People's Democratic Republic of Algeria
الجمهورية الجزائرية الديمقراطية الشعبية
Ministry of Higher Education and Scientific Research
وزارة التعليم العالي و البحث العلمي
Kasdi Merbah University of Ouargla

Dissertation Submitted to the Department of Computer Science and Information Technology in Candidacy for the Degree of Master LMD in Industry Artificial intelligence and data science branch
Master degree thesis

Students :

Sirine Bouldjemar

Chouaib Bammoune

Deep learning based 3D brain tumor segmentation

Adviser : Dr. Khaldi bilale
Co-adviser : PhD.Saber Chacha

Aknowledgments

We wish to thank first and foremost our Allah, the Almighty, the greatest of all.We would like to express our sincere gratitude to the following individuals and organizations that have contributed to the completion of our master's degree thesis: First and foremost, we would like to express our deepest appreciation to our thesis advisor, Dr. Khalidi Bilal, and co-advisor, Saber Chacha, for their guidance, expertise, and unwavering support throughout this research journey. Their valuable insights, constructive feedback, and dedication have played a pivotal role in shaping this thesis. We are also grateful to the faculty members and staff of FNTIC at Kasdi Merbah University for providing a conducive academic environment and resources that enabled us to carry out our research effectively. We would like to acknowledge the support and encouragement from our families, especially our parents and siblings, for their love, patience, and understanding. Their constant support and belief in us have been a constant source of motivation. We extend our appreciation to our friends and colleagues who have provided assistance, insightful discussions, and encouragement throughout this thesis undertaking. Their camaraderie and intellectual exchange have been invaluable. Furthermore, we are thankful to the research participants who generously shared their time, knowledge, and experiences, without whom this study would not have been possible. In conclusion, the completion of this master's degree thesis would not have been possible without the support, guidance, and contributions of all these individuals and organizations. We are deeply indebted to each and every one of them.

Sirine Bouldjemar
Chouaib Bammoune

Abstract

Brain tumor segmentation plays a crucial role in medical image analysis and assists in the diagnosis, treatment planning, and monitoring of brain tumor patients. However, accurately segmenting brain tumors from multi-modal medical images remains a challenging task due to the complex and heterogeneous nature of tumors, tissues. In this thesis, we propose a 3D U-net deep learning model for the purpose of having accurate 3D brain tumor segmentation from multi-modality data to take advantage of different levels of information that exist in the MRI sequences, in addition to reducing the diagnosis time and having an automated process to address the problem from like-real data and scale down the human bias when dealing with such a sensitive task. The 3D U-net model is trained on the BRATS2020 data-set and evaluated with segmentation volumetric metrics. The model showed promising results on the majority of the test data 63 % that reached high IOU score for the whole tumor region and showed a tolerance to variation after applying the test on noisy data. However, it miss classified the 37 % of the test data because of the imbalance between classes. as a recommendation to tackle this problem; a customized approach to deal with imbalance conditions can be proposed.

Key words: MRI Multi-modals, segmentation, Deep learning and U-net architecture.

ملخص

يلعب تجزئة ورم الدماغ دورًا حاسمًا في تحليل الصور الطبية ويساعد في التشخيص والتخطيط للعلاج ومراقبة مرضى أورام المخ ، ومع ذلك ، يظل تقسيم أورام الدماغ بدقة من الصور الطبية متعددة الوسائط مهمة صعبة بسبب الطبيعة المعقدة وغير المتجانسة من الأورام. في هذه الأطروحة ، نقترح نموذج التعلم العميق ثلاثي الأبعاد لغرض الحصول على تجزئة دقيقة لورم الدماغ ثلاثي الأبعاد من بيانات متعددة الوسائط للاستفادة من المستويات المختلفة من المعلومات الموجودة في تسلسل التصوير بالرنين المغناطيسي ، بالإضافة إلى تقليل التشخيص الوقت ووجود عملية آلية لمعالجة المشكلة من البيانات الحقيقية المماثلة وتقليل التحيز البشري عند التعامل مع مثل هذه المهمة الحساسة. يتم تدريب نموذج ثلاثي الأبعاد على مجموعة بيانات وتقييمه باستخدام المقاييس الحجمية للتجزئة أظهر النموذج نتائج واعدة على غالبية بيانات الاختبار التي وصلت إلى درجة 63 ٪ لمنطقة الورم بأكملها وأظهرت تحملاً للتباين بعد تطبيق الاختبار على البيانات الصاخبة. ومع ذلك ، فقد تصنيفها 37 ٪ من بيانات الاختبار بسبب عدم التوازن بين الفئات. كتوصية لمعالجة هذه المشكلة ؛ يمكن طرح نهج مخصص للتعامل مع ظروف عدم التوازن.

الكلمات الرئيسية: متعدد الوسائط ، التجزئة ، التعلم العميق و U-net معمارية.

Table of content

List of Figures

List of Tables

List of abbreviations

- **MRI:** Magnetic resonance imaging.
- **CT-scans:** Computed Tomography scans.
- **CAD:** Computer-Aided Design.
- **AI:** Artificial intelligence.
- **DL:** Deep learning.
- **CNN:** Convolutional neural network.
- **CSF:** Cerebrospinal Fluid.

General introduction

Cancer is one of the most deadly and pervasive diseases in the world, and early detection can help prolong survival and boost survival rates. CT or MRI imaging are the most commonly used clinical auxiliary approaches for cancer diagnosis. In despite, due to the unequal regional distribution of medical resources, physicians with the diagnostic capabilities are also unequally distributed. The distribution of cancer patients in rural areas is erratic, increasing the workload of physicians. However, due to the complexity of the medical image and the high accuracy requirements of the segmentation result, physicians must perform extensive analysis for an extended period of time. This discrepancy has emerged as one of the most pressing issues in cancer diagnosis around the world, and a high-precision automatic segmentation model is needed to reduce physician workload and increase work efficiency[**1**].

