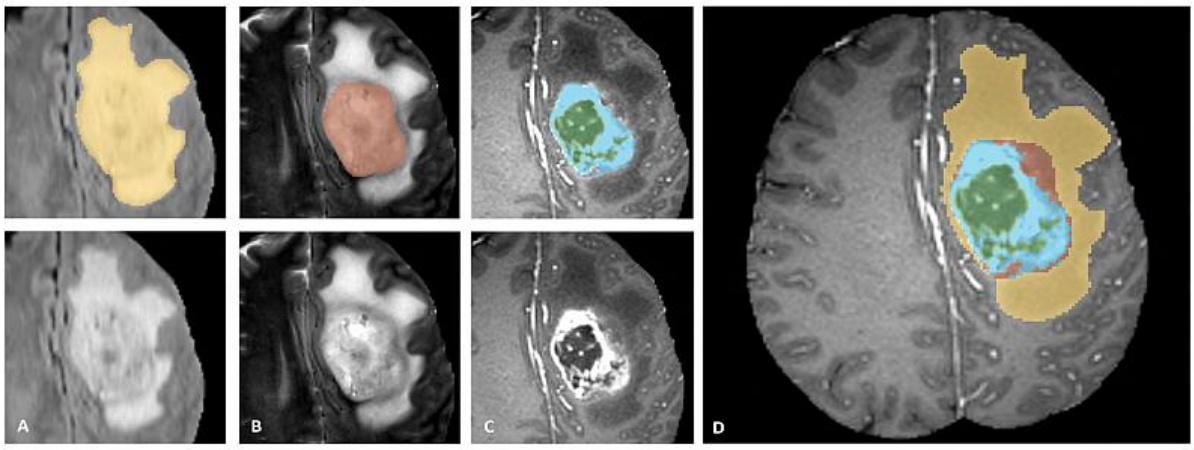

Figure 2. 4 Glioma Images and tumor sub regions annotated using different MRI modalities (Top Left), and the final labels of the entire patient brain (Figure D). Glioma images show from left to right: the whole tumor in yellow shown using T2-FLAIR (Figure A), the tumor core in red shown using T2 (Figure B), the enhancing tumor structures in light blue) shown using T1Gd, surrounding the necrotic parts of the core (green) (Figure C) [27]

2.5.2 ISBR

It is an MR Images dataset provided by Center for Morphometric Analysis at Massachusetts General Hospital. This dataset is divided into two sets which are known as IBSR18 and IBSR20. IBSR18: This dataset consists of T1-w scans with the slice thickness of 1.5mm. It does not contain any noise that can decrease the accuracy of scans. The dataset has been preprocessed with Autoseg bias field correction routines. The dataset is provided with manually labeled images as well as three class labeled images. ISBR20: This dataset contains T1-w scans with a slice thickness of 3.1mm. The dataset is labeled with the help of semi-automated algorithms. The provided labels are GM, WM and CSF tissues. The image has a lower resolution as compared to the IBSR18 dataset and are sorted by the level of difficulty with the most difficult scans containing the noise and irregularities.

2.5.3 DICOM

DICOM is the biggest online medical image and video file sharing library. It is one of the most used free and publicly available dataset platforms. It offers a wide variety of medical datasets for various diseases including glioma brain tumors. It contains numerous modality images i.e. Optical Coherence Tomography, Mammography and MR images. Many researchers prefer

DICOM images for brain tumor segmentation problem, however, their dataset has relatively smaller number of images compared to BRATS dataset

2.6 Performance Evaluation Metrics

Various performance measuring matrices are used these to evaluate and compare the accuracy of a model. These methods evaluate performance using multiple parameters i.e. true positive, true negative, false positive and false negative. A list of various performance measuring methods to evaluate brain MRI segmentation quality and their mathematical expression are listed in table 1. Table 2.3 shows Various performance evaluation metrics used for Segmentation of Glioma Tumor segmentation

Table 2. 3 Various performance evaluation metrics used for Segmentation of Glioma Tumor segmentation

Sr No.	Method	Mathematical Expression
01	Dice Coefficient	$\frac{2TP}{2TP + FP + FN}$
02	Sensitivity	$\frac{TP}{TP + FN}$
03	Specificity	$\frac{TN}{TN + FP}$
04	Precision	$\frac{TP}{TP + FP}$
05	Recall	$\frac{TP}{TP + FN}$
06	Jaccard	$\frac{TP}{TP + FP + FN}$
07	Volume Difference Rate	$\frac{FP - FN}{TP + FN}$
08	Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$

2.7 Challenges and Problems

Automatic glioma brain tumor segmentation is a challenging task due to the difference in shape, size and locality of these neoplasms. These tumors have unclear boundaries with discontinuities which makes it difficult to segment using traditional edge-based methods. Glioma tumor MRI images obtained from clinics and synthetic databases are inherently complex.

MRI devices can create Noise problems due to motion and field inhomogeneity during image acquisition. These noises could alter intensity level across an image resulting in bad segmentation performance. Also, the effectiveness of deep learning-based methods relies on high capacity of model and millions of labeled examples. Acquiring such large set of training data is not an easy task in medical imaging field. Furthermore, several image modalities are needed to segment tumor sub region which further increase the complexity. Table below shows Comparison of the various state-of-the-art brain tumor segmentation methods.

Table 2. 4 Comparison of the various brain tumor segmentation methods. Note that only Dice Score is considered for performance measure

Sr No.	Author	Classification	Method	Year	User Interaction	Dice score	Datasets	Reference
1	Havaei et al.	Neural Network	Cascaded CNN architecture	2016	Automatic	0.85	BRATS 2013	[22]
2	Jia et al.	Neural Network	Multi-scale feature fusion	2018	Automatic	0.87	3-D MPRAGE-private	[35]
3	Di et al.	Neural Network	Dilated Convolution refine network	2018	Automatic	0.87	BRATS 2015	[25]
4	Sérgio et al.	Neural Network	CNN with deeper architecture	2016	Automatic	0.88	BRATS 2015	[26]
5	Konstantinos et al.	Neural Network	3D CNN with two convolutional pathways	2016	Automatic	0.85	BRATS 2015, ISLES 2015	[24]
6	Huber et al.	Conventional Image Processing	Region-growing segmentation tool	2017	Semi-automatic	0.88	3DMPRAGE-private	[13]
7	Eman et al.	Clustering	Integrated k-mean and fuzzy c-mean	2015	Automatic	0.86	DICOM	[19]
8	Nicholas at al.	Conventional Machine Learning	Multivariate symmetric template based on random forest	2015	Automatic	0.87	BRATS 2013	[16]

9	Samya et al.	Conventional Machine Learning	Random forest transfer to support vector machine	2017	Automatic	0.83	BRATS 2012	[18]
10	Mohammad-reza et al.	Conventional Machine Learning	Classification of superpixel based on random forest	2017	Automatic	0.88	BRATS 2012	[17]
11	Ines et al.	Conventional Machine Learning	Graph cut distribution matching approach	2015	Semi-Automatic	0.77	BRATS 2012	[36]

Chapter 3 Methodology

In our proposed work we targeted the problem of glioma brain tumor segmentation on ResNet (Convolutional Neural Network) architecture. We acquired images from BRATS 2015 dataset. The proposed methodology contains three steps i.e. pre-processing, patch formation and implementation of different CNN architecture. At first, input images are preprocessed and split into patches and then these patches are passed through a CNN architecture. Figure 2 shows the flowchart of our proposed framework. The detailed description of these steps is described in the following. Figure 3.1 block diagram of our proposed model.

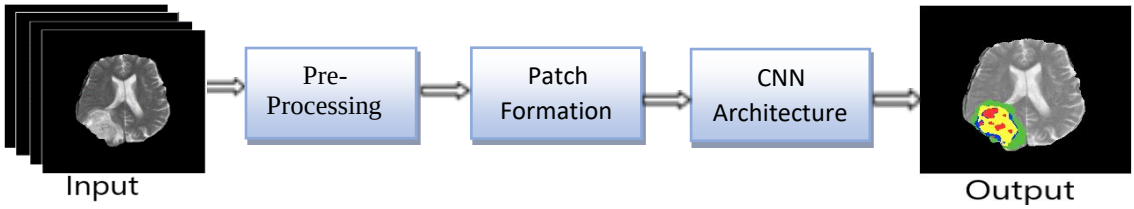

Figure 3. 1 block diagram of our proposed model.

3.1 Patch Formation

The BRATS dataset contains 3-D MRI images which is converted into 2-D image slices. Each 3-D brain image is converted into 155 2-D MRI slices of 240X240 pixel size. After that patches are extracted from each of these slices and CNN is trained on these patches. We tested on different size of patches to see which patch size give better results. After experimentation we choose patch size to be 33X33 throughout the dataset. The label of the center pixel of each patch is assigned as a label to that whole patch. Similarly, same step is repeated throughout the dataset.

3.2 Pre-Processing

MRI images come with Noise problems due to motion and field inhomogeneity during image acquisition. These noises could alter intensity level across an image resulting in bad performance. Two pre-processing techniques are applied to enhance our input images.

1. N4ITK [37], which is a bias field correction method is applied to the input images to reduce this artifact. N4ITK removes inhomogeneity in the input data created during the acquisition of MRI scans. Non-Parametric, non-uniform intensity normalization also

known as N3 algorithm is well known method to remove intensity normalization caused by the artifacts. N4ITK is the improved version of N3 normalization. To implement this technique, we used 3D slicer toolkit in python. We used its 4.6.2 version which is simpler and easy to implement. It is an open source software which helps to visualize and process the 3D MRI images [38,39].

