

# Radiogenomic classification of glioblastoma using multimodal 3D MRI

A Project Report Submitted  
in Partial Fulfillment of Requirements  
for the Degree of  
**Bachelor of Technology**

by

Deepan Maitra (2019CSB1044)  
Shivani Kumari (2019CSB1120)



Department of Computer Science & Engineering  
Indian Institute of Technology Ropar  
Rupnagar 140001, India  
May 2023

# Abstract

O6-Methylguanine-DNA methyltransferase (MGMT) promoter methylation status is an important genetic characteristic of glioblastoma and is crucial for its diagnosis and chemotherapy efficacy. Multimodal MRI imaging techniques can contribute towards monitored automation of invasive surgical approaches. In this body of work, we propose an end-to-end pipeline of tumour segmentation and subsequent radiogenomic classification to classify the given MRI scans into being methylated or non-methylated. We develop a novel Multitask-learning based model (adapted from U-Net) to simultaneously perform segmentation and classification. Further, we utilise the segmentation results to cascade lightweight classification models based on several MRI slice sampling techniques to output the final classification scores. Our resultant pipeline performs well on each of the MRI axes, and several ensembles are tried out to arrive at suitable improvements. Based on 5-fold-cross-validation, we are able to surpass the current Kaggle leaderboard AUCs of the past BraTS RSNA-MICCAI Brain tumour challenge of 2021.

## **Acknowledgements**

We would like to thank RSNA and the BraTS Kaggle team for their decision to make the competition dataset publicly available for research purposes. Dr. Deepti R. Bathula has our deepest gratitude, without whom this project would not have taken shape. Finally, we are thankful to the HPC team at IIT Ropar, for their assistance in providing computational GPU resources.

## **Honor Code**

We certify that we have properly cited any material taken from other sources and have obtained permission for any copyrighted material included in this report. We take full responsibility for any code submitted as part of this project and the contents of this report.

Deepan Maitra (2019CSB1044)

Shivani Kumari (2019CSB1120)

## Certificate

It is certified that the B. Tech. project “**Radiogenomic classification of glioblastoma using multimodal 3D MRI**” has been done by **Deepan Maitra** (2019CSB1044), **Shivani Kumari**(2019CSB1120) under my supervision. This report has been submitted towards partial fulfillment of B. Tech. project requirements.

**Dr. Deepti R. Bathula**

Associate Professor

Project Supervisor

Department of Computer Science & Engineering

Indian Institute of Technology Ropar

Rupnagar-140001

# Contents

<b>Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>Nomenclature</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 MRI Imaging . . . . .	2
1.3 Deep learning approaches . . . . .	2
<b>2 BraTS 2021 challenge and Kaggle leaderboard</b>	<b>4</b>
2.1 Current Kaggle leaderboard . . . . .	5
2.2 Observations and Scope . . . . .	6
<b>3 Dataset and Experimental set-up</b>	<b>7</b>
3.1 Dataset . . . . .	7
3.1.1 Dataset description . . . . .	7
3.1.2 MRI modalities . . . . .	8
3.1.3 Number of samples . . . . .	10
3.1.4 Tumour segmentation regions . . . . .	11
3.1.5 Methylated and Non-methylated tumours . . . . .	12
3.2 Model architectures . . . . .	13
3.2.1 UNet 2D . . . . .	13

## CONTENTS

3.2.2	Multi-task-learning (MTL) model . . . . .	14
3.2.3	ResNet18 and ResNet34 . . . . .	16
3.2.4	EfficientNet . . . . .	16
<b>4</b>	<b>Methodology</b>	<b>18</b>
4.1	Multi-Task-Learning (MTL) model . . . . .	18
4.2	Volumetric Projections . . . . .	20
4.3	Cascaded mask-cropped model . . . . .	23
4.3.1	2D mask-cropped . . . . .	23
4.3.2	3D mask cropped . . . . .	25
4.3.3	Ensemble along 3 axes . . . . .	28
<b>5</b>	<b>Results and Discussion</b>	<b>29</b>
5.1	Segmentation results . . . . .	29
5.2	MTL joint segmentation-classification results . . . . .	31
5.3	Volumetric Projection results . . . . .	33
5.4	Cascaded cropped-mask-results . . . . .	34
5.5	3-axis Ensemble Results . . . . .	37
	<b>Conclusion</b>	<b>38</b>
	<b>Appendix A : Area Under Curve (AUC)</b>	<b>39</b>
	<b>References</b>	<b>40</b>

# List of Figures

3.1	Modalities across 3 axis. <b>A1-A3</b> depict T1-weighted images, <b>B1-B3</b> as T2-weighted, <b>C1-C3</b> as T1-contrast-enhanced and <b>D1-D3</b> as Flair. The first row is the <b>axial</b> direction, then <b>coronal</b> and <b>saggital</b> showing the various positions of the tumour . . . . .	9
3.2	Tumour regions (in the axial, coronal, saggital directions). The different colours depict the multiclass pixel labels . . . . .	11
3.3	(A) Unmethylated tumours and (B) Methylated tumours . . . . .	12
3.4	A general MTL model showing shared and task-specific layers . .	14
3.5	Our U-Net adapted MTL model . . . . .	15
3.6	A simple Residual Block (cited from <b>K. He [2015]</b> ) . . . . .	16
4.1	Tumour Binary masks . . . . .	18
4.2	Projection along axis=3 . . . . .	20
4.3	<b>Row 1</b> is the Max intensity projection, <b>Row 2</b> is the Mean intensity projection and <b>Row 3</b> is the SD projection . . . . .	22
4.4	Training and Testing phases of 2D model . . . . .	23
4.5	Images with their cropped masked tumours . . . . .	25
4.6	Training and Testing phases of 3D volume model . . . . .	26
4.7	3-axis 3D volume classifier ensemble . . . . .	28
5.1	Tumour mask predictions on test images. Column 1 shows the actual image, followed by the <b>Ground truth mask</b> and the <b>Model predictions</b> . . . . .	30



# List of Tables

2.1	Kaggle leaderboard top 6 teams . . . . .	5
3.1	Brain tissue visibility in MRI . . . . .	9
5.1	Classwise IoU scores for each modality . . . . .	30
5.2	Performance of 2D middle slice MTL (T1) . . . . .	31
5.3	Performance of 2D middle slice MTL (Flair) . . . . .	32
5.4	Performance of 2D 10-slice MTL (Flair) . . . . .	32
5.5	Projection Performance (axial direction) . . . . .	33
5.6	3 Axis Performance (SD projections) . . . . .	33
5.7	Results of 2D cascaded model . . . . .	34
5.8	Results of 5-slice 3D cascaded (single model) . . . . .	35
5.9	Results of 5-slice 3D cascaded (ENSEMBLE) . . . . .	36
5.10	Results of 3-axis Ensemble . . . . .	37

# Nomenclature

## Greek Symbols

$\pi$   $\simeq 3.14 \dots$

## Superscripts

$j$  superscript index

## Subscripts

$0$  subscript index

## Acronyms

*AUC* Area Under Curve

*GBM* Glioblastoma

*MGMT* O6-Methylguanine-DNA methyltransferase

*MRI* Magnetic Resonance Imaging

*MTL* Multi Task Learning

*TMZ* Temozolomide

# Chapter 1

## Introduction

Glioblastoma multiforme (GBM) is the most probable category of brain tumour that affects adults. Over 45% of primary central nervous system tumours are accounted by GBM, and the patient survival rate is around 5.1% for a five-year period (as documented in [Morgan \[2015\]](#)).

### 1.1 Motivation

O6-methylguanine-DNA methyltransferase (MGMT) promoter is a significant molecular biomarker of glioblastoma that has relevance in cancer therapeutics. MGMT is one of the DNA repair enzymes and the methylation status of its promoter in newly diagnosed GBM tumour has been identified as a predictor of chemotherapy response ([A. L. Rivera \[2009\]](#)). The survival benefit governed by the MGMT gene methylation status during temozolomide (TMZ) treatment was observed by [Hegi ME \[2005\]](#). Further work has asserted that for patients who received chemotherapy through temozolomide, the methylation status of the MGMT promoter gene improved median survival compared with patients who had unmethylated gliomas. Thus, MGMT methylation status as reported in 30-60% of glioblastomas can enhance the response to TMZ (as recorded in [A. A. Brandes \[2008\]](#)).

Currently, surgical specimens based genetic lab testing is the standard to assess the MGMT methylation status of GBM. In this method, a large tissue

sample is required, which is tested by polymerase chain reaction (PCR) to judge the methylation status(L. Wang [2017]). The major limitations of this invasive methodology are the possibility of incomplete biopsy samples due to tumour spatial heterogeneity and the high cost involved in tissue sampling (N. R. Parker [2016]). It can lead to neurological injury and nerve damage during the operative procedure. Therefore, there is a neccessity to find non-invasive methods based on brain imaging, which can successfully lead to an effective prediction of the methylation status of glioblastoma.

## 1.2 MRI Imaging

Magnetic Resonance Imaging (MRI) is a popular diagnostic tool in detection and treatment of GBM tumours (P. Y. Wen [2010]) The monitired, coordinated use of multiple, mutually informative imaging probes to understand brain structure and function has given raise to multimodal MRI imaging techniques, that has shown greater visibility of gliomas.