Gliomas are the most frequent central nervous system (brain and spinal cord) tumors. They are caused by the uncontrollable growth of glial cells and are categorized as astrocytomas, ependymomas, and oligodendrogliomas. The most dangerous kind is glioblastoma (an astrocytoma). Gliomas can afflict people of any age, although they are more frequent in adults. They can cause symptoms by compressing or invading neighboring tissues and raising intracranial pressure.[**2**].

A complete history, neurologic exam, and neuroimaging scans are the starting point for addressing a probable glioma patient. A multidisciplinary approach to patient-centered care is advised.

Several scientific investigations have found that age, sex, genetic inheritance, radiation, food, viral infections, and stress all have varying degrees of relationship with the formation of gliomas. However, the precise reason remains unknown.

Gliomas are classified into subtypes, some of which are benign and others of which are malignant. They are classified based on the cells that give birth to them (i.e., astrocytes, oligodendrocytes, and ependymal cells). A further categorization in WHO Grades I–IV based on cell architecture is possible. The grade of a tumor is related to its aggressiveness. Observation, surgery, radiation, chemotherapy, or a combination of these are the treatment options. Treatment is chosen based on baseline patient and tumor characteristics. Lower-grade cancers have better results in general. As a result, the presence of a multidisciplinary team that collaborates with the patient to choose the optimum treatment plan is critical.

A tumor's grade is defined by how its cells appear when examined under a microscope. Lower grade, slower-growing cancers had a longer median survival time than more aggressive, faster-growing tumors. The least invasive and most likely to survive are cancers of grade I. Lesions in grade II start out as slow-growing masses but can advance to higher grades. Malignant grade III tumors have a rapid pace of growth, are more prevalent in adults, and are malignant. The worst prognosis is associated with grade IV gliomas, which are the most aggressive, quickly developing, and invasive[**2**].

- Low-Grade Glioma.
- High-Grade Glioma.

The brain tumor diagnosis requires qualified surgeons to manually contour the tumor regions from 2D MRI slices or using other systems to provide 3D view for the brain, this process requires a consensus of doctors following medical protocols to settle on an approved and accurate tumor segmentation. However this diagnosis is a time consuming process especially with hundreds of brain tumor cases, when time is a key factor for the treatment. For the sake of brain tumor patients, we develop an approach that process the targeted data and have the ability to extract relevance information that will lead to an approved tumor segmentation, taking in consideration reducing the diagnosis time by exploiting both recent powerful technologies and Medical imaging which are included in the experiment process. In addition to that, the biomedical image analysis took a real interest in the research and artificial intelligence community, hence providing promising results and

conclusions in this task will contribute to the understanding of this problem and harnessing specialized tools to build an automated and reliable systems to handle the medical image problems[**3**].
Our thesis is organized as follows:

- The first chapter **"Background"** presents the prior knowledge and overviews about the problem.
- The second chapter **"Related works"** presents the challenges that occur when dealing with brain tumor segmentation, in addition to the state of the art and how previous researchers tackled this task.
- The third chapter **"Method"** presents the proposed approach to deal with the brain tumor segmentation with explaining the different employed techniques and their impact on having accurate results.
- The last chapter **"Experimental results"** shows the final outcome of the proposed architecture and the performance of the model based on the metrics used for this type of problems. Besides to the environment and the work place adopted during the implementation. Finally we disclose with the limits and the possible future improvements that can be done to achieve more precised segmentation.

Chapter 1

Background

1.1 Introduction

Brain tumor segmentation is one of the most critical challenges in the technology field. Shedding the light on this problem would represent a huge step forward for the fields of medicine and oncology. The use of AI techniques offers accurate border detection and avoids human bias. Additionally, the 3D representation of the tumor cells spreading among the brain cells in the final phase of the experiment provides a comprehensive view. However, before delving into the technical details, this chapter will introduce some basic definitions and background.

1.2 Digital image

A digital image is a numerical representation of real-world pictures and views captured by a digital camera. The image is represented by a matrix with two dimensions for 2D images or three dimensions for 3D images. When the matrix dimension is (N x N), the matrix elements are called pixels, and when the dimension is (N x N x N), the matrix elements are called voxels. Pixels/voxels are numerical binary values that represent the intensity of color in the image. There are three types of image representation based on color:

1. **Grayscale images:** Each pixel or voxel has a single value that represents the brightness level of the image.

2. **Color images:** Each pixel or voxel has three or more values that represent the intensity of the red, green, and blue (RGB) colors or other color spaces.

3. **Multispectral images:** Each pixel or voxel has values that represent the intensity of light at various wavelengths, providing additional information about the image.

Each type represent a different level of information and used for different tasks depending on the problem.

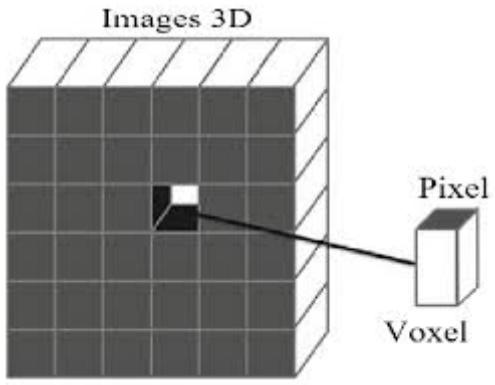

Figure 1.1: Representation of pixels and voxels

Medical 3D images are presented on a grid that can have different sizes depending on the body part image and the resolution. The grid size is denoted by (W x H x D), denoting the width, height, and depth of the 3D image.[**4**].

1.2.1 3D image vs 2D image

It is worth noting that 3D medical images outperform 2D images in medical image analysis, especially when using deep learning, for several reasons. Some of the most common reasons include:

1. **Improved accuracy:** 2D images present a flat format of the body part in question, which lead to noticeable loss of information. Unlike 3D images that give real view or modeling and give access to a complete set of features, the implementation of this option yields a favorable results.