2. We also applied intensity Normalization on each image in our dataset. Intensity normalization process transform the pixel intensities into a desirable range throughout the image. In this process we removed 1% top and bottom intensity values throughout the dataset which helps improving the learning process during training.

Normalization is performed using formula given below:

$$x_n = \frac{x - \mu}{\sigma} \tag{3.1}$$

Where X is the input slice, whereas μ and σ respresents the mean and standard deviation values in the given image. We have seen that these pre-processing steps increases the training accuracy. The effect of pre-processing techniques i.e. N4ITK and normalization is shown in the figure given below. Figure 3.2 compares the MRI scan before and after applying pre-processing on BRATS 2015 dataset.

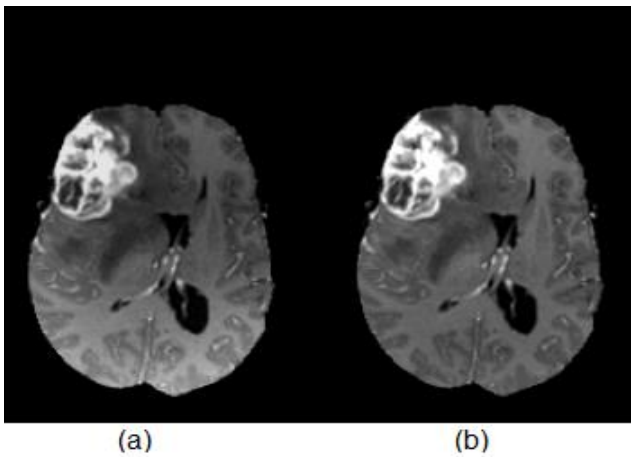

Figure 3. 2 (a) MRI scan before (b) MRI scan after bias filed correction.

3.3 Convolutional Neural Network

Convolutional Neural Network has been known for a while for their remarkable performance in pattern recognition [40,41,42]. A Convolutional Neural Network consists of input layer,

output layer and number of hidden layers. A Hidden layer may consist of layer such as pooling, activation, convolution, normalization, dropout or fully connected layer. Each of these layers performs unique functionalities. However, a CNN must have one of each convolution, activation, fully connected and output layer. These layers lie on top of each other in a hierarchical fashion resulting a feature map. Each layer takes input from its previous layer and output these features to the next layer until it reaches the final layer. Final layer is usually preferred to be Softmax layer. Softmax layer outputs the probabilistic distribution of data based upon the input feature maps. Convolution layer is the building block of Convolutional Neural Network.

A Convolution layer convolves with the convolutional filters which generates feature maps. These filters come in various sizes i.e. 3 × 3, 5 × 5 and 7 × 7. Convolution filters (also known as kernels) are usually small box shape object usually similar to the one of the objects present in the images. The resultant feature contains useful information called feature maps in the form of small box. Each feature map corresponds to some hidden units call neurons; These neurons are controlled by activation layer.

There are various kind of activation layer used in convolutional Neural Networks such as Sigmoid, Tanh, ReLU, leaky ReLU and max-out [43]. The activation influences the neighboring voxels in the feature map. The area in the feature map is called neuronal receptive fields the size of which increases in each subsequent layer. Each neuron in a layer is connected with the preceding layer through weighted connections.

The Proposed framework process four patches namely T1, T1c, T2 and Flair. These four patches are counted as one input which is forwarded to the network. The network assumes the input as the one input with four channels. Each channel represents one modality. These behaviors are similar to red, green and blue channels of color image. The process continues throw-out each slice of the entire brain of each patient. 155 2-D slices of each patient are processed in the network. A feature map M_p is obtained as.

$$M_p = b_p + \sum_r K_{ar} * L_r \tag{3.2}$$

Where K_{ar} is a convolution kernel, L_r is input plan, b_p is the bias value, and $*$ represent the convolution process.Figure 3.3 shows a simplest CNN architecture design containing one convolution, pooling and fully connected layer each.

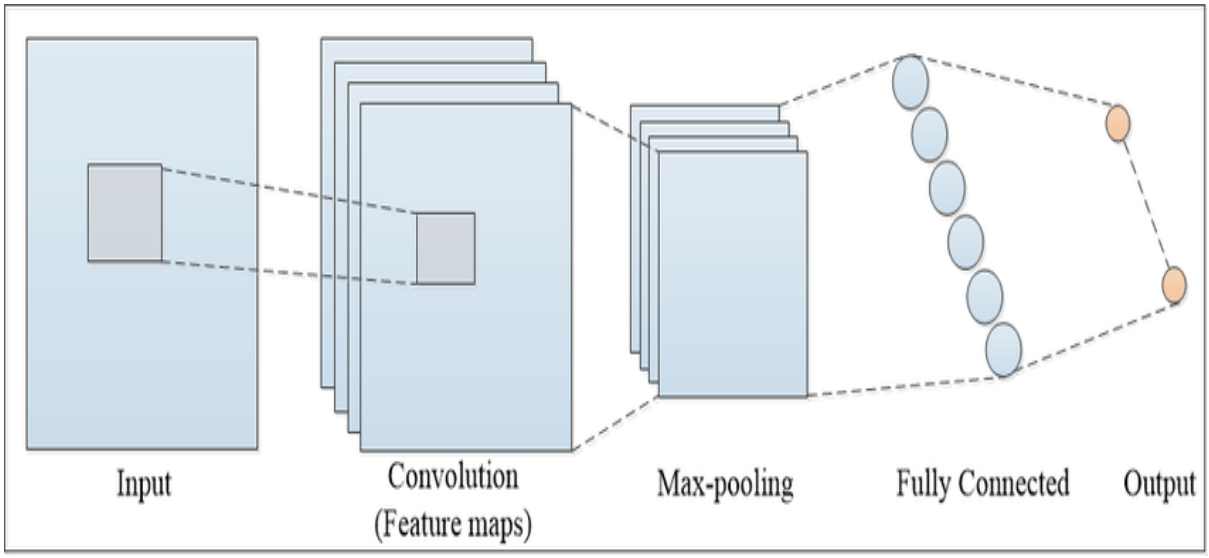

Figure 3. 3 A Simplest Convolutional Neural Network Architecture containing one of each Convolution, Pooling, fully connected and an Output layer.

The Convolutional Kernels and connected weights are update by a method called back-propagation algorithm [44]. In Back-Propagation, input images are passed through the network via feed forward pass and the network make some class prediction. These predictions are then compared with the actual prediction. The resultant error in prediction is forwarded starting from the last layer to the first layer of the network.

CNN algorithm deals with the useless feature values through pooling layers. In max-pooling layer network only keeps the useful feature map and discarded the extra values in each window. The resultant feature map has a size smaller than the original feature map which reduces the complexity and dimension as well as the load upon the network training. There are some hyper-parameters which describe the size of the resultant feature map. Let's assume, N × N is the size of the feature map before pooling, where p and s are the of pooling and stride ratio respectively. After pooling the resultant feature map R will be calculated as R = (M - p) / (s + 1). The pooling layer works in a sliding window fashion on each point $V_{b,i,j}$ of the feature map M, it keeps the maximum value in the window of length k and discard the rest as:

$$V_{b,i,j} = \max\left(M_{\mathrm{b,i+k,j+k}}\right) \tag{3.3}$$

Convolutional Neural Networks can learn complex features through a hierarchy of feature maps. The ability to learn complex feature make them very useful for image processing and pattern recognition tasks. CNN architecture consists of various activation layers. The activation influences the neighboring voxels in the feature map. The area in the feature map is called neuronal receptive fields the size of which increases in each subsequent layer. Each neuron in a layer is connected with the preceding layer through weighted connections.

The last layer of the CNN consists of a non-linear activation function that converts input feature maps into class probabilities. The class with the highest probability is assigned to the label of the input patch. The probability K of each class c from various classes E is calculated as:

$$K\ (y= c|z) = \frac{e^{zw_j}}{\sum_{E=1}^{E} e^{zw_E}} \quad (3.4)$$

Where z and w are the input feature map and weights respectively.

3.4 Architecture Setup

Convolutional Neural Network layers can be arranged in a number of ways to increase the effectiveness of learning process. Usually CNN contain tens of layer stacked over top of each other in a hierarchical manner. These layers can be arranged in sequential as well as parallel and, number of layers can be concatenated as one layer which outputs the preceding layer resulting in a complex CNN architecture. We have implemented three types of CNN model which are VGG architecture, two-phase VGG architecture and ResNet architecture. The detailed description of these deigns is described in the following.

3.5 VGG Architecture

This architecture based on Convolutional Neural Network (CNN) is our baseline architecture to solve the problem of Glioma Brain tumor segmentation. Common CNN layers i.e. convolution, pooling and ReLU are used as a building block of VGG network. These layers lie on top of each other in a hierarchical fashion resulting a feature map. Each layer takes input from its previous layer and output these features to the next layer until it reaches the final layer. We used max-pooling to drop weak features in the feature map.

Final layer is kept to be Softmax layer. Softmax layer outputs the probabilistic distribution of data based upon the input feature maps. We used ReLU as an activation function. The network process four patches namely T1, T1c, T2 and Flair. We also used dropout layer to drop weaker feature and avoid overfitting. Figure 3.4 Shows architectural details of our proposed VGG19 architecture design.