In recent studies in radiogenomics, it has been shown that quantitative and qualitative features extracted from medical images can have a direct correlation with the genetic traits of the tissues. In this regard, radiogenomic data can be directly utilized to analyse the MGMT methylation status of the GBM tumour. (L. Macyszyn [2015] and L. S. Hu [2017]). However, feature extraction from MRI images can be a cumbersome process as it involves a manual differentiation of the glioma tumour region from rest of the grey matter. Information analytics of the extracted feature is also prone to disagreement between the observers responsible for the tumour segmentation (P. Lambin [2012] and H. J. Aerts [2014])

## 1.3 Deep learning approaches

With the boom in deep learning techniques in medical imagings, several researchers have taken up the problem of GBM tumour diagnosis. Deep learning can enable automatic extraction of features based on the given loss function

optimization for the particular use case (Y. LeCun [2015]).

Convolutional Neural Networks (CNNs) make use of the convolution operation in images, and has been effective in image analysis. (Y. Yang [2018], A. Akay [2019], G. S. Tandel [2019]). As compared to traditional methods with hand-crafted features, deep learning methods has advantages of being robust to image distortions like changes in shape, orientation and involves a lower computational cost. The neural network architecture has the capacity to ‘learn’ weights on its own when supplied with a considerably large training set, and this predictive ability has been widely beneficial.

Although multiple radiomic approaches have used deep learning to predict the MGMT gene methylation, the clinical viability of the prediction and its reliability when compared to surgical diagnosis by physicians is a matter of open research. Drabycz S [2010] performed texture analysis on MR images to predict MGMT promoter methylation status, whereas Korfiatis P [2016] used combined texture features along with supervised classification schemes as imaging biomarkers.

There remains a scope for investigation of a complete end-to-end pipeline that can do automatic tumour detection, tumour region segmentation and subsequent MGMT methylation classification, with minimum user intervention. Since the MRI scans are often supplied as 3D volumes, efficient 2D sampling should be a crucial element of the model pipeline. We thereby try to come up with such an architecture that can achieve the discussed use cases.

## Chapter 2

# BraTS 2021 challenge and Kaggle leaderboard

The Radiological Society of North America (RSNA) teamed up with the Medical Image Computing and Computer Assisted Intervention Society (MICCAI) to improve the treatment planning for patients with glioblastoma. The RSNA-ASNR-MICCAI BraTS 2021 challenge (Baid [2021]) targetted the evaluation of ML algorithms assessing the tumour region segmentation, as well as the underlying tumor’s molecular characterization, in pre-operative baseline MRI imaging data. Specifically, the two tasks that BraTS 2021 focuses on are: a) the segmentation of the histologically distinct brain tumor sub- regions, and b) the classification of the tumor’s O[6]-methylguanine-DNA methyltransferase (MGMT) promoter methylation status.

The part 2 of the competition which involves predicting the MGMT methylation status of the given patient MRI smaples was hosted on Kaggle<sup>1</sup>. Kaggle is an online community of data scientists and machine learning practitioners.

---

<sup>1</sup><https://www.kaggle.com/competitions/rsna-miccai-brain-tumor-radiogenomic-classification>

## 2.1 Current Kaggle leaderboard

The Kaggle 2021 competition saw the participation of 1555 teams. The following table summarizes the model architectures and scores of the top six teams of the currently featuring [leaderboard](#).

Current BraTS 2021 leaderboard		
Rank	Model used	AUC performance
1	3D CNN + ResNet	0.621
2	EfficientNet + LSTM	0.618
3	4 parallel EfficientNet	0.617
4	YOLO (2D) + 3D CNN	0.615
5	EfficientNet	0.607
6	UNet + LSTM	0.606

Table 2.1: Kaggle leaderboard top 6 teams

The **Rank 1** team uses a combination of 3D CNNs and a popular ResNet architecture to stack the 2D scans around an appropriately chosen 2D middle slice (from given DCM files) and then pass the 3D volume for classification. 3D CNNs are computationally exhaustive and require a heavy GPU for the larger convolutional load. The **Rank 2** team uses a complex sampling strategy to introduce sequentiality into the data. They sample 10 scans from all patients (uniform temporal subsampling) across the 4 MRI modalities (T1cE, T2cE, T1wCE, Flair) to get 4-slice stacks which were fed into an LSTM network.

The **Rank 3** team uses 4 EfficientNet networks parallelly for each of the 4 given modalities. The networks are simultaneously trained during the training phase, but in the testing phase, for each modality, predictions are obtained on a slice-

by-slice basis and the most confident prediction probability is selected. Finally on the patient level, the average across 4 modality is obtained. The **Rank 4** team uses YOLO models on the 2D slices to detect the slices which have the best tumour visibility and then feed them into pre-trained classification networks. The **Rank 5** team sampled 10 images from each modality (patient level), took pixel intensity average of the 10 images to obtain one single image per modality, then stacked them together to get 4-slice-stacks. It was then fed to an EfficientNet model. Lastly, the **Rank 6** team sampled 2D slices from segmentation masks (with high tumour visibility) to get 2 masks (Tumour core, Enhancing tumour). The 2D slice was stacked with 2 kinds of mask to get 3-slice input, which was then fed to an LSTM network in chunks.

## 2.2 Observations and Scope

The leaderboard test dataset AUCs saturate at around 0.61. Observing the leaderboard approaches, it can be sensed that less focus is being given to localize the tumour voxel and then feed the Region-of-interest into a suitable classification network. Hence, the utility of an efficient Tumour segmentation task becomes apparent. Localization of the tumour region can be done in 2D slice basis or by taking into account the whole 3D volume. In this report, we highlight our investigations in this direction.

It is to be noted that the tumour segmentation and Multi-task model based experiments was conducted in the Autumn semester (Sem 1 2022-23) while the volumetric projections and cropped-cascaded model tests were conducted in the Spring semester (Sem 2 2022-23). The source code, implementations, models can be found in the GitHub<sup>1</sup> repository.

---

<sup>1</sup><https://github.com/Deepan2486/Radiogenomic-classification-glioblastoma-multimodal-3D-MRI>



# Chapter 3

## Dataset and Experimental set-up

### 3.1 Dataset

The Brain Tumour Segmentation Challenge (BraTS) dataset describes a collection of brain tumor MRI scans acquired from multiple institutions under standard clinical conditions, but with different imaging protocols. The data inclusion criteria comprised pathologically confirmed carcinoma diagnosis and available MGMT promoter gene methylation status (Baid [2021])

#### 3.1.1 Dataset description

The BraTS 2021 task was classified into two phases:

1. **Tumour region segmentation (Phase 1):** This phase was hosted on the Synapse platform. The mpMRI scans describe a) native (T1) and b) post-contrast T1-weighted, c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes. After standard pre-processing, the DICOM data files were converted to NIfTI file format (R. Cox [2004]). Co-registration to the same anatomical template (SRI24) was performed (as in T. Rohlfing [2010]) and the scans were resampled to a uniform isotropic resolution (1mm<sup>3</sup>), and finally skull-stripped. To construct the training set, all imaging volumes were segmented using the STAPLE fusion (S. K. Warfield [2004]) and then manually annotated by radiological experts.

2. **Radiogenomic classification (Phase 2)** : This phase was hosted on Kaggle. The annotated NIfTI files were again converted into a series of DICOM scans (in the patient space) using CaPTk's NIfTI to DICOM conversion engine. This process is repeated across the 4 modalities. The Pyromark CpG MGMT kit detected the average level of methylation on CpG 74–81 sites located in the MGMT gene. The percent methylation above 10% was interpreted as positive. A sample below 10% methylation was interpreted as negative

### 3.1.2 MRI modalities

MRI is based on the magnetization properties of atomic nuclei. By changing the sequence of Radio Frequency pulses applied, different types of images are created. Repetition Time (TR) is defined as the time duration between successive pulse sequences applied to a particular slice. Time to Echo (TE) is the time duration between the delivery of the RF pulse and the receipt of the echo signal. The most common MRI sequences are T1-weighted and T2-weighted scans. **T1-weighted** images are produced by using short TE and TR times. **T2-weighted** images are produced by using longer TE and TR times. In general, T1- and T2-weighted images can be easily differentiated by looking at the Cerebro-spinal-fluid (CSF), which is dark on T1-weighted imaging and bright on T2-weighted imaging. A third modality is **Fluid Attenuated Inversion Recovery (Flair)**. The Flair sequence is similar to a T2-weighted image except that the TE and TR times are very long. T1-weighted imaging can also be performed while infusing a non-toxic paramagnetic contrast enhancement agent Gadolinium (Gad). When injected, Gad changes signal intensities. It is termed as **T1-contrast-enhanced** images and are especially useful in looking at vascular structures. The visibility of brain structures in the modalities can be described as:

	T1w	T2w	Flair
CSF	Dark	Bright	Dark
White matter	Light	Dark grey	Dark grey
Cortex	Grey	Light Grey	Light Grey

Table 3.1: Brain tissue visibility in MRI

The MRI scan modalities are depicted below, represented across the 3 axis (**axial**, **coronal** and **saggital**).