2. **Reduced noise:** 3D medical images typically have less noise than 2D image as they contain more information about the underlying structure or tissue.

3. **Consistency across slices:** 3D medical images provide a more consistent representation of the structure or lesion across slices, which can improve model performance.

4. **Reduced manual intervention:** When it comes to annotation, it is difficult to contour lesion across slices in case of 2D medical images and could require significant manual intervention. Unlike 3D images case when annotation is easy and consistent, as the annotation can be applied on the entire volume.

At last, 3D medical images offer a variety of advantages when processing medical analysis tasks with deep learning. In addition, 3D images provide a more comprehensive view of the anatomy and pathology of the body than 2D images, which can be limited in their ability to capture the full complexity of the human body from different angles. However, it's important to note that working with 3D images can also introduce some challenges, such as increased computational requirements and the need for specialized software tools.

1.2.2 Medical image

The images we see nowadays are produced based on different energy sources. The principal energy source for images in use today is the electromagnetic energy spectrum. Other important sources of energy used for producing images include acoustic, ultrasonic, and electronic. Additionally, synthetic images used for modeling and visualization can be generated by computers [**4**].

Medical images are digital images that are captured by various medical equipment such as MRI and CT scans. These types of equipment use different wavelengths to capture images, with each wavelength showing a different level of tissue or information. Figure 1.2 shows examples of medical images.

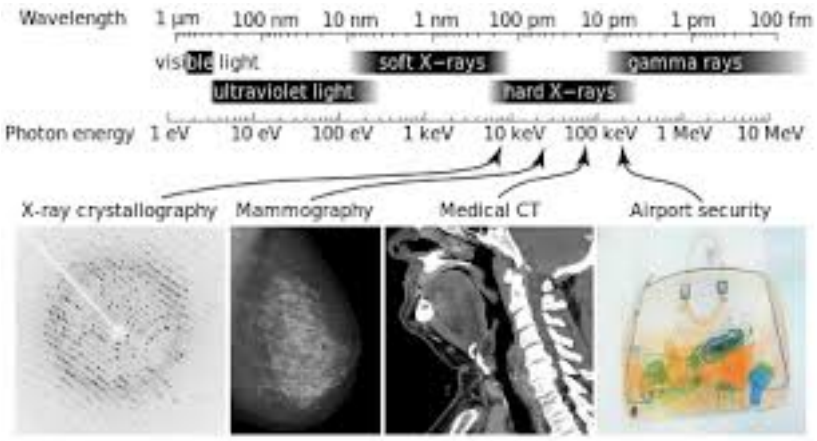

Figure 1.2: wavelength and there different medical images

1.3 Segmentation

Image segmentation is a sub-domain of computer vision and digital image processing that aims to group similar regions or segments of an image under their respective class labels. Since the entire process is digital, the representation of the analog image in the form of pixels is available, making the task of forming segments equivalent to grouping pixels.

Image segmentation is an extension of image classification, where in addition to classification, we perform localization. As such, image

segmentation is a super-set of image classification, with the model pinpointing where a corresponding object is present by outlining its boundary [**4**].

1.3.1 Semantic segmentation

Semantic segmentation refers to the classification of pixels in an image into semantic classes. Pixels belonging to a particular class are simply classified to that class with no other information or context taken into consideration.

However, this task can be challenging when there are closely grouped multiple instances of the same class in the image, as it can result in poor definition and provides very little in-depth details or information about the image.

1.3.2 Medical segmentation

Medical image segmentation is the process of extracting areas of interest (ROIs) from medical image data, such as CT or MRI scans. The primary objective of this process is to locate the anatomical regions required for a particular study, such as simulating physical attributes or accurately placing implants with CAD designs (Computer Aided Design) inside patients. Recent advancements in AI techniques are simplifying tasks such as medical image segmentation, which can be a time-consuming process.

1.3.3 The benefits of medical image segmentation

The primary goal of segmentation is to identify the anatomical regions required for a given study, such as implant design or simulating physical properties. One of the key advantages of medical image segmentation is that it enables a more precise examination of anatomical data by isolating only the relevant areas. Segmentation also facilitates the removal of unwanted scan components, such as air, and allows for the differentiation of various tissues, including bone and soft tissues. When combined with various software processing options, researchers and physicians can generate multiple segmented masks that are ready for further analysis.

1.3.4 Applications and impacts

1. Increase the speed of tumor diagnosis.
2. Identify abnormalities that are undetectable by the naked eye.
3. Early diagnostic.
4. Assist other deep learning methods to improve accuracy (preprocessing).

1.4 Image segmentation before deep learning

Image segmentation originally started in digital image processing coupled with optimization algorithms. These primitive algorithms made use of methods like region growing and snakes algorithm where they set up initial regions, and the algorithm compared pixel values to gain an idea of the segment map. These methods took a local view of

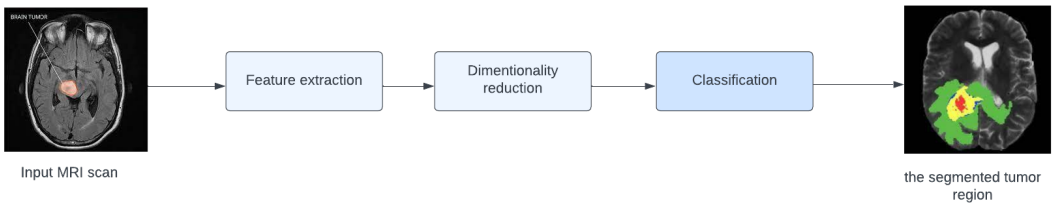

Figure 1.3: Machine learning pipeline for segmentation

the features in an image and focused on local differences and gradients in pixels [**4**].