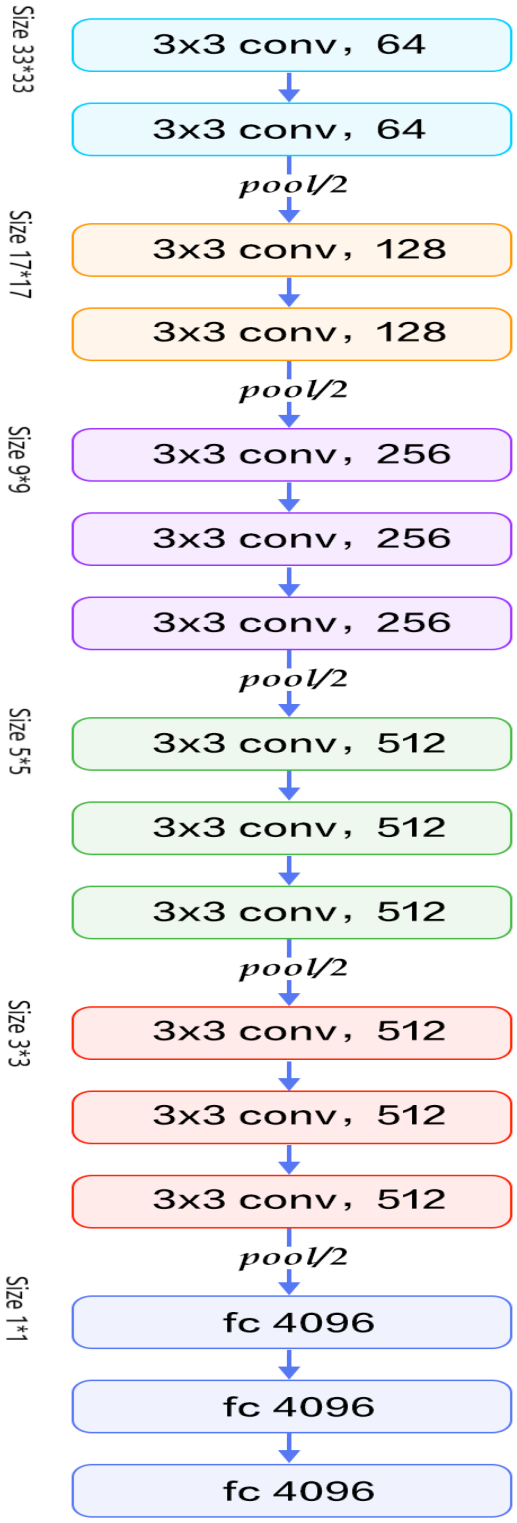

Figure 3. 4 Basic design of VGG architecture

3.6 Proposed Modified ResNet Architecture

Convolutional neural network (CNN) consists of multiple layers i.e. convolution, pooling, activation, dense, batch normalization and dropout. Each of these layers perform unique

functionalities. These layers lie on top of each other in a hierarchical fashion resulting a feature map. Each layer takes input from its previous layer and output these features to the next layer. Convolutional layer is the most important layer in a CNN as they are the building blocks of Convolutional Neural Network. The convolution layer convolves with convolution filter to produce feature maps which are propagated to the pooling layer. Pooling layer keep useful feature and discard the rest. These maps pass through each layer in the network until it reaches the final layer which is usually preferred to be Softmax layer. Softmax layer outputs the probabilistic distribution of data based upon the input feature maps. Figure 3.5 shows details of our proposed ResNet18 architecture. It can be seen in the figure that two pooling layers are used first one is max pooling and second one is average pooling.

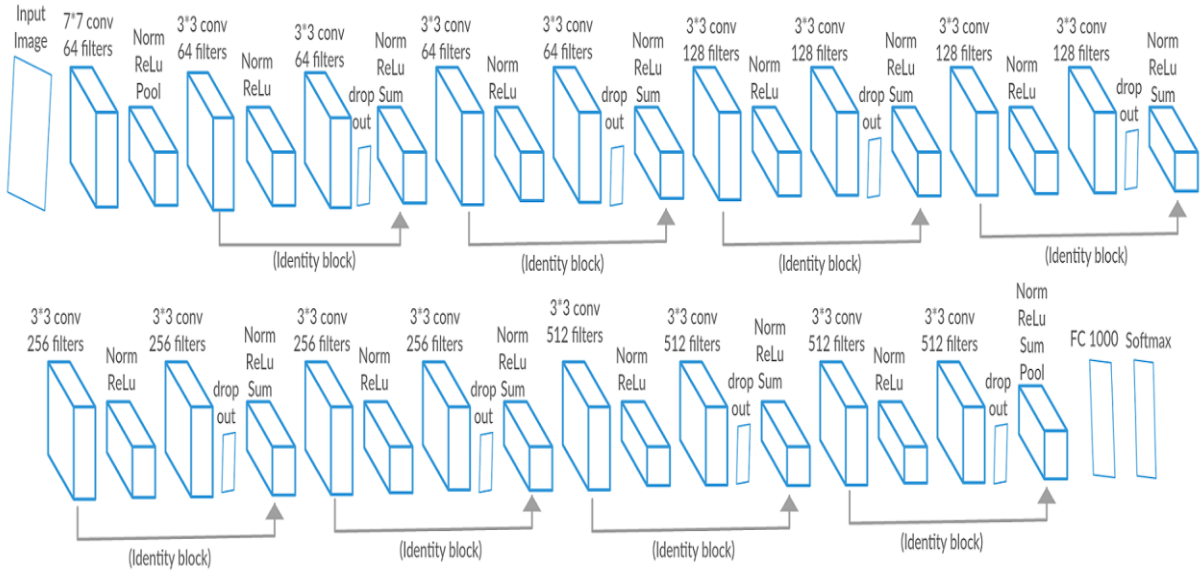

Figure 3. 5 Our Proposed ResnNet18 architecture for glioma brain tumor segmentation. In this diagram two pooling layers are used first one is max pooling and second one is average pooling.

A ResNet [1] architecture consists of stacked layers on top of each other with each combination of layers called a module. Each module consists of convolution layer, batch normalization and activation layer. Furthermore, it contains a shortcut connection which skips multiple layers and its output is added to the output of stacked layers. Typically, a module in ResNet architecture takes input X and produce f(x) through a series of convolutions, batch normalization and ReLu layers. The resultant f(x) gives F'(x)=f(x)+X, when added with the input x. This residual mapping is shown in figure given below. Figure 3.6 shows identity block of typical ResNet architecture.

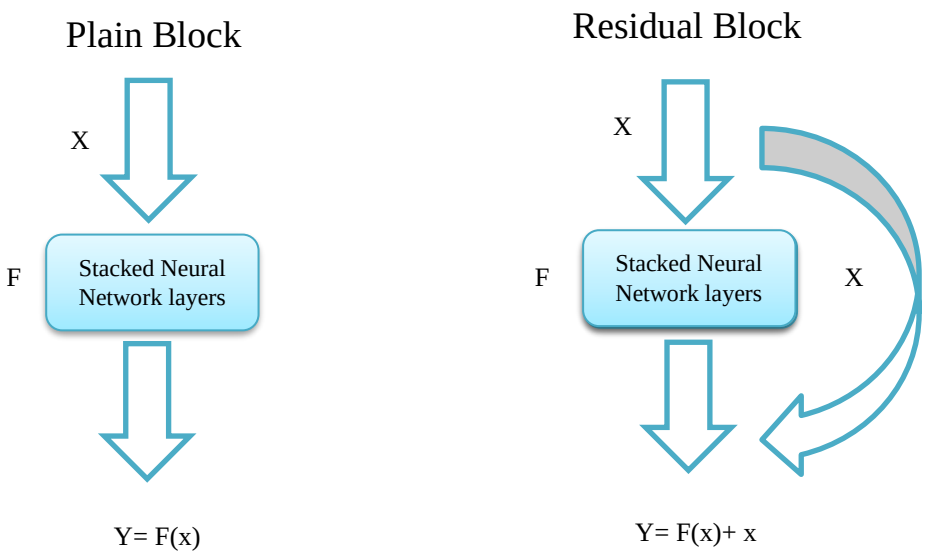

Figure 3. 6 Identity mapping in ResNet Architecture

We have added a dropout layer after every other convolution layer in our proposed ResNet architecture. Due to this addition each module generates better output and increased regularization. Dropout layer is used by many architectures in order to prevent overfitting. Some researcher use dropout as a substitute of batch normalization, however, few studies [45] proposed that Dropout performs well with batch normalization in generalizing the output. Detailed Diagram of our modified ResNet18 architecture is given in figure 2. Last layer was kept Softmax layer as default. Table 3.1 shows the Layers details of our modified ResNet18 model.

Table 3. 1 Layers detail of our ResNet18 model

Layer	**Output Size**	**Layer detail**
Conv1_x	17 × 17	7 × 7, 64, S=2
Conv2_x	9 × 9	3 × 3 max-pool, S=2
		3 × 3, 64
Conv3_x	5 × 5	3 × 3, 128
Conv4_x	3 × 3	3 × 3, 256
Conv 5_x	2 × 2	3 × 3, 512
	1 × 1	Avg. Pool, 1000-d, FC, Softmax
# Params.		1.1×10^7

3.7 Training

The Convolutional Neural Network training is required to increase the correct training probabilities of labels throughout the dataset and minimizing the loss function. The goal is to maximize the Softmax layer probabilities for the true label of each training patch in the entire dataset. Various loss function optimizers were tested on the network for maximizing the training performance and reducing the time overhead in the training. Figure 3.7 shows randomly selected kernels at various point throughout both of proposed VGG-19 and modified ResNet architecture.