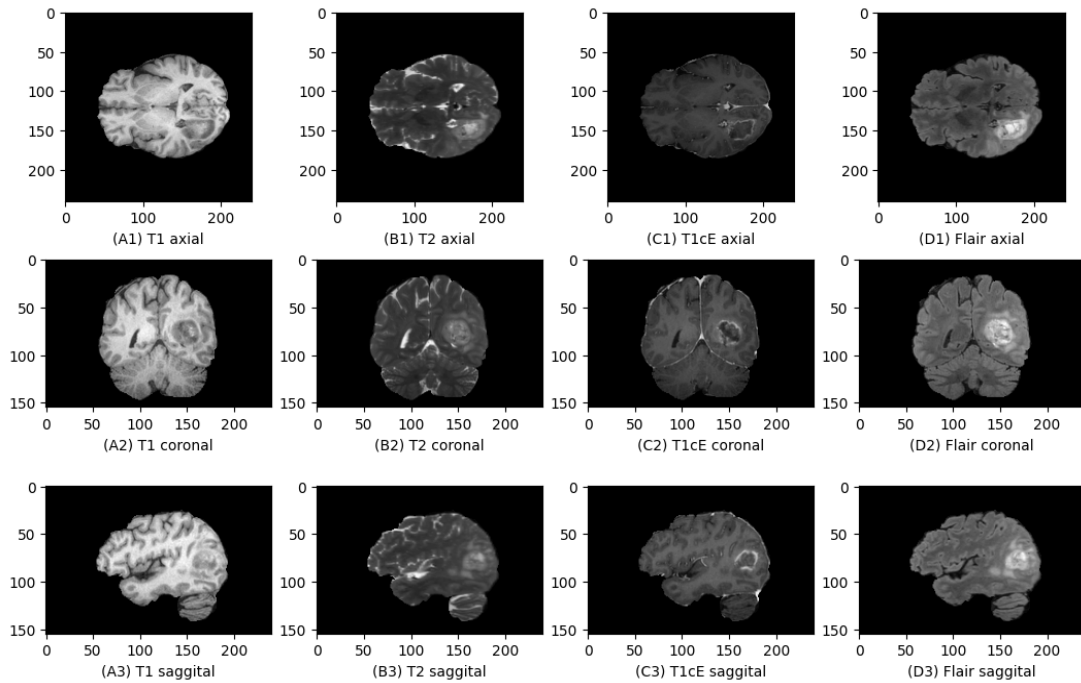


Figure 3.1: Modalities across 3 axis. **A1-A3** depict T1-weighted images, **B1-B3** as T2-weighted, **C1-C3** as T1-contrast-enhanced and **D1-D3** as Flair. The first row is the **axial** direction, then **coronal** and **saggital** showing the various positions of the tumour

As observed, the Flair scan images gives the best tumour voxel visibility by illuminating the tumour tissue with pixel intensity very different from the surrounding region. The T1cE scans outline the tumour regions and depict the vascularity. The T1 and T2 scans are not representative enough of the tumour location, as assessed by monitoring multiple cases from the dataset.

### 3.1.3 Number of samples

The patients were uniquely labelled with a number called the BraTS21-ID. There were 595 datapoints provided as segmentation masks and 585 classification ground-truth labels were given. In all, **573 unique patient IDs** could be utilised for experimentation purposes. Out of these, **298** were labelled as methylated and **275** were unmethylated. So, class imbalance was minimal.

Each of the unique BraTS21-ID had five NIfTI files (corresponding to T1-weighted, T2-weighted, T1-contrast-enhanced, Flair and the multiclass segmentation mask) along with the classification true labels. All the NIfTI files were volumetric with dimensions **(240,240,155)** which meant that 2D slices of dimensions **240 x 240** were stacked on top of one another to form a height of 155 slices. The methylation status labels were 0 or 1 (binary classification)

In experimentations without a cross-validation step, 20% of the datapoints were reserved for testing and the rest 80% for training purposes. Further, during training the validation split was 0.2. So, among 573 samples, **458** were used for training, **115** for testing (within training, 91 datapoints were used internally for validation).

In experiments where we used 5-fold-cross validation, the entire dataset was divided into 5 chunks of {114,114,115,115,115} datapoints. During each fold, one of the sets was used as a test set and a stratified combination of the other 4 as the training set.

### 3.1.4 Tumour segmentation regions

The tumour segmentation part of the task involves dividing the localized tumour regions into pixel categories of  $\{0,1,2,3\}$  based on the kind of tumour tissue. The classes are as follows:

1. **Label 0** : The background region
2. **Label 1** : Necrotic Tumour Core NCR
3. **Label 2** : peritumoral edematous invaded tissue (ED)
4. **Label 3** : Gd enhancing tumour (ET)

ET is the enhancing portion of the tumor, described by areas with faint enhancement on T1Gd (T1CE) MRI. NCR is the necrotic core of the tumor, the appearance of which is intense on T1cE MRI. ED is the peritumoral infiltrated tissue, defined by the abnormal hyperintense envelope on the T2-weighted and FLAIR volumes. They include the infiltrative non enhancing tumor.

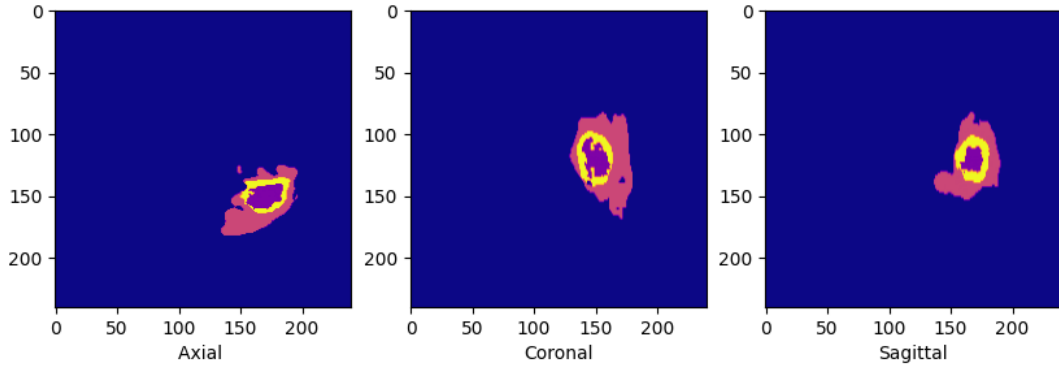


Figure 3.2: Tumour regions (in the axial, coronal, sagittal directions). The different colours depict the multiclass pixel labels

As can be seen, the tumour regions are distinctly prominent in all the 3 directional axes. These mask categories are independent of the MRI modality. They show the ground-truth labels of the pixels that depict the tumour tissue.

### 3.1.5 Methylated and Non-methylated tumours

According to the discussed background of the problem, the methylation status of the MGMT gene can be correlated with biomarkers in the MRI scan. The methylation status can be informally judged by looking at the distribution of the tumour pixel categories, and in the way the tumour region is spread out.

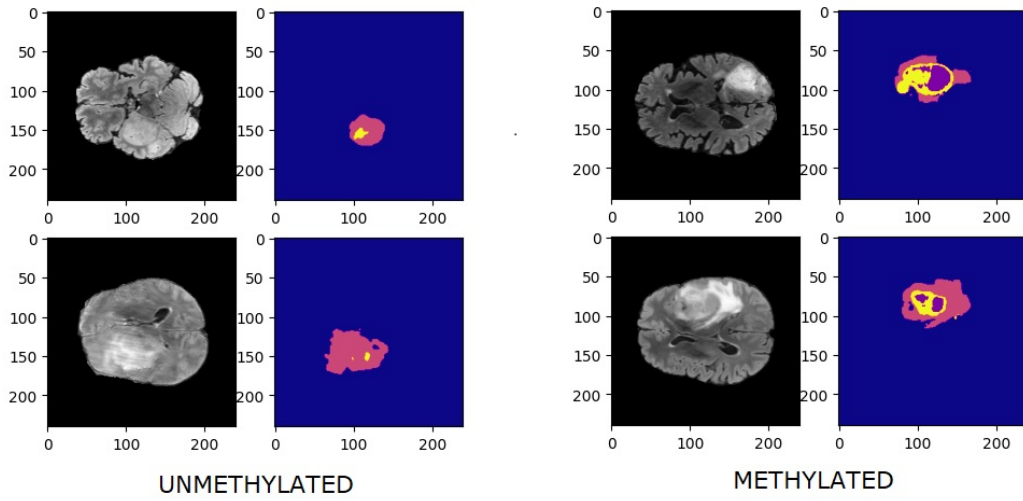


Figure 3.3: (A) Unmethylated tumours and (B) Methylated tumours

The unmethylated tumour examples in (A) have a less-defined tumour core region, and although the invaded tissue extent is large, the necrosis of the core is subdued. Whereas in methylated samples (B), the tumour core is starkly prominent and the tumour growth is visible from inside-out.

The methylation classification of the tumour is challenging because the tissue biomarker of the underlying genomic characteristic is not directly inferrable. In the methylation ground-truth labels provided as training data, a 10% benchmark across the gene sites was considered for the binary classification. Since the MRI scans can have varying levels of carcinoma and the MGMT methylation percentage can vary, therefore there remains some sort of ambiguity in the visible differences.

## 3.2 Model architectures

Various model architectures were experimented for the segmentation and classification tasks. All the models that were based on 2D Convolutional Neural Networks (CNN) and the input data had multiple channels based on the requirement.

### 3.2.1 UNet 2D

U-Net was introduced by [O. Ronneberger \[2015\]](#) for highly accurate biomedical semantic image segmentation. It gained popularity because it required less number of training samples.

UNet is based upon the encoder-decoder model of deep CNN architecture and incorporates an encoder block, skip connections and a decoder block. It combines feature extraction and feature localization strategies.

1. **Encoder block:** Encoder captures the most relevant information in the input image into a lower-dimensional space, by extracting the most definitive features. It uses a series of convolution, pooling and fully-connected layers to arrive at a compact representation.
2. **Decoder Block:** Decoder block is the exact opposite of the encoder operation. By using the compact feature representation, it reconstructs the image by using deconvolutional and upsampling operations which increase the image dimensionality.
3. **Skip connections:** This step is crucial for preserving the locational or spatial information of the image. In U-Net, skip connections are used to pass information from earlier convolutional layers to the deconvolution layers. What is passed is the location of the feature extracted by convolutional layers. This is done by concatenating the last layer in the convolutional block with the first layer of the opposite deconvolutional block.