Algorithms that took a global view of the input image came much later, with methods like adaptive thresholding, Otsu's algorithm, and clustering algorithms being proposed among classical image processing methods. However, the use of machine learning methods in complex segmentation problems comes with limitations. The process of feature engineering is entirely done by human intervention, with parameters such as filter size tuned to extract different levels of information to provide a feature space that can be fed to a machine learning algorithm for pixel-wise classification. However, this process is entirely dependent on the human perspective and their idea of the suitable feature space that will lead to promising results.

The extraction of feature space to offer relevant information to train a model presents a sensitive question about the quality and quantity of data provided by human intervention. The task of adding more and generalizing accurate data will lead to the curse of dimensionality. The solution is to seek techniques and algorithms that provide a more general representation of the data used for training.

1.5 Deep learning for image segmentation

The idea behind using DL for image segmentation is to have an automated feature extraction system with no human intervention at all as shown in 1.4, with more general and accurate information. To illustrate the point, consider these listing [**5**, **6**, **7**]:

1. **Ability to learn complex features :**

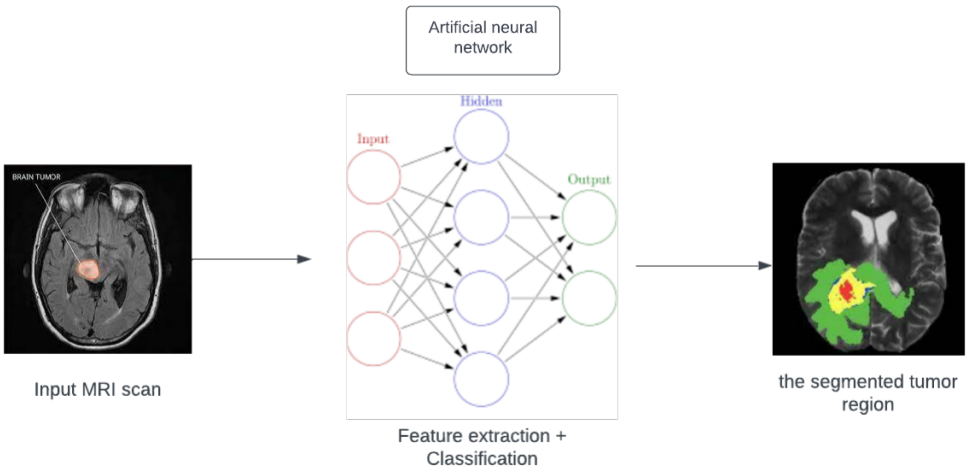

Figure 1.4: Deep learning learning pipeline for segmentation

thanks to the learn-able parameters in the convolution operation (illustrated in the convolution block section), Convolution neural networks can learn automatically complex features from raw data without the need of an explicit feature engineering. In contrast to machine learning, where this task can be time consuming and error prone.

2. **Robustness to variation :**
   Due to the hierarchical learning approach that consists of learning features at multiple levels of abstraction (starting with low level features such as edges and corners and gradually building up to high level features as texture and shapes)[**4**] which is originally inspired from the primary cortex in the human brain [**8**]. This propriety specifically is needed when working on multi-modal segmentation where different imaging parameters are selected to generate various MRI sequences. For the same reason this kind of data can miss-lead a machine learning algorithm because of notable variation in the multi-modal data.

3. **Scalability :**
   Deep learning models can be scaled to handle large datasets with

millions of samples, specifically this propriety is a crucial option when addressing a sensitive task as medical image analysis problems in parallel with the availability of these kind of data from research centers and medical institutions.

4. **State of the art :**
   Previous papers and biomedical image analysis searches dealt with in comparison between the machine learning and deep learning and the outcome confirmed the above points.

In deep learning, a CNN is a class of deep neural networks that are typically used to recognize patterns present in images; they are also used for spatial data analysis, computer vision, natural language processing, signal processing, and various other purposes. The architecture of a convolutional network resembles the connectivity pattern of neurons in the human brain and was inspired by the organization of the visual cortex. This specific type of artificial neural network gets its name from one of the most important operations in the network: convolution.

1.5.1 CNN

One of the important types of Deep learning algorithms that are used to handle computer vision problems is CNN, this specific type uses what known as a stack of convolution blocks to extract features and treat a wide range of computer vision tasks. Convolution is an orderly procedure where two sources of information are intertwined. Convolutions have been used for a long time, typically in image processing to blur and sharpen images but also to perform other operations (e.g., enhance edges and emboss). CNNs enforce a local connectivity pattern between neurons in adjacent layers. CNNs make use of filters (also known as kernels) to detect what features, such as edges, are present throughout an image. There are four main operations in a CNN:

1. Convolution
2. Non Linearity (ReLU)

3. Pooling or Sub Sampling

4. Classification (Fully connected layer)

1.5.2 Convolution opertion

The first layer of a Convolutional Neural Network is always a Convolutional Layer. Convolutional layers apply a convolution operation to the input, passing the result to the next layer. A convolution converts all the pixels in its receptive field into a single value. For example, if you would apply a convolution to an image, you will be decreasing the image size as well as bringing all the information in the field together into a single pixel. The final output of the convolutional layer is a vector. Based on the type of problem we need to solve and on the kind of features we are looking to learn, we can use different kinds of convolutions [**4**].

The most common type of convolution that is used is the 2D convolution layer, which is usually abbreviated as conv2D. A filter or a kernel in a conv2D layer "slides" over the 2D input data, performing an element-wise multiplication. As a result, it will sum up the results into a single output pixel. The kernel will perform the same operation for every location it slides over, transforming a 2D matrix (image) into a different 2D matrix of features[**4**]. In the case of a 3D image, the same process is applied with a kernel of (n*n*n) size; henceforward,the receptive field becomes a grid, and the pixels are called voxels.
The filter size plays a very important role in the quality of the extracted features, where large-sized kernels tend to find global features (such as tumor location and size) with a large receptive field and small kernels tend to contain local features (such as boundary and texture) with a small receptive field, as illustrated in the figure.1.5.