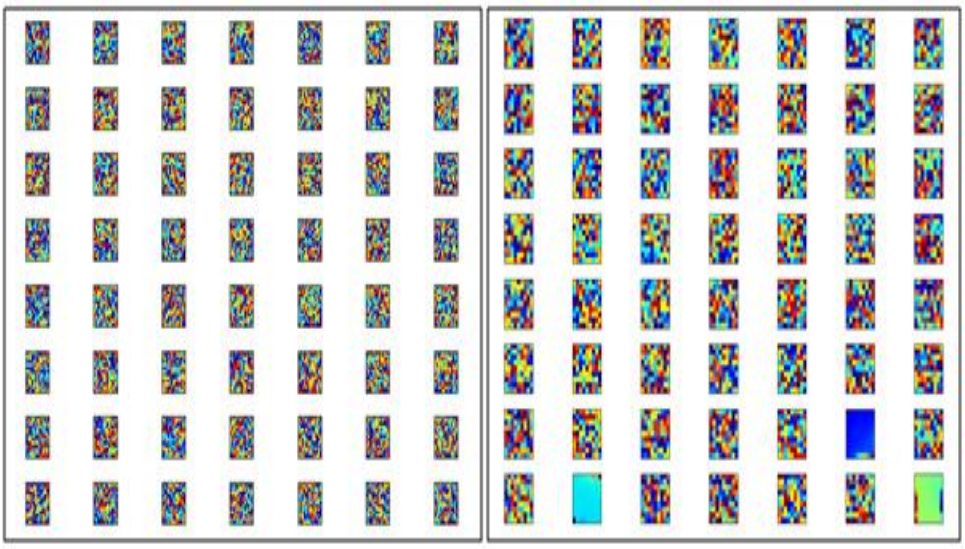

Figure 3. 7 randomly selected kernels at various layers throughout both of proposed VGG-19 and modified ResNet architecture

3.8 Loss Function

Loss function is the essential element in Neural network as well as in Convolutional Neural Network. They are described as a parameter against which the performance of the network is measured. We choose mean square loss function in our research because we needed pixel buy pixel information.

3.8.1 Loss function Optimizers

We trained network with three loss function optimizer to finetune of our result.

- First, we used stochastic gradient descent (SGD) which is generally considered as default optimizer.
- Then we further experimented our network on Adam optimizer.

- Finally, Root Mean Square Propagation (RMSProp) optimizers to were tested to check its effect on finding the global minimum point of loss function.

We find out that Adam optimizers not only outperformed SGD and RMSProp but it is also faster and memory efficient. Figure 3.8 shows various kernels at last layer of our proposed VGG19 and Modified ResNet architecture.

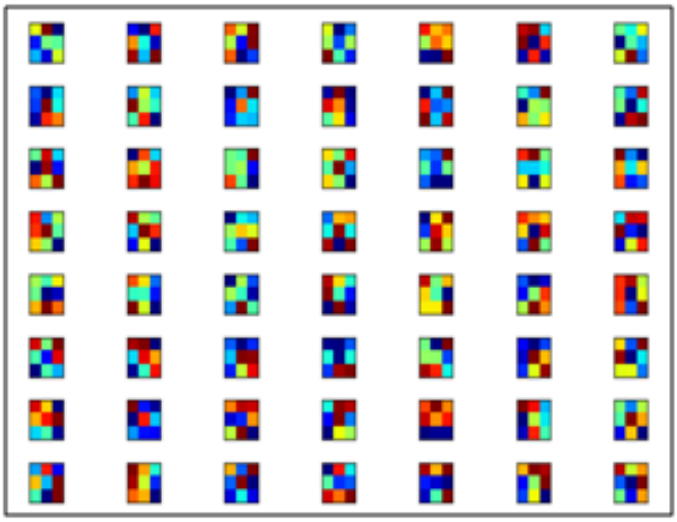

Figure 3. 8 Kernel selected randomly at last layer of ResNet Architecture

3.9 Two-phase Training

Glioma tumor segmentation is highly class imbalance problem, as most of the brain image tissues belongs to a healthy class. The model should train significant number of patches from each class in order to effectively learn data. With class imbalance problem the training process become difficult as it creates biasness towards healthy classes. Whereas if we train the network with equal number of classes then it creates biasness toward the turmeric classes. Therefore, we trained network both with actual ratio of patches and then their true ratio of patches.

In two phase model first, we trained the data using equal number of patches from each class in the dataset. During second phase, patches are made randomly throughout the training thus 98% of the patches belongs to the healthy class whereas only 2% patches represents the four turmeric classes i.e. edema, necrotic, enhancing and non-enhancing tumor.

Weighted training method is used during second phased in which each class is assigned a weight based on their true distribution in the dataset. In this way all the classes have equal effect on network training. During first phase the network is trained on 10 hundred thousand patches on 70 epochs.

Then using the previous training outcomes, we again trained the network with true distribution of patches from each class. During second phase training network is trained on two hundred thousand patches for 25 epochs only for the output layer. Only the output layer is trained for true distribution of patches in the dataset. Therefore, network learns through true class probabilities making it more effective for segmentation process. Figure 3.9 shows block diagram of our 2-phase network training model.

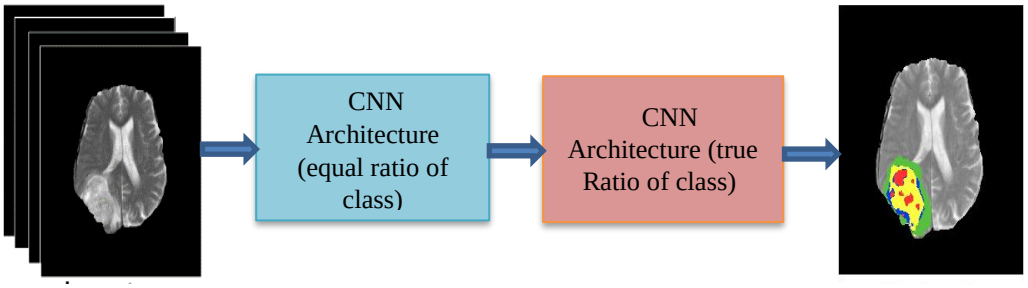

Figure 3. 9 Shows the propose design of two-phase training Architecture

Chapter 4 Experimental Results

4.1 BRATS 2015 Dataset

We performed our experimentation on BRATS 2015 [29] dataset which is used as a benchmark dataset for evaluation of Brain Glioma tumor segmentation. The training dataset contains MRI images of 220 HGG and 54 LGG patients while test data contain 110 images. There are no ground truth (label) images for the test data. The dataset contains four modality images of each patient i.e. T1, T1C, T2 and Flair MRI with a ground truth image. Ground truth images shown in figure 1 and 5 contain 5 classes i.e. Healthy image, necrotic core (red), edema(green), non-enhancing (blue) and enhancing (yellow) tumor labelled as 0,1,2,3 and 4 respectively.

4.2 Experimentation

We experimented glioma brain tumor segmentation using various architecture including VGG19 and proposed modified ResNet architecture. We also used a two-phase training model to reduce the effect of class imbalance problem as most patches belongs to class 0 and 1. In two phase we first trained the model using their true ratio in the dataset. Then using the previous training outcomes, we again trained the network with equal number of patches from each class. On experimenting we saw that network showed 1% increase in performance using this technique. Table 2 compares performance of our various implemented methods.

4.2.1 VGG-19 Architecture

We chose VGG19 as our baseline models to start the segmentation efforts. Although VGG is a little older architecture but we modified it a little bit in a sense that we used latest trend in CNN i.e. Regularization layer, hyperparameter tuning and two-phase training in order to enhance the segmentation results.

The experimentation is performed on 220 HHG and 54 LGG glioma brain tumor images. The dataset is actually in 3-D form and BRATS dataset lacks resolution in the third dimension therefore we transformed the 3D data into 2-D image dataset before training process. The network is trained on ten hundred thousand patches for 50 epochs. We also saw implement the our modified VGG architecture on two phase training which is explained in later.

The proposed model performed well in segmenting the glioma tumor. It has been seen that by using dropout layer network were able to train better. As it significantly reduced the overfitting in the training. The results of VGG architecture is shown in the table 4.1.

4.2.2 Proposed Modified ResNet Architecture

Residual Networks tends to be deeper than any other available networks. They tend to be as deep as 1000 layers. The main idea behind training residual connection architecture is utilizing the multiple paths in the network. Same concept is utilized by DensNet and Google Inception architecture. This implies that sure the 'deeper is better' but it is even better for a network to be wider. Another benefit those 'shortcut connection' gives is the ability to train much deeper network without worrying about the degradation problem. Figure 4.1 shows flow chart diagram of our proposed framework.