The UNet architecture is symmetrical and the dimensions of the opposite layers/blocks will be the same. If we input a (X,Y) dimensional 2D image which

is to be semantically segmented to 4 classes, then the output image will be (X,Y,4) in the categorical data format.

### 3.2.2 Multi-task-learning (MTL) model

A novel Multi-task-learning model was developed from an UNet adaptation. Multi-task-learning models refers to networks which are optimized for more than one task, and they simultaneously train the network parameters for all the tasks. By sharing representations between the related tasks, it avoids overfitting and enables the model to learn better.

In our problem statement, the **segmentation** and **classification** tasks are related to one another and can be learnt jointly. We employ **Hard parameter sharing** in which there are task-specific individual layers and shared layers. Shared layers have to learn parameters that fit both the given tasks. The model back-propagates along both the individual and the shared layers.

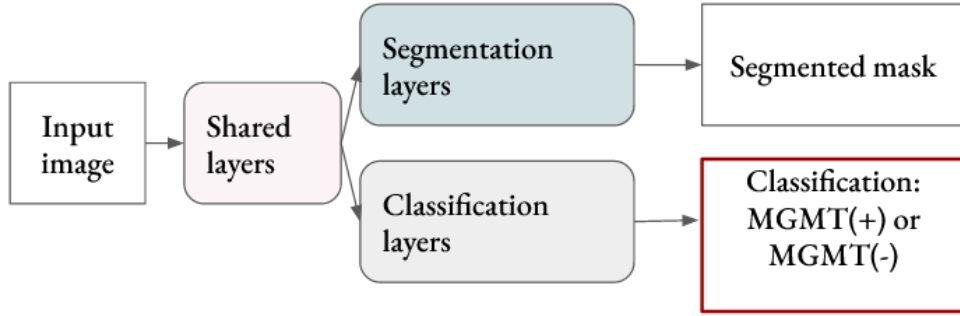


Figure 3.4: A general MTL model showing shared and task-specific layers

The semantic segmentation model of UNet was adapted into an MTL model by adding the **classification branch**. The classification branch utilizes the compact feature representations from the lowest 3 blocks of the UNet (one from encoder, one from decoder and the bridge layer). Since these 3 blocks have outputs of





### 3.2.3 ResNet18 and ResNet34

A new category of deep networks called **Residual networks (ResNet)** was introduced in [K. He \[2015\]](#). The layers are reformulated to learn residual functions with reference to the layer inputs, instead of learning unreferenced functions. These ResNets are easier to optimize, and can gain accuracy from considerably increased depth. By using '**skip connections**' periodically in the model architecture, activations of a layer are connected to further layers by skipping some layers in between, thus forming a '**Residual Block**'. Identity shortcut connections add neither extra parameter nor computational complexity. The entire network can still be trained with back propagation.

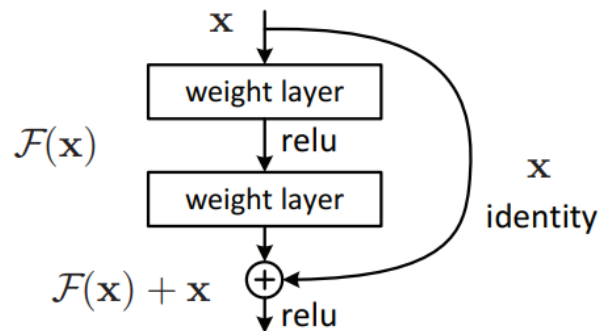


Figure 3.6: A simple Residual Block (cited from [K. He \[2015\]](#))

ResNet uses a network architecture inspired by VGG-19 ([K. Simonyan \[2015\]](#)) in which shortcut connections are added. On the ImageNet dataset, the authors of the paper uses a 152-layers ResNet, which is 8 times more deep than VGG19 but still have less parameters. In general, **ResNet18** and **ResNet34** are used for smaller classification tasks which are 18 layers and 34 layers deep respectively.

### 3.2.4 EfficientNet

A new family of models called EfficientNet was introduced in [M. Tan \[2019a\]](#). The EfficientNet architecture uniformly scales all dimensions of depth/width/resolution

using a compound coefficient (**model scaling**). The EfficientNet scaling method uniformly scales network width, depth, and resolution with a set of fixed scaling coefficients, rather than using arbitrary scaling factors.

The baseline model **EfficientNet B0** is based on the inverted bottleneck residual blocks that was demonstrated in MnasNet (M. Tan [2019b]), in addition to the squeeze-and-excitation blocks. The building block of this architecture is the mobile inverted bottleneck MBConv (inverted residual block) with an additional Squeeze-and-Excitation block. Using compound scaling, **EfficientNet B1** to **EfficientNet B7** are developed, where the number of model parameters increase exponentially, but give higher accuracy.

# Chapter 4

## Methodology

### 4.1 Multi-Task-Learning (MTL) model

As depicted in Figure 3.5, the segmentation-classification MTL model performs multiclass segmentation to get the detected tumour region from the MRI, and also gives the prediction probability of the binary classification of the glioblastoma being methylated/unmethylated. (Results are summarized in section 5.2)

We only need an accurate tumour localization in the segmentation step, so we change the tumour masks from being 4-class of labels  $\{0,1,2,3\}$  to being a binary mask of labels  $\{0,1\}$ . This is done by simply equating the non-zero pixel labels to 1. The below figure shows some examples:

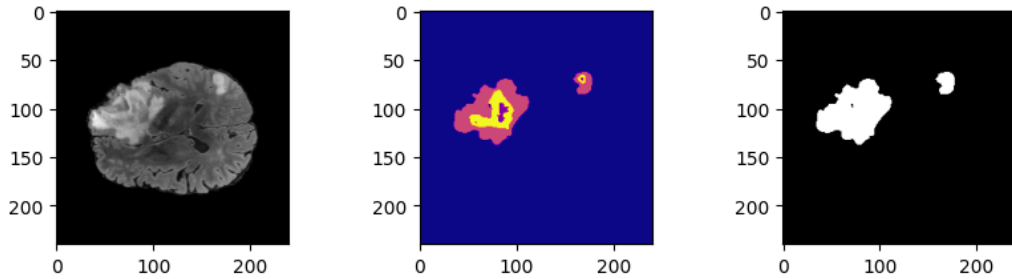


Figure 4.1: Tumour Binary masks

$$L_{joint} = \lambda L_{cls} + (1 - \lambda) L_{seg}$$

The loss function of an MTL model is jointly given by the loss of the classification  $L_{cls}$  and the loss of segmentation  $L_{seg}$ , weighted by a parameter  $\lambda$ . The variable  $\lambda$  ranges from  $[0,1]$  and is used as a hyperparameter.

For the binary classification task, we use **Focal Loss** as the loss function. For segmentation, we experimented with both **Binary Cross Entropy** and **Dice Loss**.

$$BCELoss = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i))$$

$$FocalLoss = \sum_{i=1}^N (1 - p_i)^\gamma \log(p_i)$$

$$DiceLoss = 1 - \frac{1 + 2 * P_s * Y_s}{1 + P_s + Y_s}$$

In Focal Loss, the **Focusing Parameter**  $\gamma$  controls the class imbalance in the classification labels and also reduces the influence of easy examples on the loss function, resulting in more attention being paid to hard examples (**Down Weighting**).  $N$  is the total number of classes (here 2),  $y_i$  is the true label of the classification and  $p_i$  is the prediction probability.

In BCE Loss (used for segmentation),  $y_i$  represents the true tumour label mask, and  $p_i$  represents the predicted tumour label mask. For the segmentation task, the imbalance between foreground and background may cause segmentation bias. To account for this problem, a segmentation loss based on the Dice coefficient (Dice Loss) is utilized to emphasize shape similarity between actual and predicted segmentation maps. For Dice Loss,  $Y_s$  is the actual tumour mask whereas  $P_s$  is the predicted tumour mask.

Two approaches were taken for the MTL experimentation. Since the model is trained for 2D images, fitting in the entire 3D volume is not feasible. We therefore employ 2 sampling methods:

1. **Middle slice:** We use the middle slice (slice=70) from each NIfTI scan, and then pass this 2D image into the network. The mask for training is also sampled by taking slice=70 from the corresponding true mask.
2. **10 slices from each volume:** In order to introduce variability in the dataset, we sample 10 slices from the entire 155 slices. The samples are taken by eliminating the upper and lower 20% of the slices (since they hardly have any relevant tumour information). Iteratively we save one slice after another by spacing out 2 slices in between. This way, the dataset slices increases 10-fold. During the sampling, a check is employed to ensure that the slice has the requisite tumour visibility. If not the case, then we move to a new slice. This method trains the MTL model towards predicting unseen tumours of varying shapes and structure.

## 4.2 Volumetric Projections

Since the dataset dimensions are (240,240,155) where 155 is the number of slices in each scan, there is a need to capture the entire 3D spatial information into a 2D dimension. We therefore utilize various projection approaches to experiment.