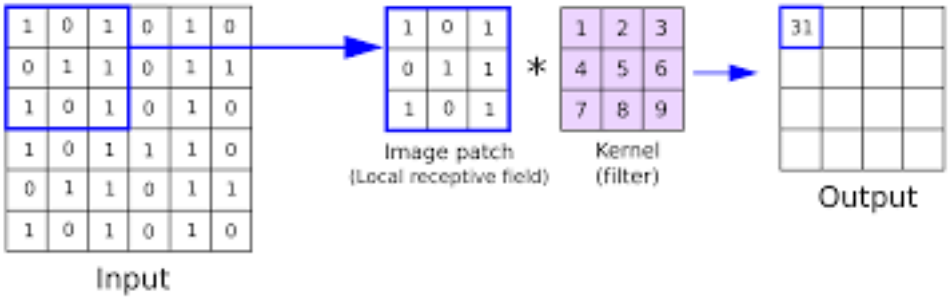

Figure 1.5: Convolution operation

1.5.3 Non linearity

After completing the convolution operation and resulting feature maps, the next step is applying a non linear activation function for the reason that the non linearity concept improve neural networks by speeding up training [**4**]. Depending on the sign of x, the gradient computation is relatively straightforward (either 0 or 1). Moreover, a ReLU's computation phase is simple: all negative elements are set to 0.0, requiring no use of exponential functions or multiplication or division operations. "$RelU(x) = max(0, x)$" Despite that, the ReLU function discards the negative values, which can hold some information. This is why a new activation function is proposed by[**9**] called 'Leaky ReLU', "$Leaky - ReLU(x) = max(0, x) + min(0, x)$", $where$" is a predetermined parameter that can also be learned.[**10**].

1.5.4 Max Pooling

This step consists of replacing a (n x n) region by the max value within[**4**], it is also called the max filter. This operation is done for the purpose of:

1. Choosing the highest activation in a local region, thereby providing a small degree of spatial invariance.

2. It reduces the size of the activation for the next layer by a factor of n^2. With a smaller activation size, a smaller number of parameters need to be learned in the later layers.

There are other types of polling such as global-average-pooling, winner-takes-all-pooling and stochastic-pooling [**11**, **12**].

1.5.5 Fully connected layer

The final phase in the CNN architecture is the fully connected layer or a sub neural networks that has the right to decide or make segmentation after getting features from the previous convolution blocks.

1.6 Conclusion

In this chapter, we have explored the fundamental concepts and background of medical image segmentation in the context of deep learning. We started by discussing the availability of digital images and their properties, which lead to various tasks such as semantic segmentation, and then we discussed medical image segmentation as a crucial task in various clinical applications, enabling accurate delineation of anatomical structures and pathological regions within medical images. We then delved into the foundations of deep learning, highlighting its revolutionizing impact on image segmentation and why we prioritize it over machine learning. Furthermore, we showed the basic and fundamental concepts of deep learning and what a deep learning algorithm is composed of for both neural and CNN architectures. Overall, this chapter provided a comprehensive overview of the background of medical image segmentation and deep learning. It laid the foundation for understanding the upcoming chapters, which will discuss more advanced titles.

Chapter 2

Related works

2.1 Introduction

Magnetic resonance imaging is one of the most common imaging methods used before and after surgery and is used to diagnose a wide range of medical conditions, such as joint and bone disorders, cardiovascular disorders, and brain and nervous system disorders. Aiming at providing fundamental information for the treatment plan. The availability of multi-modal MRI data motivated many deep learning researchers to develop techniques and approaches that are discussed in this chapter for medical image segmentation, which plays an active role in diagnosis and treatment. An accurate segmentation mask may help surgery planning and improve survival rates[**13**].

2.2 Research challenges

It is to mention the challenges that are held by brain tumor segmentation that still state-of-the-art and deep learning methods experience. Due to the wide spatial distribution of tumorous cells and their gluey nature, a location uncertainty issue arises, especially when dealing with cells that may appear at any location in the brain. The morphological uncertainty is also a huge challenge. Despite being a rigid object, the morphology (shape and size) of different brain tumors varies with such large uncertainty that it barely provides any prior information for describing the tumor shapes and structures. High-resolution and high-contrast images are recommended to get very accurate image seg-

mentation while containing diverse image information. However, it is not the case with MRI images with low quality and low contrast. Due to the image projection and tomography, the boundary between biological tissue tends to be blurred and hard to detect.[**14**]. In addition to the biological nature and imaging conditions, the manual annotation (producing the ground truth mask) presents a real challenge, which is the annotation bias because of the individual experience. This particular point would have a notable impact on the segmentation algorithm during the learning process.

2.3 Related works

To have a better understanding of the existing work and research in deep learning-based brain tumor segmentation and a systematic study, it is better to categorize the existing work based on the task and the purpose of each DL network. The figure 2.1 shows a comprehensive taxonomy that includes designing effective segmentation networks in the section2.3.1, dealing with imbalanced conditions in the section 2.3.2, and finally learning from multi-modality in the section 2.3.3.