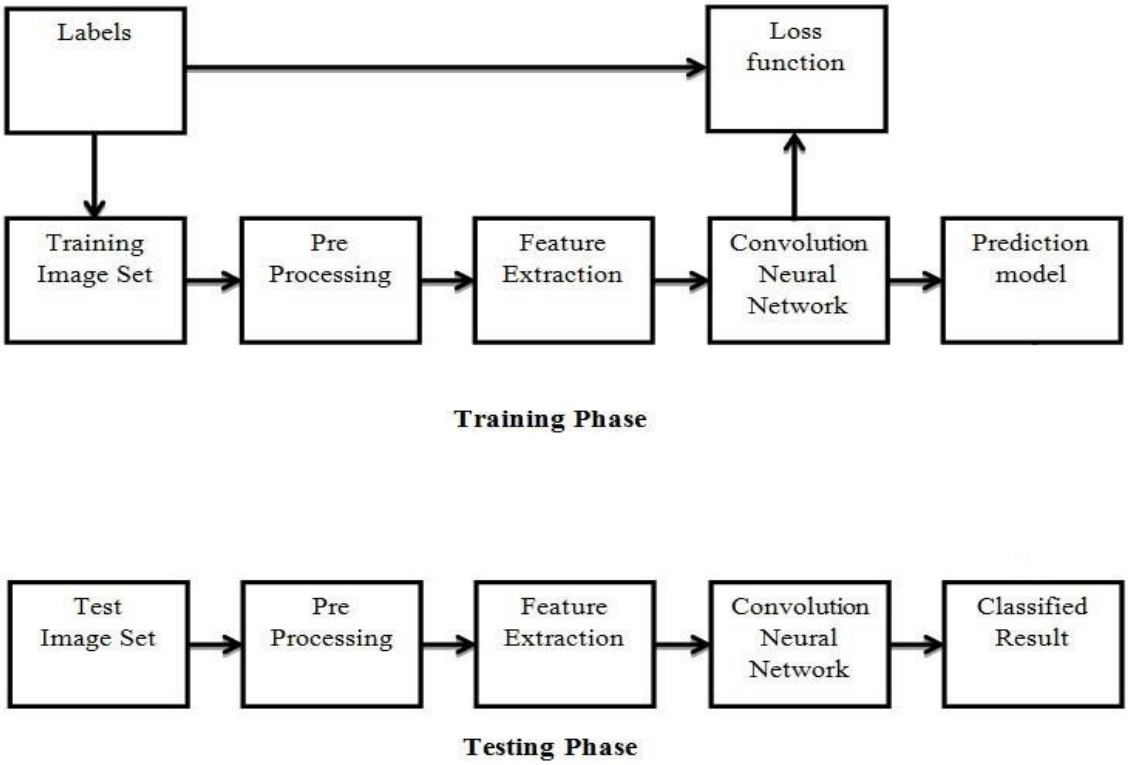

Figure 4. 1 Flowchart of our proposed Framework

It has been evident from the results that our proposed ResNet Architecture works well as it learns good local as well as global feature during the training. We trained our proposed network on 70 epochs for 10 hundred thousand training patches.

We observed that after 70 epochs there is no benefit of training and the network had already achieved convergence. Table 2 gives detailed comparison of our proposed models with the state-of-the-art techniques of BRATS 2015 Dataset.

4.2.3 Two phase Training

Glioma tumor segmentation datasets has unequal class distribution, as most of the brain image pixels comprise healthy tissues. The model should train significant number of patches from each class in order to effectively learn data. This is not true under default one phase training process. Therefore, we employed two phase training to train our modified VGG and ResNet network.

We saw too as the feature maps increases the chances of overfitting increases significantly, while if the feature is greatly reduced it causes underfitting in the network. Therefore, dropout value of 0.4 is adjusted to drop weak feature in the training.

It has been noted that fully connected layer takes most of the time in network training therefore an optimal size of 1024 fully connected (FC) feature was used to achieve the stability. We saw improvement in performance by using two phase training method for both VGG and ResNet architecture.

The comparison of results of training with default and two-phase network is shown in table 4.1. While figure 4.2 shows various components i.e. pre-processing, post-processing, patch extraction, both training phases, classification and evaluation of our proposed framework for the problem of Glioma Brain tumor segmentation.

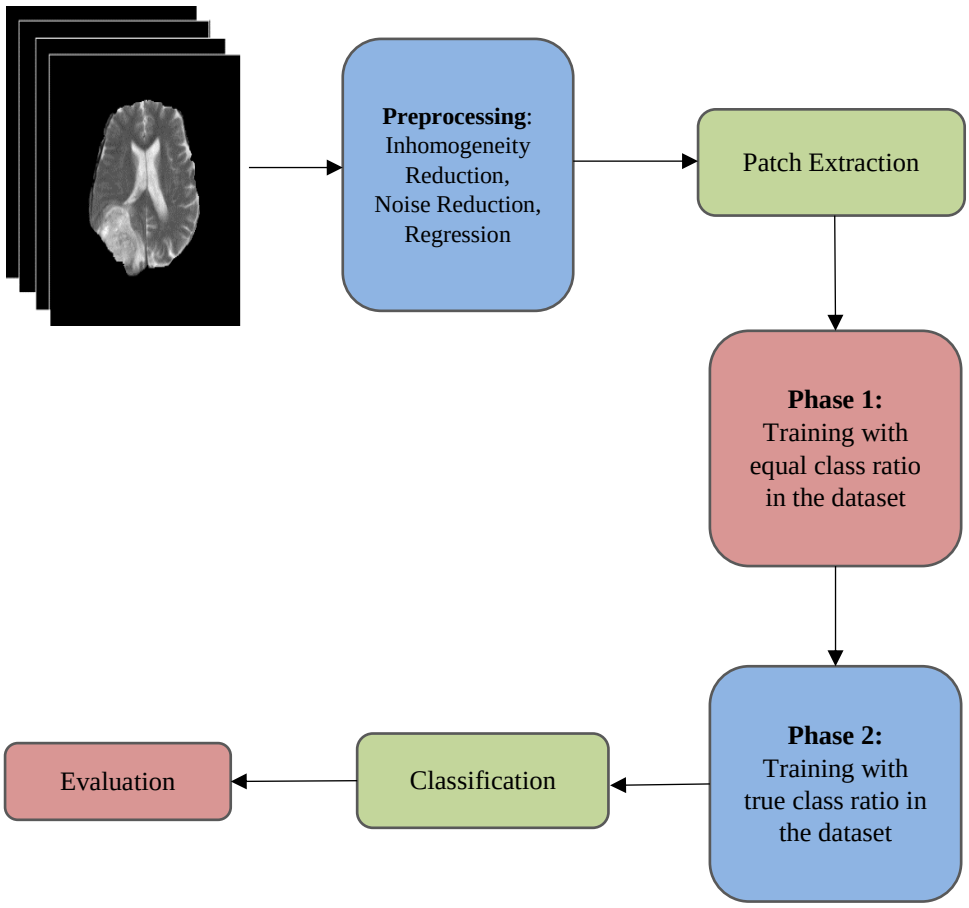

Figure 4. 2 Flow chart of two-phase training

4.3 Testing

There are 110 test images provided in BRATS 2015 dataset. Before network is provided with the test image same image processing technique is applied on the images. The 3-D test data is converted into 2D and then preprocessed in order to perform testing. Same pre-processing steps Intensity Normalization and N4ITK bias field correction techniques are applied. The testing process take 4-5 minutes to generate results on our machine. We have evaluated the test results on the basis of three parameters i.e. Dice Coefficient, Sensitivity and Specificity. For further comparison three type of tumor sub region are evaluated for performance which are Complete (Complete tumor include all tumor classes i.e. 1,2,3 and 4), Core (Include the tumor from the class 1) and Advancing tumor (Include the tumor from the class 4).

Table 4. 1 Effect of Two-phase Network Training on BRATS 2015 dataset

Network	Dice Score	Sensitivity	Specificity
Plain VGG	0.84	0.85	0.81
Two-phase VGG	0.85	0.83	0.84
ResNet18	0.88	0.89	0.88
2-phase ResNet18	0.90	0.90	0.89
ResNet50	0.89	0.90	0.89
2-phase ResNet50	**0.90**	**0.91**	**0.90**

4.4 Hardware Specification

The above experimentation was performed using a NVIDIA 1080Ti GPU with 16 GB of RAM on an Intel core i7 processor. In 2007 NVIDIA started production of Compute Unified Device Architecture (CUDA) in order to facilitate programming for their NVIDIA graphical Processing Unit (GPUs). Since then NVIDIA GPUs has been top on the market with high performance GPU specifically designed for larger amount of matrix multiplication which can execute in parallel as each NVIDIA graphical processing unit contain huge number of parallel cores.

With the advent of high-performance graphical processing unit, the field of machine learning gained a significant performance elevation since deep learning techniques requires huge amount of matrix multiplication and those GPUs facilized it. These availabilities of high performing GPUs and their CUDA API were critical for ImageNet [46]. As before the emergence of theses GPUs graphic performance was strictly dependent on the CPUs power. The performance power of the earlier models of NVIDIA GPUs was as much as 70x higher than the standard personal computers. Since the launch of the earlier model of NVIDIA GPUs in 2012 both memory capacity and number of cores of the NVIDIA graphic cards has increased significantly. Table 4.2 compares of results with state-of-the-art methods on BRATS 2015 dataset.

Table 4. 2 A comparison of various state-of-the-art architectures with our top performing models in term of Dice Score, Specificity and Sensitivity on BRATS 2015 dataset

Model	Dice			Sensitivity			Specificity		
	Complete	Core	Enhancing	Complete	Core	Enhancing	Complete	Core	Enhancing
Havaei et al.	0.85	0.78	0.73	0.93	0.80	0.72	0.80	0.76	0.75
Konstantinos et al.	0.90	0.75	0.72	0.89	0.71	0.74	0.92	0.85	0.75
Pereira et al.	0.88	0.76	0.73	0.91	0.90	0.72	0.86	0.74	0.81
V. Shreyas et al.	0.83	0.75	0.72	0.79	0.74	0.76	0.89	0.78	0.73
Di Liu et al.	0.87	0.62	0.68	0.92	0.65	0.86	0.80	0.62	0.60
VGG	0.86	0.79	0.80	0.87	0.83	0.78	0.89	0.78	0.80
ResNet18	0.88	0.85	0.83	0.89	0.84	0.75	0.88	0.83	0.81
ResNet50	**0.90**	**0.87**	**0.84**	**0.91**	**0.84**	**0.81**	**0.90**	**0.84**	**0.82**

4.5 Neural Network Hyperparameters

We implemented our framework on ubuntu machine with Keras [47] library which uses tensor flow as back-end. Four images of 33*33 pixel are passed through the network as input. Batch size of 64 and 128 were used during training.

We trained the model till the convergence was achieved. We used max and average pooling layer in our model. While max pooling is used through the both network while in proposed ResNet architecture the last pooling layer was used to be global average pooling. The convolution and pooling stride size are kept as 2.