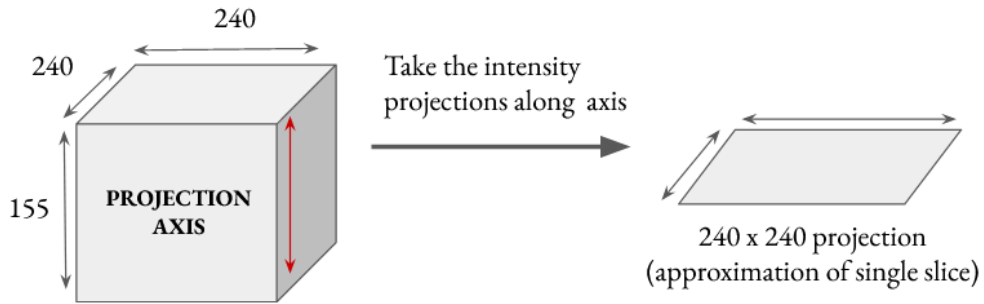


Figure 4.2: Projection along axis=3

The technique of projections is fairly simple. First, an axis is decided (let us assume we have chosen axis=3). Then for each and every particular pixel intensity on a plane  $(x_i, y_j, z)$  we consider all the values along axis=3, and then take the projection operation across all such values. For instance, here we take all the values of  $(x_i, y_j, z)$  where  $z \in [0, l]$  and  $l$  = number of slices along the axis. This process is repeated for all  $x_i$  and  $y_j$  where  $j \in [0, m]$  and  $i \in [0, n]$ ,  $m$  and  $n$  being the dimensions of the 2D plane we are considering. At the end of this projection operation, we arrive at the 2D plane with pixel intensity corresponding to all the resultant values.

Based on the kind of projection operation we can perform, we consider 3 kinds of projections:

1. **Mean Intensity** : Here the mean of the intensity values of  $\{z_0, z_1, \dots, z_l\}$  is taken where  $l$  is the number of slices along the axis.

$$z_{mean} = \frac{1}{L} \sum_{i=1}^l z_i$$

2. **Max Intensity**: The maximum intensity values of among  $\{z_0, z_1, \dots, z_l\}$  is taken where  $l$  is the number of slices along the axis.

$$z_{max} = \max(z_i) \text{ for } i \in [1, l]$$

3. **Standard Deviation of Intensity**: The Standard Deviation (SD) values of among intensities  $\{z_0, z_1, \dots, z_l\}$  is taken where  $l$  is the number of slices along the axis.

$$z_{SD} = \sqrt{\frac{1}{L} \sum_{i=1}^l (z_i - \bar{z})^2} \text{ where } \bar{z} \text{ is the mean intensity}$$

Various levels of tumour tissue information, vascularity and carcinoma invasion can be inferred from the kind of projections. The below figure depicts one patient exmaple (across 4 modalities).

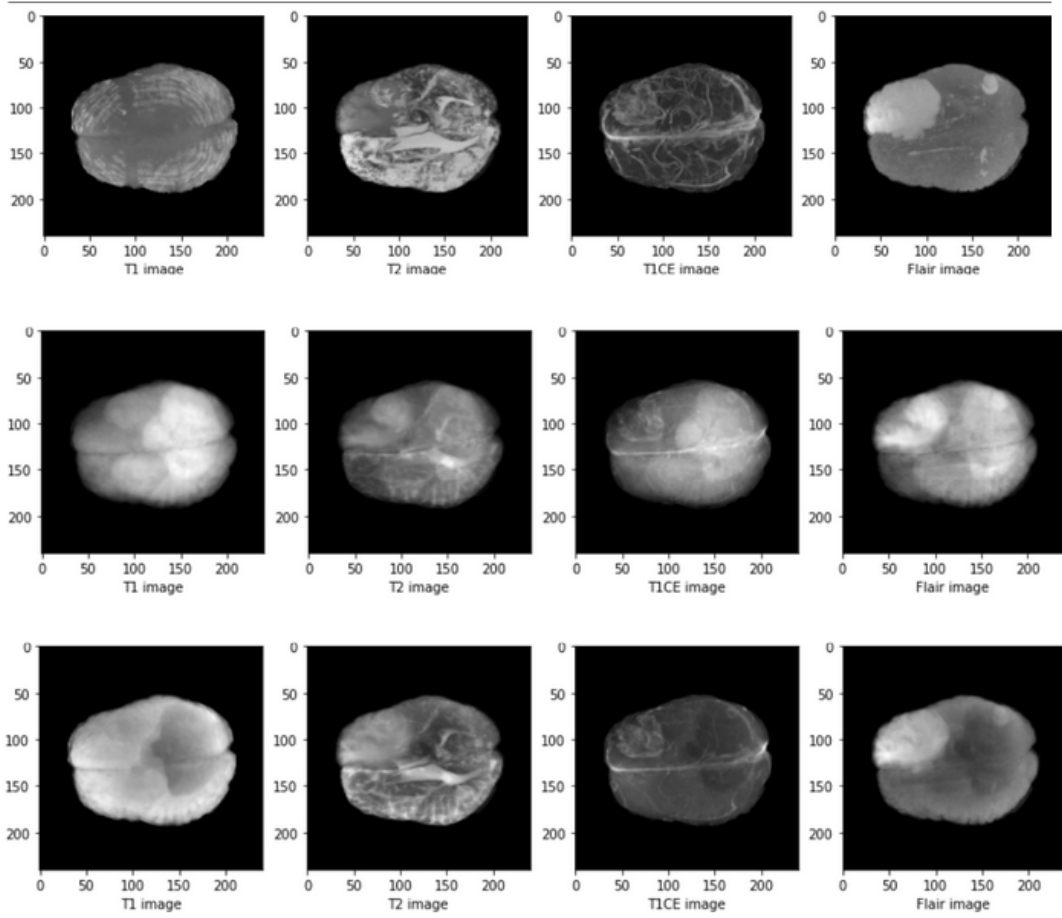


Figure 4.3: **Row 1** is the Max intensity projection, **Row 2** is the Mean intensity projection and **Row 3** is the SD projection

Classification accuracies was compared with 3 kinds of projections, as well as on 3 different axes (axial, coronal and Saggital). **EfficientNet B7** was used as a baseline model with **Focal Loss** as the loss function. Generally, EfficientNet



is trained for dimensions of  $(x, y, 3)$  to incorporate 3-channel RGB, but in this case we used it for predicting 2D images of dimensions  $(x, y, 1)$ . The results are summarized in the later section 5.3.

### 4.3 Cascaded mask-cropped model

One drawback of using the entire tumour image for classification is that it encodes redundant information. The pixels which correspond to the non-tumour tissues can be disposed. Therefore, there is a strong need for a tumour localization strategy that solves this issue. The solution is to train separate classification and segmentation models in a cascaded fashion.

The tumour localization can be done on a pre-determined 2D level or an adaptive 3D level. The 2D approach can have a large bias since the selection of the 2D slice isn't dependent upon the patient scan sample, hence ascertains no effect to the end-to-end pipeline.

#### 4.3.1 2D mask-cropped

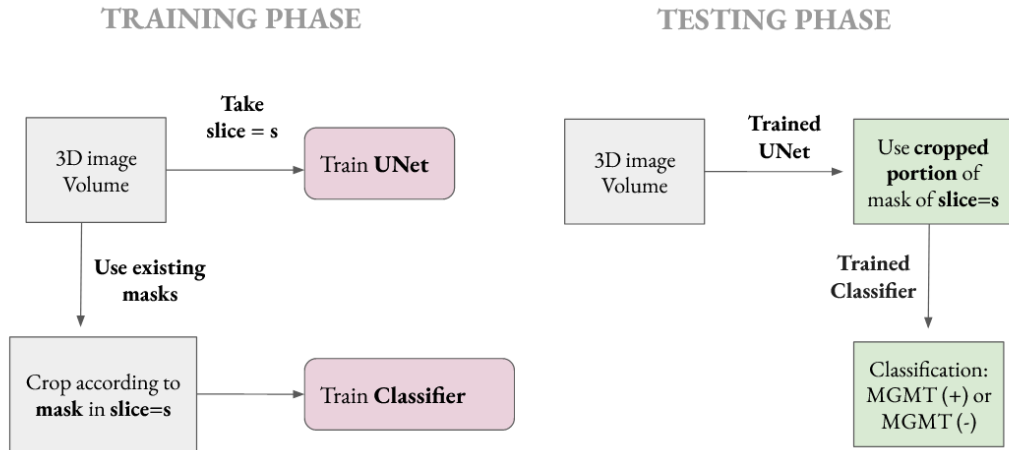


Figure 4.4: Training and Testing phases of 2D model

---

**Algorithm 1** 2D cascaded: SEGMENTATION

---

**Require:** Slice = s (pre-determined by user)

- 1: **for** <Each sample in TRAIN set > **do**
  - 2:     Extract the slice = s (pre-defined)
  - 3:     Pass that slice to UNet
  - 4:     Train Unet on all the slices = s
  - 5:
  - 6: **for** <Each sample in TEST set > **do**
  - 7:     Pass the slice = s to UNet
  - 8:     Get the tumour mask
  - 9:     Get the best-bounded-box of the tumour (crop)
  - 10:    Save it in a separate folder
- 

After we have all the tumour-cropped portions of the train and test set, we move towards the Classifier training. It is to be noted that the cropped regions of the train set are directly obtained from ground-truth masks whereas the cropped regions of the test set are obtained by using the UNet predictions.

---

**Algorithm 2** 2D cascaded: CLASSIFICATION

---

- 1: **for** <Each sample in TRAIN set > **do**
  - 2:     Use ground-truth mask to get cropped image
  - 3:     Pass that cropped slice to Classifier
  - 4:     Train Classifier
  - 5:
  - 6: **for** <Each sample in cropped TEST set > **do**
  - 7:     Pass that cropped slice to Classifier
  - 8:     Get methylation prediction probability and output (**patient level**)
- 

The following images show some examples with the original tumour image, the binary mask and the cropped mask region.