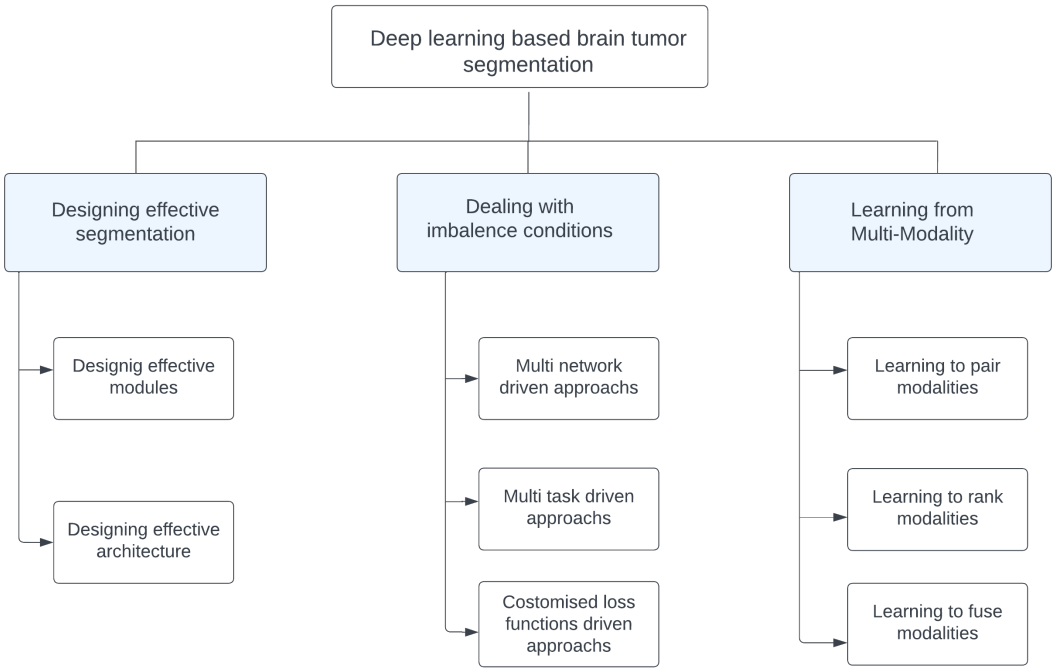

Figure 2.1: Taxonomy of deep learning based brain tumor segmentation methods.

2.3.1 Designing effective segmentation modules and architectures

Designing effective segmentation modules and architectures consists of using different deep learning techniques and even different neural network concepts and structures to extract high discriminitive data features for the sake of achieving accurate performance for the tumor segmentation.

We can see through the literature developments that affect the design of deep learning in the transition from single-channel networks [**15**] to multi-channel networks [**16**], from networks with fully connected layers to CNN. As the network goes deeper, the ability of the deep learning model to learn relevance features will be enhanced. One of the most common ways to deepen the neural network is by stacking convolution blocks with different kernel sizes; this will get the network to learn complex features gradually when moving from the input side to the output side of the network.

As the network goes deeper, it will face the vanishing or exploding gradient issue, which can be solved by the residual connections technique, which helped learn meaningful representation of the data. This solution is expanded also in [**17**] as dense connections that gave the segmentation deep learning algorithm a huge step forward in different aspects as follow [**18**]:

- Improved gradient flow.
- Feature reuse and combining.
- Parameters efficiency.
- Enhanced information flow.
- Regularization and over-fitting reduction.

Other convolution methods, such as dilated convolution, that preserve the spatial resolution and enlarge the receptive field with the introduction of additional parameters to capture large area targets as edema [**19**]. Also, there is the anistropic convolution, which is employed in [**20**] that helped in extracting features that aligned with dominant orientations in the image and allowed the kernel to be more sensitive to specific orientations.

When mentioning designing networks for accurate segmentation, encoder-decoder based algorithms take the lead when it comes to getting accurate feature-to-label mapping and localization of the tumor region, as in [**21**] which adopted the U-net that used the skip connection method to help layers in the network in contracting path recover details. Other research [**22**, **23**] proposed brain tumor segmentation based on U-net with a change in loss function, setting the padding to zero to avoid over-fitting and preserve the identical output, respectively.

2.3.2 Segmentation under imbalanced conditions

Data imbalance is defined as the number of pixels in the sub-region of the brain tumor that is totally different from patient to patient, in addition to the class imbalance within tumor classes. At the same time, labeling biases that are introduced by manual experts are also counted as data imbalances that will affect the false-positive rate.

Inspired by expert systems and ensembles of trees, the multi-network approach has come with the same principle that even complex models and architectures [**19**] can learn the best discriminative features, yet a single network cannot be protected from the data imbalance.An example of a network ensemble is a network cascade, a series of top-down networks with the principle that the output of the upstream network is the input of the downstream network. Where the upstream network extracts features (a coarse-to-fine strategy) and the downstream network subdivides the input to achieve a fine-grained segmentation. As proposed in [**24**] a cascade network of three layers (streams) of anisotropic and dilated convolution is combined with multi-view fusion to reduce the false-positive rate.

Another solution for the imbalance conditions in brain tumor segmentation is the multi-task-driven approach that divides the whole segmentation task or the task in question into other related tasks, as in [**25**, **26**], which splits the brain tumor segmentation into different sub-region segmentations, i.e., segmentation of the whole tumor, tumor core, and enhancing tumor individually. One further multi-task approach was discussed in [**27**, **28**] that introduced localization, or first detection, then segmentation, to perform precise tumor segmentation. In addition, [**29**] manages three auxiliary tasks, among them reconstruction, edge segmentation, and patch comparison. The authors consider the auxiliary task to be regularization of the main brain tumor segmentation.

2.3.3 Learning from multi-modalities

In this section, we tackle the task of learning from a complete multimodality for the purpose of a better understanding of the brain tissue and lesion and make use of the whole available BRATS dataset. The designed work can be categorized as learning to rank, learning to pair,

and learning to fuse modalities, as shown in the figure. In learning to rank, the data is sorted by relevance, as in [**30**], where the features are extracted from different embedding modalities and the relationship between modalities and the segmentation of different tumor sub-regions are modeled, so that the data of distinct MRI sequences are weighted and ordered accordingly to individual tasks. In addition to that, a cross-modality feature is employed to pair between modalities, also processed in [**30**]. Another aspect is to fuse modality by feature concatenation from each modality and send it to a downstream classifier, as in [**30**]. Although concatenation and addition are used, these two Fusion methods do not change the semantics of learned features and cannot highlight or suppress features. To tackle this problem, many research studies in recent years have adopted attention mechanisms to strengthen the learned features. [**31**, **32**, **33**] used a spatial and channel attention-based fusion module. The proposed attention mechanism highlights useful features and suppresses redundant features, resulting in accurate segmentation. In the case of missing modalities scenarios, [**34**] learned the explicit relationship between modalities and examined all possible missing scenarios. The results show that multi-modality has an important influence on accurate segmentation. As [**35**] proposed an intensity correction algorithm for different cases to distinguish the tumor and non-tumor regions in synthetic data.