4.5.1 Neuronal Activation

An activation function is used to control the output values of neuron after each layer. There are many activation layers been proposed by the researcher to control the value of neuron i.e. Linear, Tangent (tanh), sigmoid Rectified Linear unit (ReLU) and leaky ReLU and max-out. The tangent activation function output a value between [-1,1]. ReLU activation function is a nonlinear activation function that outputs either a zero or a positive value. Given input values a ReLU activation function outputs maximum of 0 and input value as:

$$f(s) = \max(0, s) \tag{4.1}$$

Compared to ReLU activation function a leaky ReLU activation function out a small value when input is smaller than zero.

Max-out activation function takes feature map as input and outputs the maximum feature value. Given the input set of feature X the max-out activation function goes through all spatial location of the feature map and generate the output as:

$$Y_i = \max\ (S_1, S_2, S_3 \dots \dots \dots S_i) \quad (4.2)$$

We analyzed various activation unit in order to enhance the network outcome while ReLU activation function performed best for the problem at hand.

4.5.2 Normalization

We used Batch Normalization (BN) [48] to normalize each layer through the input data. Batch normalization uses an activation function in order to control the feature's mean and standard deviation close to 0 and 1, respectively. Sometimes layers weights change significantly due to large learning rate, causing the small changes to amplify results in bad training. BN control the gradient in acceptable range during back-propagation.

A feature map M is normalized as:

$$M = Rnl\ (Bn\ (w_i)) \quad (4.3)$$

Where nrl represents ReLU non-linearity, Bn represents Batch Normalization and w_i represent weight parameters.

4.5.3 Regularizer

A regularization layer is used in order to reduce overfitting od data by applying punishment on layer weight during training [49]. A dropout layer [50] is usually used as regularization unit after every convolution layer throughout the network. The dropout layers functions by removing certain amount of network weights during training. A vector containing random variables r_i is created and each element of vector has a certain probability of being equal to one. This vector V_i is multiplied with the output O_i feature map of each layer jo generates a sparse output S_i. The resultant feature maps are used as input to the upcoming layer. The formula for calculation dropout function is given below:

$$S_i = V_i * O_i \quad (4.4)$$

A Dropout value of 0.4 is adjusted to drop weak features and reduce overfitting. A dropout value of 0.3 is used before the final layer of the network in order to generate a strong feature value-based output probability.

4.5.4 Loss Function

A Loss function is the essential element in Neural network as well as in Convolutional Neural Network. They are described as a parameter against which the performance of the network is measured. A We choose mean square loss function in our research because we needed pixel buy pixel information. The loss function L_a can be calculated as:

$$L_a = -\frac{1}{B}\sum_{a=1}^{B} \log(P(Y_a = y_a)) \quad (4.5)$$

Where Y_a is the label of the target class, y_a is the predicted label of the class and B is the given batch of the data.

4.5.5 Loss function Optimizers

An optimizer computes the loss function value at the output of the network and these updated values are propagated back to the network in order to optimize the training. We analyzed various optimizer in our network to accelerate. the network training

We trained network with three loss function optimizer to finetune of our result.

- First, we used stochastic gradient descent (SGD) which is generally considered as default optimizer.
- Then we further experimented our network on Adam optimizer.
- Finally, Root Mean Square Propagation (RMSProp) optimizers to were tested to check its effect on finding the global minimum point of loss function.

We find out that Adam optimizers not only outperformed SGD and RMSProp but it is also faster and memory efficient.

Figure 4.3 shows training accuracy on our top model while fine tuning with various optimizers for better training outcomes.

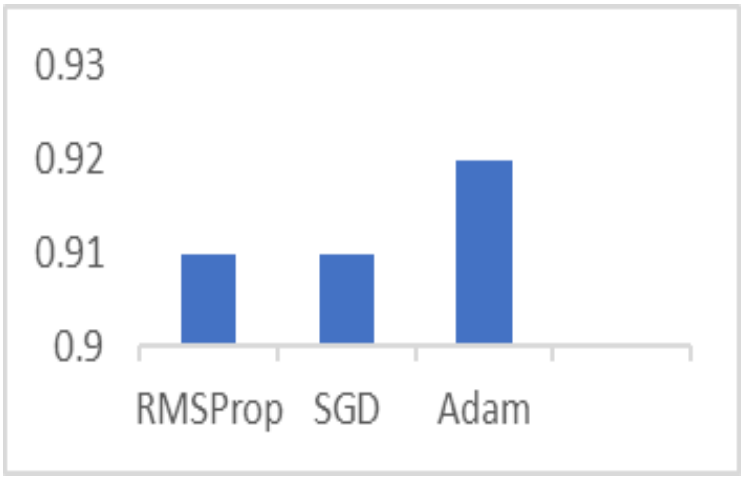

Figure 4. 3 Training accuracy on our top model while fine tuning with various optimizers

Figure 4.4 shows various stages of Hyperparameter selection for our propose framework for glioma brain tumor segmentation.

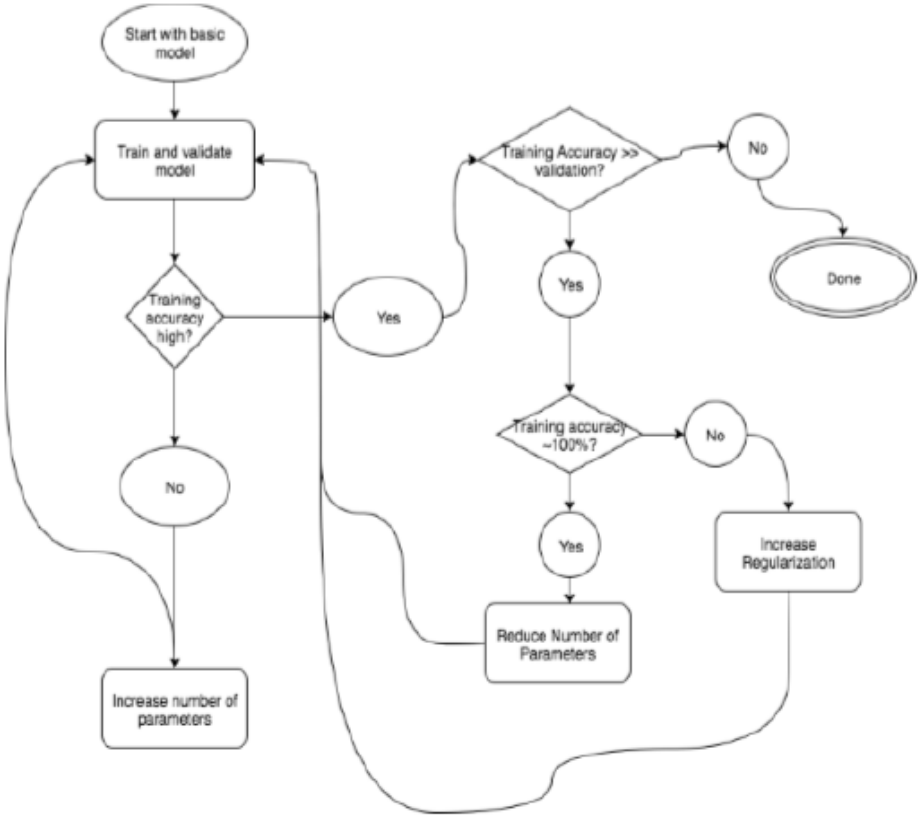

Figure 4. 4 Hyperparameter selection flowchart

We did further experimentation to see the effect of transfer learning on our proposed architecture. Transfer learning can perform well where a number of training examples are not adequate. Transfer learning is a technique in which network learn some information from one

problem and use that knowledge to solve another problem. Some architectures ResNet Inception-V3 [51] and DensNet are most often used for fine-tuning. We performed transfer learning on our modified ResNet18 and Resnet50 architecture. We saw an improvement in the classification accuracy by using pre-trained weights on our dataset. Using ResNet architecture our network achieved the best Dice Score Coefficient value of 0.90. Table below shows comparison of our model with best performing model in term of Dice score, Sensitivity and Specificity on BRATS 2015 dataset.

Table 4. 3 comparison of our model with best performing model in term of Dice score, Sensitivity and Specificity on BRATS 2015 dataset

Model	**Dice Score**	**Sensitivity**	**Specificity**
Havaei et al.	0.88	0 .91	0.87
Konstantinos et al.	0.90	0.92	0.89
Di et al.	0.87	0.83	0.90
Sérgio et al.	0.88	0.86	0.91
Proposed	**0.90**	**0.91**	**0.90**

4.6 Evaluation Parameters

Segmentation results are commonly evaluated using three metrices i.e. Dice score, sensitivity and specificity [52]. Dice Score is given as:

$$DSC = 2 \times \frac{(|L \cap P|)}{|L| + |P|} \tag{4.6}$$

Where L and P stands for actual label for tumor region and predicted label for tumor region respectively. Sensitivity is measure of tumor regions which are correctly classified. Its formula is given as:

$$\text{Sensitivity} = \frac{|P \cap L|}{|L|} \tag{4.7}$$

Where P and L stands for predicted and actual regions respectively.

Third and last metric is Specificity. It is the measure of healthy pixels that are classified as healthy. It is given as:

$$\text{Specificity} = \frac{|P_0 \cap L_0|}{|L_0|} \tag{4.8}$$

Where P_0 and L_0 stands for predicted and actual healthy tissues respectively.

4.7 Results Comparison

We have compared the results of our study with various state-of-the-art techniques on BRATS 2015 dataset in table 2. Our proposed methodology outperforms the presents stat-of-the-art method in term of dice score for all three-tumor region i.e. complete, core enhancing. The increase in the dice score values is the main highlight of this research work.