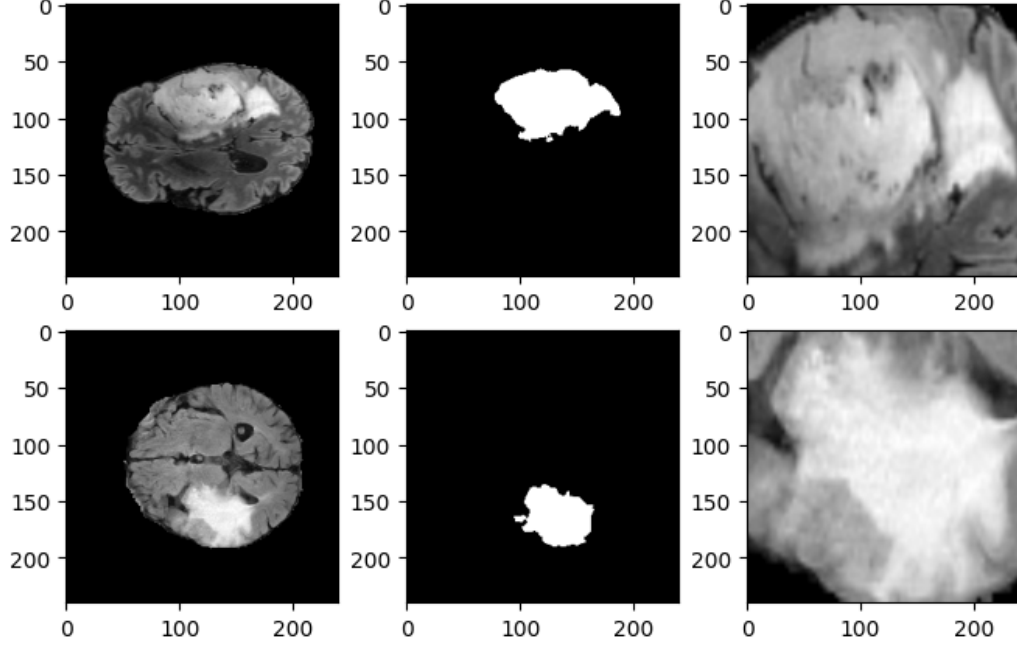


Figure 4.5: Images with their cropped masked tumours

As classifier, we have used **ResNet18** and **EfficientNet B7** with Focal Loss as the classification loss function. The results are summarized in the subsequent section. However, using a pre-determined slice for the pipeline is not desired because it leads to a high bias. Hence, we move to 3D tumour localizations.

### 4.3.2 3D mask cropped

The 3D tumour localization version takes into account a 3D stacked tumour voxel instead of just a single slice. The central location of the voxel (from the 155 slices) is judged according to the tumour visibility of the slices. The slice with the best tumour visibility is chosen as the middle slice and other slices are taken around it. The tumour visibility is judged by the percentage of '1' labelled pixels in the predicted mask.

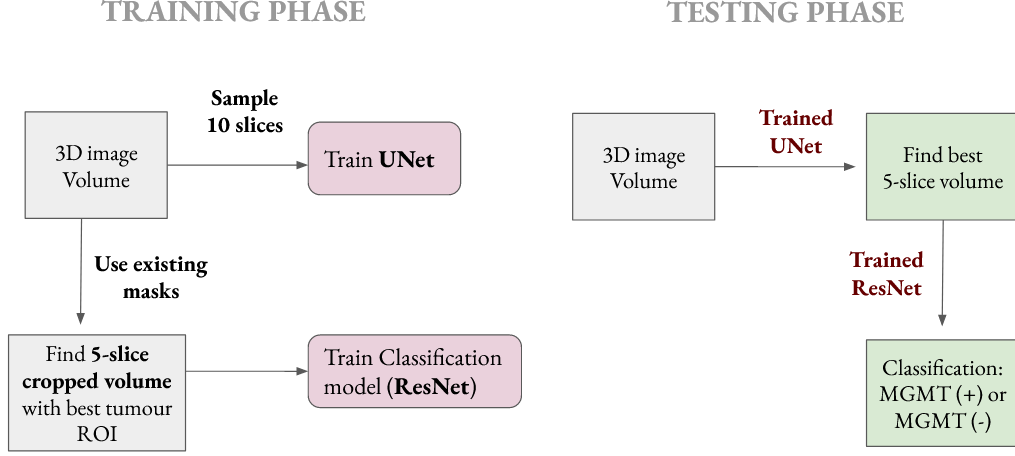


Figure 4.6: Training and Testing phases of 3D volume model

After repeated experimentation across all 3 axes (axial, coronal and sagittal) we decided to use a **5-slice volume** for the 3D localization of the tumour. The middle slice from the 5-sliced stack has the best tumour visible and 2 slices each are chosen from the top and bottom of the chosen best slice.

---

**Algorithm 3** 3D cascaded: SEGMENTATION

---

- 1: **for** <Each sample in TRAIN set > **do**
  - 2:     Sample **10 slices** (except from top 20% and bottom 20% slices)
  - 3:     Train Unet on all the sampled slices
  - 4: **for** <Each sample in TEST set > **do**
  - 5:     Pass **each slice** to UNet
  - 6:     Get the tumour mask for each slice individually
  - 7:     Find the best slice with most tumour visible, Say  $slice = x$
  - 8:     Crop the tumour portions from slices  $[x - 2, x - 1, x, x + 1, x + 2]$  to get 5-stacked volume
  - 9:     Save 5-stacked volume in a separate folder
-

---

**Algorithm 4** 3D cascaded: CLASSIFICATION

---

```
1: for <Each sample in TRAIN set > do
2:   Use ground-truth mask (slice-by-slice) to get  $slice = a$  with best tumour
   visibility
3:   Crop the tumour portions from slices  $[a - 2, a - 1, a, a + 1, a + 2]$  to get
   5-stacked volume
4:   Pass that 3D volume to Classifier
5:   Train Classifier
6:
7: for <Each sample in 3D cropped TEST set > do
8:   Pass that 3D 5-stacked volume to Classifier
9:   Get methylation prediction probability and output (patient level)
```

---

The benefit of extending the 2D slice version to a 3D volume version is that the pipeline now decides which tumour slice to extract the most information from. A quick iteration across all the given datapoints told us that the tumour location is fairly variable and hence using a pre-set slice is not the best option. Also, training the Unet on 10 samples from each patient scan makes the segmentation more robust for unseen data.

The 3D version was tried with various models (across all the 3 axes): **ResNet18**, **ResNet34**, the EfficientNet model group (**EfficientNet B0**, **EfficientNet B3**, **EfficientNet B7**). Ensembles of ResNet18 + ResNet34 and ResNet + EfficientNet was also experimented with. The AUCs and accuracies are presented in section 5.4.

### 4.3.3 Ensemble along 3 axes

The final part of our methodology involved using a 3-axis-ensemble, in which the same patient scan is utilized along the 3 axes (by using 3 separately trained UNet and 3 separately trained Classifiers) and the final predictions are ensemble together to get the final prediction probability. The same 3D MRI scan is sampled along the 3 axis to train the Unet and then used sliced by slice (in testing phase) to get the 5-stack volume which has the best tumour visibility along that particular axis.

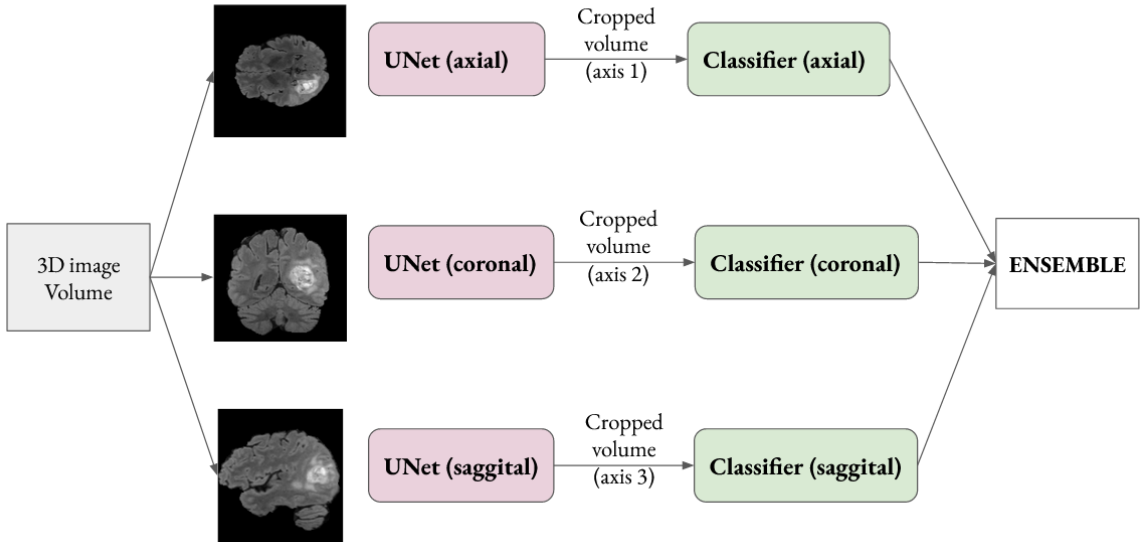


Figure 4.7: 3-axis 3D volume classifier ensemble

The classifiers used in each case for each branch was **ResNet18** and **ResNet34** since they were lightweight. The number of model parameters becomes three times due to the increased branches of axis 2 and axis 3. The results of the ensemble are shown in section 5.5.