2.4 Summery

Designing effective modules with good results and small amounts of data to understand the learning process and performance is encouraging work. However, the need to design comprehensive architectures to deal with large-scale data remains necessary, especially since the deeper the neural network, the more problems emerge that deserve treatment and solutions that do not affect the accuracy of the network or the sensitivity of the task. While using network ensembles and cascades shows good results in decreasing the false negative rate, the output of the downstream network is completely dependent on the upstream output, in addition to the memory and computation resources needed for the training. Learning from multi-modality is a great as-

pect when most of the research works model the implicit ranking while learning the modality-aware feature; however, existing pair works show modality pairing through exhausting combinations with large computing resources. In the fusion of modalities, addition or concatenation does not introduce additional parameters, despite the lack of physical expression of features. the need for a process to combine modalities to help the back propagation phase learn features at different levels and relate them to build feature awareness for the available MRI modalities in the BRATS data set.

2.5 Table description

We categorize the methods based on their main contributions. In column **Input**, 'P' means patch and 'I' means image. 'Dim' means the dimension of the network. Column **Dataset** indicates the the year of the BRATS data release which the paper worked on. The **DCS** column indicates the dice similarity metric value for the whole tumor segmentation.

Aspects	**paper**	**Input**	**Dim**	**Dataset**	**DCS**
Designing effective modules and architectures	*Gaffari et al.2020* [**36**]	I	3D	2020	0.9
	Yuan et al.2020 [**37**]	I	3D	2020	0.91
	Henry et al.2020 [**38**]	I	2D	2020	0.91
	Rehan et al.2022 [**39**]	-	3D	2020	0.86
	Yuan et al.2022 [**40**]	-	3D	2020	0.89
Dealing with imbalanced conditions	*Jia et al.2020* [**41**]	I	3D	2020	0.91
	Cirillo et al.2020 [**42**]	I	3D	2020	0.89
	Weninger et al.2019 [**43**]	P	3D	2019	0.85
	Liu et al.2020 [**44**]	I	2D	2020	0.88
	Huang et al.2020 [**45**]	I	3D	2020	0.9
Learning from Multi-Modality	*Zhang et al.2020* [**30**]	P	3D	2020	0.9
	Zhou.2020 [**34**]	I	3D	2020	0.88
	Yi et al.2018 [**35**]	I	3D	2018	0.91
	Li et al.2019 [**31**]	P	3D	2019	0.9
	Islam et al.2020 [**32**]	I	2D	2020	0.88

Table 2.1: State of the art Table

2.6 Conclusion

In this chapter, we conducted a comprehensive review of the related work in deep learning-based brain tumor segmentation, in addition to the existing aspects and purposes of the works reviewed. The aim was to explore the advancements, methodologies, and key findings from existing studies in this field. By synthesizing and analyzing a wide range of research papers, we gained valuable insights into the current state-of-the-art techniques, challenges, and future directions in deep learning-based brain tumor segmentation.

The reviewed literature showcased the tremendous progress made in recent years, highlighting the effectiveness of deep learning models in accurately segmenting brain tumors from medical images. Various architectures, including U-Net, cascade networks, and 3D CNNs, have been successfully employed to tackle this task. These models leverage the hierarchical features learned from large-scale data-sets to achieve high segmentation accuracy, capturing both the spatial and contextual information present in brain tumor images.

While significant progress has been made, several challenges remain in deep learning-based brain tumor segmentation. Class imbalance, dealing with small tumor sub-regions, handling multi-modal imaging data, and addressing interpretability, uncertainty estimation, and morphological nature are among the key areas that require further exploration and development.

In conclusion, reviewing the literature gave us insights about the approach we are working on and a clear representation of the existing challenges, which by solving them would make significant contributions to clinical practice by assisting radiologists and oncologists in accurate diagnosis, treatment planning, and monitoring of brain tumor patients.

Chapter 3

Method

3.1 Introduction

In this chapter, we will introduce the data used for multi-modal brain tumor segmentation, in addition to explaining the techniques used to build the deep learning architecture that can produce insightful results.

3.2 Methodology

In order to organize the process of our study, we must follow a certain methodology that defines the different phases and stages of our project. A well-known method CRISP-DM stands for cross-industry process for data mining. It provides an Structured approach to planning a data exploration project. By CRISP-DM Divide the project into six stages, as shown in the **figure 3.1**.

3.2.1 Data Description for Imaging

The BraTS2020 dataset is a well-known and widely used dataset in medical imaging research, particularly in the field of brain tumor segmentation and classification. It is made up of multimodal MRI im-

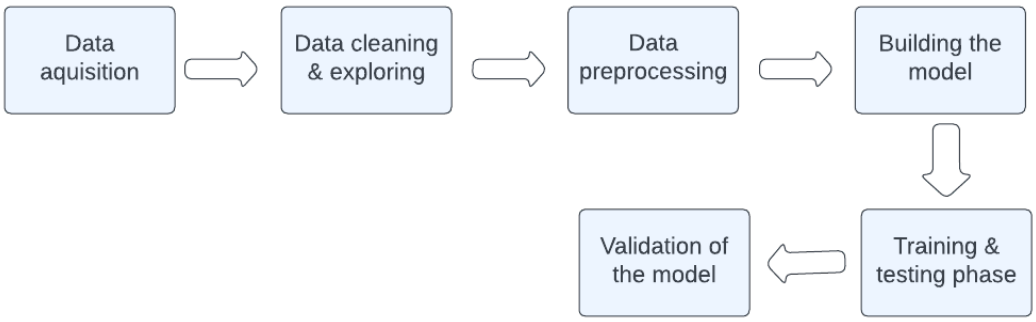

Figure 3.1: The methodology and the development process adopted by a artificial intelligence research.

ages from various institutions obtained from individuals diagnosed with various forms of brain tumors, with a focus on gliomas. T1-weighted (T1), T1-weighted with contrast enhancement (T1ce), T2-weighted (T2), and Fluid Attenuated Inversion Recovery (FLAIR) MRI scans are included in the collection. Expert radiologists use manual annotations to delineate the various tumor locations. The BraTS2020 dataset can be used by researchers and data scientists to develop and evaluate algorithms for brain tumor segmentation and classification, and it has been instrumental in advancing the state-of-the-art in medical image analysis and facilitating the development of automated methods for tumor detection, segmentation, and treatment planning. Researchers should go to the official website provided before to access the dataset or participate in the challenge[**45**].