Label 2 is edema class in our proposed model which has the highest ratio in all of the tumor sub region therefore, it affects the most among other classes in computing whole tumor results. As label 2 is not present in the results of core and enhancing tumor therefore network gained good performance for these regions.

Our best performing model is the proposed ResNet 50 architecture. It is to be noted here that increasing number of layers can improve the network training outcomes but there are several other factors that are involves in it. Usually deeper networks work better when the training dataset is very large. However, gathering such large amount of data is very difficult for medical imaging problem. Due to this reason we limited our experimentation to only as deep as 50-layer architectures i.e. ResNet50 architecture.

Another significance of this research is the implementation time. While the architecture reported in [53] takes 100 minutes to segment a brain our model on the other hand propose model can do this task in 4-5 minutes using a standard person computer PC. This time reduction is of significant importance in medical field along with the performance improvement. Figure 4.5 shows results of our model on BRATS 2015 dataset.

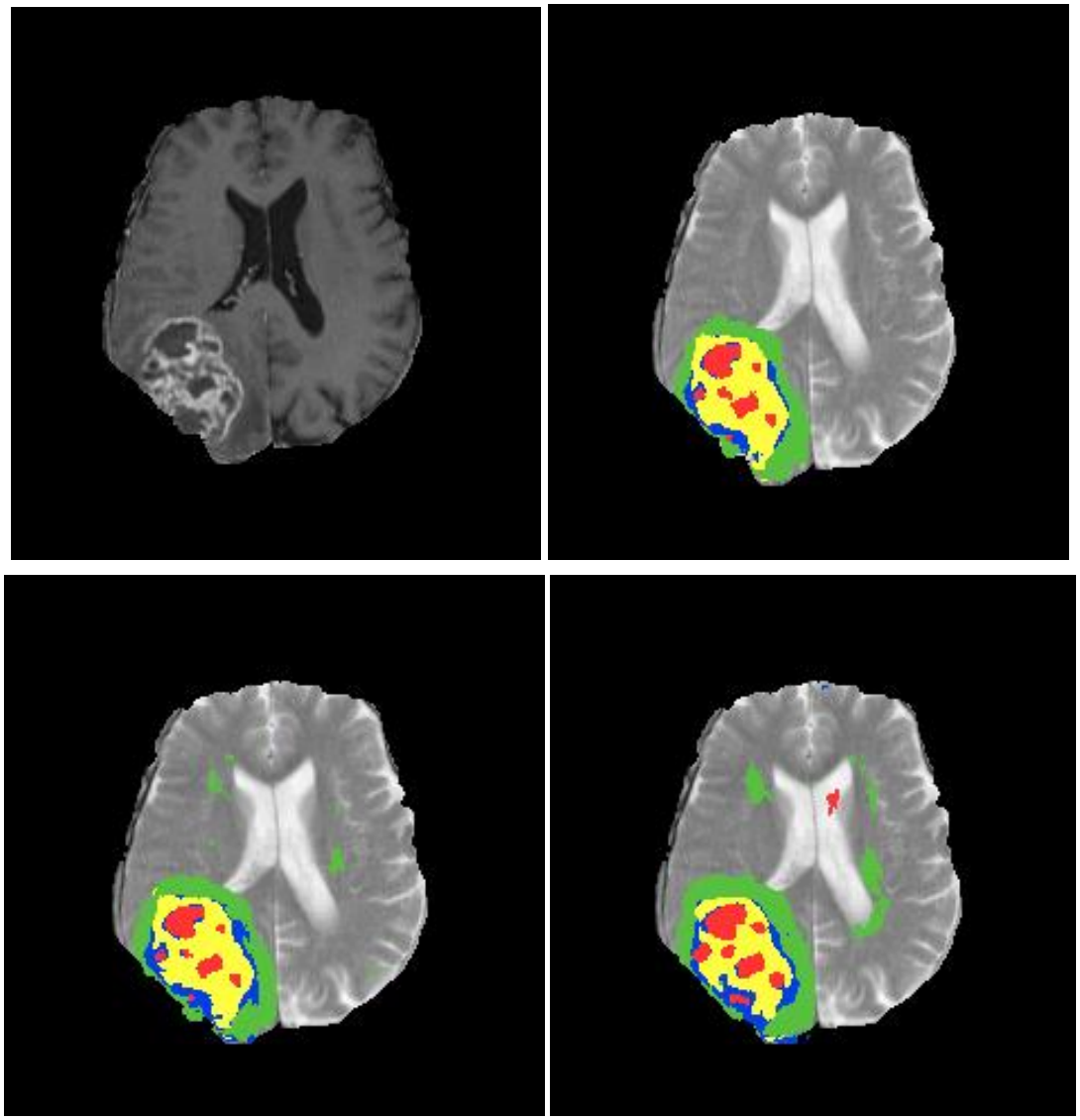

Figure 4. 5 Segmentation results on BRATS 2015 dataset using modified ResNet models. Top Row: On left is Image modality and on right is ground truth image. Bottom Row: On left is segmentation results of our Proposed modified ResNet Architecture while on right is segmentation result of VGG architecture.

Chapter 5 Conclusion and Future Work

5.1 Conclusion

In this research, we have proposed a novel framework for glioma tumor segmentation. Our framework consists of ResNet architecture which is based on Convolutional Neural Network. We used a pre-processing step for bias field reduction on our dataset. We added dropout layer after second convolution layer to reduce overfitting. Experiment results shows that our network outperforms the current state-of-the-art architectures in term of dice score. We used BRATS 2015 dataset for our experimentations on which ResNet50 with achieved best performance with dice score value of 0.90.

We gave a comprehensive review of most state-of-the-art segmentation methods for glioma tumor segmentation and classifies all the techniques into four-sub types i.e. Deep Learning, Conventional Image Processing, Clustering and Conventional Machine learning methods. We have also briefly discussed some most used publicly available datasets for brain segmentation. We also gave a brief analysis of the performance of various methods and techniques on the basis of dice score which is shown in table 4. We have seen that deep learning algorithms can give very good results but they are slow in processing time and needs large training datasets.

We have enlisted many performance measuring matrices that are present these days to the measure performance of segmentation. It is evident that the segmentation of glioma brain tumor is one the most difficult task in image processing due to the irregular shape and varying size and location of these tumors. Low brightness MRI images are also one of the major hurdles in this segmentation problem. However, in recent times with the evolution of neural networks and the availability of more publicly available dataset, a large number of researches are been carried out resulting in a reasonable efficiency.

Another significance of this research is the implementation time While many architectures takes up to 100 minutes to segment a complete brain our model on the other hand propose model can do this task in 4-5 minutes using a standard person computer PC. This time reduction is of significant importance in medical field along with the performance improvement.

5.2 Future Work

For the feature purpose there are many advancements which can be made to enhance the Glioma Brain tumor segmentation process.

5.2.1 Dataset

Most automatic segmentation methods have promising results in tumor segmentation and analysis, however, further improvement in these algorithms and availability of additional image information from new image modalities may improve these methods and prove to be helpful in the development of large-scale, clinically acceptable method for better tumor segmentation. The Deep Convolutional Neural Network performs better with larger dataset therefore in future there should be the addition of more training examples for the purposed of network training. Another crucial aspect in developing a brain-tumor segmentation model is its robustness, meaning algorithm should be able to work on different datasets. Therefore, researchers should need to test their algorithm on different datasets with different modalities in order to design a robust brain tumor segmentation model.

5.2.2 Computation Time

Researchers should more emphasize on the computation time as current most state of the art algorithm i.e. latest CNN architectures has a computation time of few minutes which is unacceptable in developing a real time CAD system. It may be difficult to achieve a real time computation time however, the improvements in this domain is still open for the future researcher.

5.2.3 Transfer Learning

Transfer learning is a technique in which network learn some information from one problem and use that knowledge to solve another problem. In experimentation, we explored transfer learning across entire network training. However, in literature there are many other transfer learning paradigms which are known to provide great results. A good change in transfer learning mechanism could be to freeze the last layer of the network during fine-tuning. In this way fine tuning will be performed on all the layer till the fully connected layer of the network. The benefit of this approach is the reduction of overfitting during fine tuning process which is also a problem in our network. Another improvement could seem to provide better results is applying transfer leaning on various medical dataset and then using those learning outcomes on the problem on hand. There are some barrier in applying transfer learning on Brats dataset. As BRATS dataset is a multimodality imaging dataset so network training includes for channels per input i.e. T1, T1C, T2 and Flair However in case of applying fine tuning most dataset has only three channels or one channel only in case of gray level images. There are many possible solutions to cater this problem. One solution is to simply discard a channel while

training on BRATS dataset to match the feature channels size. The research on this domain of Glioma Brain tumor segmentation is still open for new researchers.