# Chapter 5

## Results and Discussion

### 5.1 Segmentation results

Before we proceeded to the MTL model inferences, we tested tumour segmentation on the baseline UNet model. Based on a 2D single slice of the MRI scan, we obtained tumour segmentation by categorising each pixel into the labels of  $\{0,1,2,3\}$ . As an accuracy metric for segmentation, we used **IoU (Intersection Over Union)** score. IoU score gives the overlap of the predicted segmentation mask with the actual segmentation mask (This is done classwise. So for 4-class segmentation, 4 IoU scores will be shown). In terms of true-labels and predicted labels (prediction of some label  $x$  versus the actual label  $x$ ), we can come up with another notion of IoU too.

$$IoU\ score = \frac{|A \cap B|}{|A \cup B|}$$
$$IoU\ score = \frac{TP}{TP + FP + FN}$$

The following table represents the IoU scores obtained solely through UNet when tested across the 4 modalities: T1-weighted, T2-weighted, T1-contrast-enhanced and Flair.

Modality	Class 0	Class 1	Class 2	Class 3
T1	0.982	0.209	0.303	0.105
T2	0.983	0.223	0.371	0.089
T1cE	0.985	0.435	0.385	0.467
Flair	0.991	0.067	0.472	0.113

Table 5.1: Classwise IoU scores for each modality

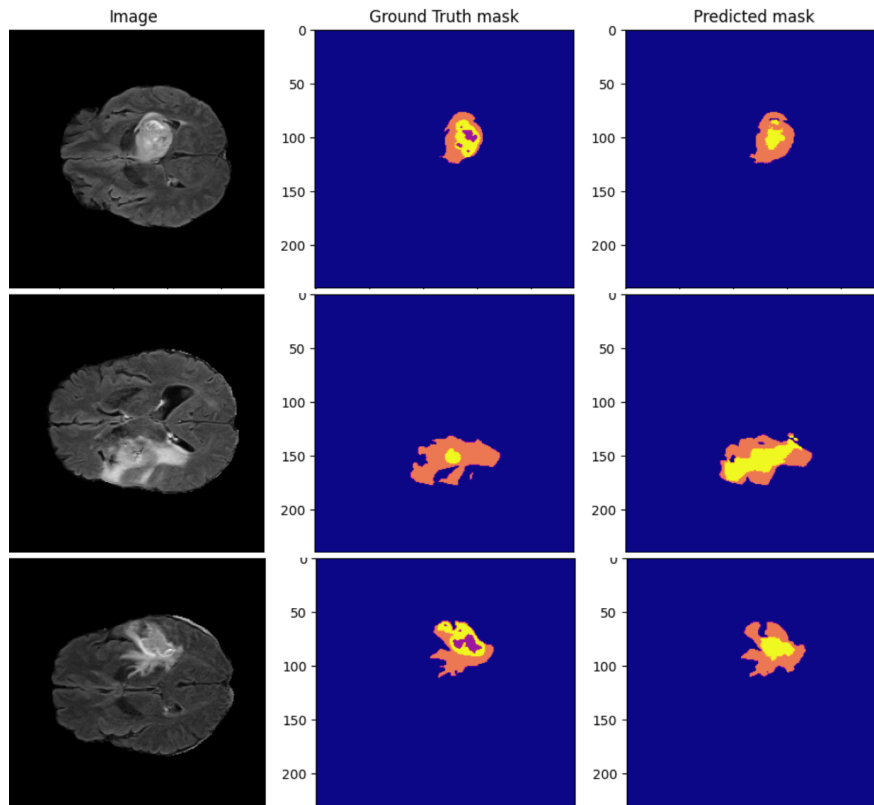


Figure 5.1: Tumour mask predictions on test images. Column 1 shows the actual image, followed by the **Ground truth mask** and the **Model predictions**



The above picture shows some examples of ground-truth segmentation masks and the masks predicted by the UNet model.

## 5.2 MTL joint segmentation-classification results

We experimented with 2 versions of the MTL model (as explained in section 4.1). One was with just the middle slice = 70 being used as the training set, and another where 10 slices were sampled across all the scans. To get the best possible extent of the nature of model discrimination, we performed the experiment with **T1-weighted** and **Flair** images separately. The results are summarized as:

Middle slice : T1-weighted				
$\lambda$	Mean IoU	AUC	Accuracy	Precision
0.1	0.735	0.526	0.51	0.55
0.3	0.714	0.531	0.45	0.48
0.5	0.719	0.512	0.46	0.52
0.7	0.702	<b>0.599</b>	0.50	0.58
0.9	0.695	0.541	0.55	0.60

Table 5.2: Performance of 2D middle slice MTL (T1)

Middle slice : FLAIR				
$\lambda$	Mean IoU	AUC	Accuracy	Precision
0.1	0.868	0.542	0.58	0.58
0.3	0.845	0.587	0.53	0.60
0.5	0.858	0.536	0.52	0.56
0.7	0.851	0.534	0.51	0.56
0.9	0.793	0.441	0.45	0.50

Table 5.3: Performance of 2D middle slice MTL (Flair)

10 sampled slice : Flair				
$\lambda$	Mean IoU	AUC	Accuracy	Precision
0.1	<b>0.898</b>	0.4712	0.46	0.47
0.3	0.890	0.516	0.53	0.52
0.5	0.882	0.497	0.51	0.50
0.7	0.846	0.548	0.56	0.53

Table 5.4: Performance of 2D 10-slice MTL (Flair)

As can be inferred, the highest AUC **0.599** is obtained for middle-slice MTL version for the **T1-weighted** images. The highest mean IoU score over 2 classes segmentation **0.898** is obtained for the **Flair** modality by sampling 10 slices.

### 5.3 Volumetric Projection results

We experimented with 3 kinds of projections: **Mean, Maximum and Standard Deviation** projections (section 4.2). The results of axis 1 (axial direction) for all 3 kinds of projections are depicted below:

3 kinds of projection (axial)			
Projection	AUC	Accuracy	Precision
Maximum	0.5416	0.53	0.51
Standard Deviation	0.5467	0.49	0.49
Mean	0.5312	0.51	0.51

Table 5.5: Projection Performance (axial direction)

It was observed that the Standard Deviation projection was giving the best results, and so SD projections was tested along all 3 axes: axial, coronal and saggital.

Standard Deviation Projection: 3 axes			
Axis	AUC	Accuracy	Precision
AXIAL	<b>0.5884</b>	0.50	0.43
CORONAL	0.4905	0.49	0.49
SAGGITAL	0.5058	0.49	0.49

Table 5.6: 3 Axis Performance (SD projections)

**EfficientNet B7** with **SD projection (along axial direction)** was the best performance with AUC of **0.5884**.

## 5.4 Cascaded cropped-mask-results

In the cascaded cropped segmentation-classification model (section 4.3), we experimented using 2 strategies: (a) With pre-determined slice number across train and test datasets (section 4.3.1) (b) With 5-stack volume by finding the slice with best tumour visibility (section 4.3.2)

**EfficientNet B3**, **EfficientNet B7** and **ResNet32** were used for the pre-determined slice version with **slice = 70** used as the unanimous slice number. The performances are summarized below:

2D slice cascaded model			
Model	AUC	Accuracy	Precision
EfficientNet B3	0.6250	0.47	0.47
EfficientNet B7	0.4783	0.47	0.48
ResNet34	<b>0.6342</b>	<b>0.57</b>	<b>0.53</b>

Table 5.7: Results of 2D cascaded model

We proceeded to the 3D 5-slice stacked volume version of the cascaded model, and the classifiers were changed as: **EfficientNet B0**, **EfficientNet B3**, **EfficientNet B7**, **ResNet18** and **ResNet34**. 5 fold cross-validation was utilized with **Stratified K-folds** strategy to keep the class label ratio similar along all the fold partitions. Henceforth, we used **5-fold-cross-validation** for each phase. The data mentioned in the tables mention the average AUC/Accuracy with the Standard Deviation obtained across all 5 folds.

5-slice 3D cascaded (single classifier)			
Axis	model	AUC	Accuracy
AXIAL	EffNet B0	$0.532 \pm 0.038$	$0.522 \pm 0.046$
	EffNet B3	$0.4942 \pm 0.034$	$0.506 \pm 0.021$
	EffNet B7	$0.4455 \pm 0.058$	$0.5099 \pm 0.017$
	ResNet18	$0.553 \pm 0.062$	$0.510 \pm 0.047$
	ResNet34	<b><math>0.641 \pm 0.187</math></b>	<b><math>0.620 \pm 0.150</math></b>
CORONAL	EffNet B0	$0.546 \pm 0.052$	$0.493 \pm 0.019$
	EffNet B3	$0.521 \pm 0.044$	$0.513 \pm 0.018$
	EffNet B7	$0.475 \pm 0.055$	$0.511 \pm 0.017$
	ResNet18	$0.640 \pm 0.190$	$0.608 \pm 0.182$
	ResNet34	$0.586 \pm 0.121$	$0.574 \pm 0.090$
SAGGITAL	EffNet B0	$0.490 \pm 0.086$	$0.511 \pm 0.018$
	EffNet B3	$0.505 \pm 0.054$	$0.511 \pm 0.017$
	EffNet B7	$0.509 \pm 0.060$	$0.488 \pm 0.017$
	ResNet18	$0.582 \pm 0.147$	$0.576 \pm 0.118$
	ResNet34	$0.575 \pm 0.183$	$0.551 \pm 0.150$

Table 5.8: Results of 5-slice 3D cascaded (single model)

The highest AUC of  $0.641 \pm 0.187$  is obtained for **ResNet34** in the **axial direction** (as highlighted in the tables). Various kinds of ensembles were also tested with, and in majority of the cases they gave better performances.