Reference

1. K. He, X. Zhang, S. Ren, and J. Sun.: Deep Residual Learning for Image Recognition, Jan. 2016J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
2. I G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger.: Densely Connected Convolutional Networks, Jan. 2017.
3. C. Szegedy.: Going deeper with convolutions, Jan. 2015.
4. Brain Cancer Statistics and Brain Cancer Awareness." Tocagen, tocagen.com/patients/brain-cancer/brain-cancer-statistics.
5. 7T., Benefits, M., & of, B. (2017, June 27). The 7 Most Meaningful Benefits Of Meditation. Max Lugavere.
6. Ali I and Cem Direko. "Review of MRI-based brain tumor image segmentation using deep learning methods." (2016).
7. Rao, B. Srinivasa and Dr. E. Sreenivasa Reddy. "A survey on Glioblastoma Multiforme Tumor Segmentation through MR images." (2016).
8. Albelwi, S.; Mahmood, A. A Framework for Designing the Architectures of Deep Convolutional Neural Networks. *Entropy 19*, 242,2017.
9. Lu, W., Miklau, G., & Gupta, V. Generating private synthetic databases for untrusted system evaluation. (2014).
10. N. Sharma and L. M. Aggarwal, Automated medical image segmentation techniques, Journal of Medical Physics/Association of Medical Physicists of India, vol. 35, no. 1, p. 3, 2010.
11. Dubey, Rash & Hanmandlu, Madasu & Gupta, S.K. (2009). Semi-automatic Segmentation of MRI Brain Tumor. ICGST-GVIP Journal. 9.
12. D. L. Pham, C. Xu, and J. L. Prince, Current methods in medical image segmentation 1, Annual Review of Biomedical Engineering, vol. 2, no. 1, pp. 315-337, 2000
13. Huber, T, et al. "Reliability of Semi-Automated Segmentations in Glioblastoma." Clinical Neuroradiology, vol. 27, no. 2, pp. 153–161,2015.
14. K. Sudharani, T. Sarma, and K. S. Prasad, "Advanced Morphological Technique for Automatic Brain Tumor Detection and Evaluation of Statistical Parameters", vol. 24, pp. 1374-1387, Jan. 2016.Top of Form

15. I. Zabir, S. Paul, M. A. Rayhan, T. Sarker, S. A. Fattah, and C. Shahnaz, “Automatic brain tumor detection and segmentation from multi-modal MRI images based on region growing and level set evolution”, Jan. 2015
16. N. J. Tustison, “Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with ANTsR”, vol. 13, no. 2, pp. 209-225, Jan. 2015
17. M. Soltaninejad, “Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI”, vol. 12, no. 2, pp. 183-203, Jan. 2017.
18. S. Amiri, I. Rekik, and M. A. Mahjoub, “Deep random forest-based learning transfer to SVM for brain tumor segmentation”, Jan. 2016
19. E. Abdel-Maksoud, M. Elmogy, and R. Al-Awadi, “Brain tumor segmentation based on a hybrid clustering technique”, vol. 16, no. 1, pp. 71-81, Jan. 2015.
20. J. S. Cordova, “Quantitative Tumor Segmentation for Evaluation of Extent of Glioblastoma Resection to Facilitate Multisite Clinical Trials”, vol. 7, no. 1, pp. 40-W5, Jan. 2014
21. https://medium.com/@sidereal. (2017, November 16). CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more …. Medium. Retrieved February 27, 2019, from https://medium.com/@sidereal/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5
22. Havaei, Mohammad, et al. “Brain Tumor Segmentation with Deep Neural Networks.” Medical Image Analysis, vol. 35, 2017, pp. 18–31.
23. Hussain, Saddam, et al. “Segmentation of Glioma Tumors in Brain Using Deep Convolutional Neural Network.” Neurocomputing, vol. 282, 2018, pp. 248–261.
24. Kamnitsas, Konstantinos, et al. “Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation.” Medical Image Analysis, vol. 36, 2017, pp. 61–78.
25. D. Liu, H. Zhang, M. Zhao, X. Yu, S. Yao and W. Zhou, "Brain Tumor Segmention Based on Dilated Convolution Refine Networks," 2018 IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA), Kunming, China, 2018, pp. 113-120.
26. 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

27. R.K.-S. Kwan, A.C. Evans, G.B. Pike: “MRI simulation-based evaluation of image-processing and classification methods” IEEE Transactions on Medical Imaging. 18(11):1085–97, Nov 1999.
28. Valverde, Sergi et al. “Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations.” Journal of magnetic resonance imaging: JMRI 41 1 (2015): 93-101.
29. Menze, Bjoern H., et al. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS).” IEEE Transactions on Medical Imaging, vol. 34, no. 10, 2015, pp. 1993–2024., doi:10.1109/tmi.2014.2377694.
30. Liew, Sook-Lei, et al. “A Large, Open Source Dataset of Stroke Anatomical Brain Images and Manual Lesion Segmentations.” Scientific Data, vol. 5, 2018, p. 180011., doi:10.1038/sdata.2018.11
31. O. Commowick, F. Cervenansky, R. Ameli, MSSEG Challenge Proceedings: Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure (2016).
32. I. Iˇsgum, M. J. Benders, B. Avants, M. J. Cardoso, S. J. Counsell, E. F. Gomez, L. Gui, P. S. H″uppi, K. J. Kersbergen, A. Makropoulos, et al., Evaluation of automatic neonatal brain segmentation algorithms: the NeoBrainS12 challenge, Medical image analysis 20 (1) (2015) 135–151
33. A. M. Mendrik, K. L. Vincken, H. J. Kuijf, M. Breeuwer, W. H. Bouvy, J. De Bresser, A. Alansary, M. De Bruijne, A. Carass, A. El-Baz, et al., MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans, Computational intelligence and neuroscience 2015
34. 2015B., BRATS, M., & 2017B. (1970, January 1). BraTS 2015 - MICCAI BRATS 2017. Retrieved February 27, 2019, from https://sites.google.com/site/braintumorsegmentation/home/brats2015.
35. J. Liu, “A Cascaded Deep Convolutional Neural Network for Joint Segmentation and Genotype Prediction of Brainstem Gliomas”, vol. 65, no. 9, pp. 1943-1952
36. I. Njeh, “3D multimodal MRI brain glioma tumor and edema segmentation: A graph cut distribution matching approach”, vol. 40, pp. 108-119, Jan. 2015. 2018
37. N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, et al., "N4ITK: improved N3 bias correction," IEEE transactions on medical imaging, vol. 29, pp. 1310-1320, 2010.
38. S. Pieper, B. Lorensen, W. Schroeder, and R. Kikinis, "The NA-MIC Kit: ITK, VTK, pipelines, grids and 3D slicer as an open platform for the medical image computing

community," in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, 2006, pp. 698-701

39. A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, *et al.*, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magnetic resonance imaging*, vol. 30, pp. 1323-1341, 2012.
40. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
41. S. Dieleman, K. W. Willett, and J. Dambre, "Rotation-invariant convolutional neural networks for galaxy morphology prediction," *Monthly notices of the royal astronomical society*, vol. 450, pp. 1441-1459, 2015
42. E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE transactions on pattern analysis and machine intelligence*, 2016.
43. I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. C. Courville, and Y. Bengio, "Maxout Networks," *ICML (3)*, vol. 28, pp. 1319-1327, 2013
44. G. W. Burr, R. M. Shelby, S. Sidler, C. Di Nolfo, J. Jang, I. Boybat, *et al.*, "Experimental demonstration and tolerancing of a large-scale neural network (165 000 synapses) using phase-change memory as the synaptic weight element," *IEEE Transactions on Electron Devices*, vol. 62, pp. 3498-3507, 2015
45. S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He.: Aggregated Residual Transformations for Deep Neural Networks, Jan. 2017.
46. C. K. D. Z. R. S. R. L. G. P. A. Schl¨ogl, "A fully automated correction method of EOG artifacts in EEG recordings.," Clin.Neurophys, pp. 118(1):98-104., 2007
47. P. Charles, Project title, 2013, (https://github.com/charlespwd/project-title
48. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015
49. N. H. G. K. A. S. I. &. S. R. Srivastava, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," Journal of Machine Learning Research, , vol. http://doi.org/10.1214/12AOS1000 , no. 15, p. 1929–1958., 2014.
50. N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929-1958, 2014.
51. https://medium.com/@sh.tsang. (2018, September 10). Review: Inception-v3—1st Runner Up (Image Classification) in ILSVRC 2015. Medium. Retrieved February 27, 2019, from

https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c.

52. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, pp. 1993-2024, 2015.
53. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, pp. 1993-2024, 2015.

ABBREVIATION

DCNN: Deep Convolutional Neural Network

BRATS: Brain Tumor Segmentation Challenge

MRI: Magnetic Resonance Imaging

CAD: Computer Aided Diagnosis

CNN: Convolutional Neural Network

GT: Ground Truth

LGG: Low Grade Glioma

HGG: High Grade Glioma

CT: Computed Tomography

PET: Position Emission Tomography

SPECT: Single Photon Emission Computed Tomography

NMR: Nuclear Magnetic Resonance

MRS: Magnetic Resonance Spectroscopy

EEG: Electroencephalography

WM: White Matter

SVM: Support Vector Machine

GM: Gray Matter

FCM: Fuzzy C-Means

KIFCM: called k-mean integrated with fuzzy C-mean clustering

IRT: Infrared thermography

DICOM: Digital Imaging and Communications in Medicine

ILSVRC: ImageNet Large Scale Visual Recognition Competition

N4ITK: Improved N3 Bias Correction

FLAIR: Fluid-attenuated inversion recovery

ISBR: Internet Brain Segmentation Repository

DCR: Dilated convolution refined

ROI: Region of Interest

TP: True Positive

TN: True Negative

FP: False Positive

FN: False Negative

ReLU: rectified linear unit

DNN: Deep Neural Network

CSF: Cerebrospinal Fluid

MEG: Magnetoencephalography

DF: Decision Forests

RF: Random Forests

SGD: Stochastic Gradient Descent

GPU: Graphical Processing Unit

DSC: Dice Similarity Coefficient

SR: Sparse Representation

ERT: extremely randomized trees

RDF: Random Decision Forests

CPU: Central Processing Unit

PCA: Principle Component Analysis

VGG: Visual Geometry Group

FC: Fully Connected

CUDA: Compute Unified Device Architecture

API: application program interface

BN: Batch Normalization

TL: Transfer Learning

PC: Personal Computer

DL: Deep Learning

GT: Ground Truth

SR: Sparse Representation

RDF: Random Decision Forests