5-slice 3D cascaded (ensemble)			
Axis	model	AUC	Accuracy
AXIAL	ResNet18 + ResNet34	<b><math>0.736 \pm 0.160</math></b>	<b><math>0.662 \pm 0.148</math></b>
	ResNet18 + EffNet B3	$0.692 \pm 0.169$	$0.653 \pm 0.140$
	ResNet34 + EffNet B3	$0.713 \pm 0.160$	$0.662 \pm 0.148$
CORONAL	ResNet18 + ResNet34	<b><math>0.721 \pm 0.174</math></b>	<b><math>0.693 \pm 0.140</math></b>
	ResNet18 + EffNet B3	$0.680 \pm 0.199$	$0.637 \pm 0.172$
	ResNet34 + EffNet B3	$0.637 \pm 0.177$	$0.608 \pm 0.136$
SAGGITAL	ResNet18 + ResNet34	$0.707 \pm 0.202$	$0.684 \pm 0.154$
	ResNet18 + EffNet B3	$0.629 \pm 0.189$	$0.609 \pm 0.135$
	ResNet34 + EffNet B3	$0.715 \pm 0.146$	$0.671 \pm 0.104$

Table 5.9: Results of 5-slice 3D cascaded (ENSEMBLE)

For ensembles, the highest AUC  $0.736 \pm 0.160$  is obtained for the **ResNet18 and ResNet34 ensemble** in the **axial direction** and the highest Accuracy  $0.693 \pm 0.140$  is obtained for the same emsemble in the **coronal direction** (as highlighted in the tables).

## 5.5 3-axis Ensemble Results

The 3-axis-ensemble (section 4.3.3) uses the same MRI scan to sample slices across 3 different axis directions, and then trains separate UNet and classifiers for each direction.

3-axis ensemble		
Model	AUC	Accuracy
ResNet18	$0.711 \pm 0.198$	$0.665 \pm 0.171$
ResNet34	$0.721 \pm 0.231$	$0.700 \pm 0.205$

Table 5.10: Results of 3-axis Ensemble

As a conclusion across all models tested, we were able to attain the highest AUC **0.736** in **ResNet18 + ResNet34 ensemble** in the axial directional case. This surpasses the current Kaggle leaderboard AUC of abc.

AUC in the experimentation refers to Area under curve, which has been explained in Appendix A (5.5). AUC is calculated as the Area Under the ***Sensitivity***(TPR)-***(1 - Specificity)***(FPR) Curve. True Positive refers to the samples which were classified as '1' (methylated) and it was a correct classification, while FP refers to a '1' classification which was False (truth was unmethylated).

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

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

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

# Conclusion

The proposed end-to-end tumour segmentation pipeline can be used assistively to aid medical practioners in detecting and assessing the methylation of glioblastoma, in order to increase the chemotherapy benefits of TMZ. Since our model is sufficiently lightweight, it can save time in processing a large quantity of MRI scans. As limitations to the proposed approach, it should be noted that because the dataset size was quite small, an experimental bias could have been incurred. Further research could be done to build more nuanced models by combining the MRI modalities and give more emphasis on predicting the methylation status through a countinuos value (across all the gene markers) rather than limiting to a simple binary classification.



# Appendix A : Area Under Curve (AUC)

The ROC: Receiver Operating characteristic curve analyses discriminatory capability of any binary classifier in machine learning. Area Under the ROC Curve (AUC) has been widely used to judge the performance of classification models.

Normally, while using a Softmax layer in the last layer of any binary classification model, we get the output as the vector  $[x,y]$  where  $x$  and  $y$  represent the classification probability of the sample belonging to class 0 or class 1 respectively. Based on a threshold value (given), the activation function decides which class to assign. Thereafter when we calculate the Accuracy, Precision, Recall we are generally counting how many samples were classified correctly and how many were classified incorrectly. The count of assigned samples depends upon the threshold value assigned during classification.

$$\text{Sensitivity} = \text{Recall} = \frac{TP}{TP + FN}$$
$$\text{Specificity} = \frac{TN}{FP + TN}$$

AUC is calculated as the Area Under the **Sensitivity**(TPR) versus the  $(1 - \text{Specificity})$ (FPR) Curve. With an increasing sets of thresholds (0% to 100%) it calculates the TPR and FPR in each case, and plots a point in the AUC curve. When all the thresholds have been plotted, it calculates the integral area under the curve. AUC 0 signifies a model with 100% wrong predictions while AUC 1 means a model with 100% correct predictions.

# References

- A. Tosoni A. A. Brandes, E. Franceschi. “MGMT Promoter Methylation Status Can Predict the Incidence and Outcome of Pseudoprogression After Concomitant Radiochemotherapy in Newly Diagnosed Glioblastoma Patients”. *Journal of Clinical Oncology*, 26:2192–2197, 2008. 1
- H. Hess A. Akay. “Deep learning: current and emerging applications in medicine and technology”. *IEEE journal of biomedical and health informatics*, 23, 2019. 3
- M. R. Gilbert H. Colman C. De La Cruz E. P. Sulman B. N. Bekele K. D. Aldape A. L. Rivera, C. E. Pieloski. “MGMT promoter methylation is predictive of response to radiotherapy and prognostic in the absence of adjuvant alkylating chemotherapy for glioblastoma”. *Neuro Oncology*, 12:116–121, 2009. 1
- U. Baid. “The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification”. *arXiv*, 2021. 4, 7
- De Robles P Drabycz S, Roldan G. “An analysis of image texture, tumor location, and MGMT promoter methylation in glioblastoma using magnetic resonance imaging”. *Neuroimage*, 49, 2010. 3
- O. G. Kakde G. S. Tandel, M. Biswas. “A review on a deep learning perspective in brain cancer classification”. *Cancers*, 11, 2019. 3
- R. T. Leijenaar H. J. Aerts, E. R. Velazquez. “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach”. *Nature Communications*, 5, 2014. 2

## REFERENCES

- Gorlia T Hegi ME, Diserens AC. “MGMT gene silencing and benefit from temozolomide in glioblastoma”. *N Engl J Med*, 352:997–1003, 2005. 1
- Shaoqing Ren Jian Sun K. He, X. Zhang. “Deep Residual Learning for Image Recognition”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. vii, 16
- A. Zisserman K. Simonyan. “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION”. *ICLR*, 2015. 16
- Coufalova L Korfiatis P, Kline TL. “MRI texture features as bio- markers to predict MGMT methylation status in glioblastomas”. *Med Phys*, 43, 2016. 3
- J. M. Pisapia L. Macyszyn, H. Akbari. “Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques”. *Neuro-oncology*, 18, 2015. 2
- J. M. Eschbacher L. S. Hu, S. Ning. “Radiogenomics to characterize regional genetic heterogeneity in glioblastoma,”. *Neuro-oncology*, 19, 2017. 2
- C. Liu L. Wang, Z. Li. “Comparative assessment of three methods to analyze MGMT methylation status in a series of 350 gliomas and gangliogliomas”. *Pathology-Research and Practice*, 213:1489–1493, 2017. 2
- Quoc V. Le M. Tan. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”. *ICML*, 2019a. 16
- Ruoming Pang Vijay Vasudevan Mark Sandler Andrew Howard Quoc V. Le M. Tan, Bo Chen. “MnasNet: Platform-Aware Neural Architecture Search for Mobile”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019b. 17
- L. L. Morgan. “The epidemiology of glioma in adults: a ”state of the science” review”. *Neuro Oncology*, 17:623–624, 2015. 1
- P. Khong N. R. Parker, A. L. Hudson. “Intratumoral heterogeneity identified at the epigenetic, genetic and transcriptional level in glioblastoma”. *Scientific Reports*, 6, 2016. 2

## REFERENCES

- Thomas Brox O. Ronneberger, Philipp Fischer. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. *MICCAI*, 2015. 13
- R. Leijenaar P. Lambin, E. Rios-Velazquez. “Radiomics: extracting more information from medical images using advanced feature analysis”. *European Journal of Cancer*, 48, 2012. 2
- D. A. Reardon P. Y. Wen, D. R. Macdonald. “Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group”. *Journal of clinical oncology*, 28, 2010. 2
- H. Breman K. Fissell C. Haselgrove C. Holmes J. Lancaster D. Rex S. Smith J. Woodward R. Cox, J. Ashburner. “A (sort of) new image data format standard: Nifti-1: We 150”. *Neuroimage*, 22, 2004. 7
- W. M. Wells S. K. Warfield, K. H. Zou. “Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation”. *IEEE transactions on medical imaging*, 23, 2004. 7
- E. V. Sullivan A. Pfefferbaum T. Rohlfing, N. M. Zahr. “The sri24 multichannel atlas of normal adult human brain structure”. *Human Brain Mapping*, 31, 2010. 7
- G. Hinton Y. LeCun, Y. Bengio. “Deep Learning”. *Nature*, 521, 2015. 3
- X. Zhang Y. Yang, L. F. Yan. “Glioma grading on conventional MR images: a deep learning study with transfer learning”. *Frontiers in Neuroscience*, 12, 2018. 3