3.2.2 Comparative analysis with earlier BraTS datasets

Each NIfTI file from a BraTS multimodal scan includes a volume description for the native (T1), post-contrast (T1Gd), T2-weighted (T2), and T2-FLAIR volumes. Using the same annotation technique, one to four raters manually segmented each imaging dataset, and their annotations were approved by skilled neuroradiologists. The supplied data have been stripped of the skull, co-registered to the same anatomical template, and interpolated to the same resolution (1 mm3) [**45**].

3.2.3 Dataset information

1. Multimodal scans available as NIFTI files (.nil.gz).
2. Four 'channels' of information - 4 different volumes of the same region [**45**]:

   (1) L Native (T1).
   (2) Post-contrast T1-weighted (T1CE).
   (3) T2-weighted (T2).
   (4) T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes.

3.3 Data pre-processing

3.3.1 Loading and data undersatanding

1. Load the the MRI modalities in an appropriate data structure.
2. Choose a slice to display (e.g. Slice 50 of the enhancing tumor sub-region).
3. Visualize the MRI modalities.

- All the imaging datasets have been segmented manually and were approved by experienced neuro-radiologists[**46**].
- Annotations (labels):
  - Label 0: Unlabeled volume.
  - Label 1: Necrotic and non-enhancing tumor core (NCR/NET).
  - Label 2: Peritumoral edema (ED)
  - Label 3: Missing (No pixels in all the volumes contain label 3)
  - Label 4: GD-enhancing tumor (ET)

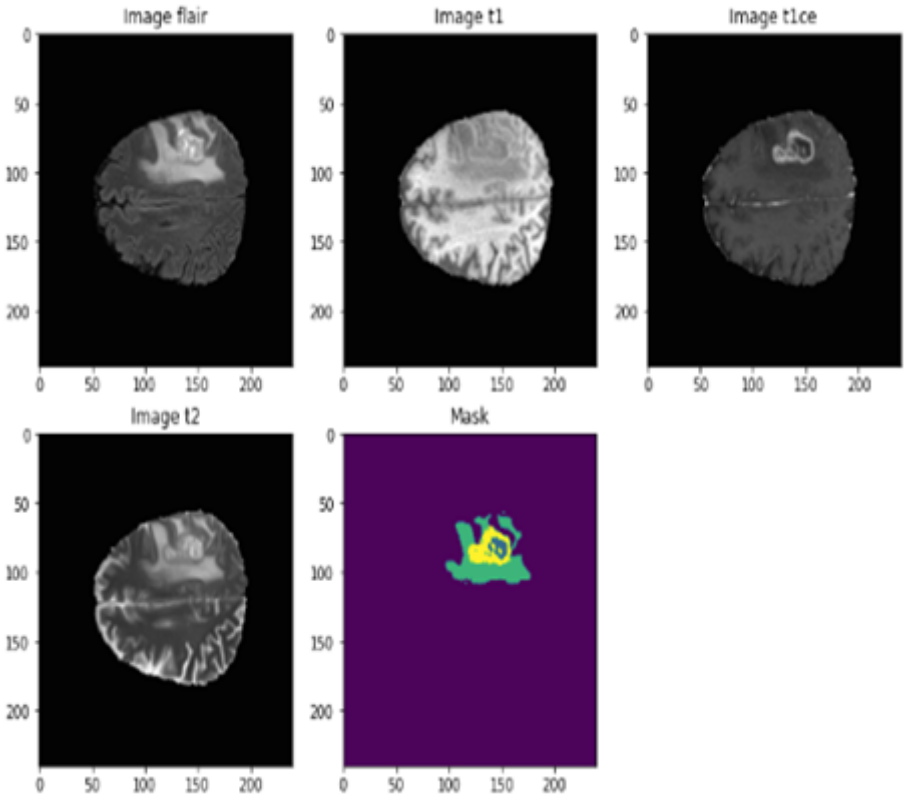

Figure 3.2: Visualistaion of the MRI modalities

3.3.2 Physics of MRI Imaging Sequences

- **FLAIR :** FLAIR is a highly effective MRI sequence to distinguish the edema region from the CSF with white matter appearing dark grey, cortex light grey and fat light [**47**] .

- **T1 :** T1-weighted MRI sequences are used to analyze brain tumor patterns, allowing for easy annotation of healthy cells.

- **T2-weighted :** T2-weighted sequences generate long TE and TR times, making CSF brighter than other MRI sequences.

- **T1-ce :** T1ce sequences are more sensitive than other sequences, showing details of regional angiogenesis and the integrity of the blood-brain barrier.

  The MRI modality characteristics are organised in the table 3.1

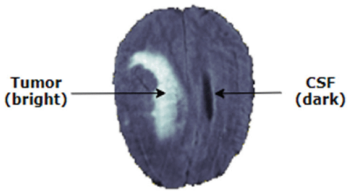

Figure 3.3: The FLAIR modality

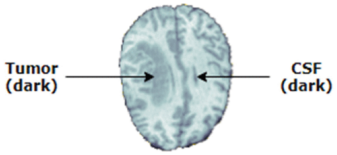

Figure 3.4: The T1 modality

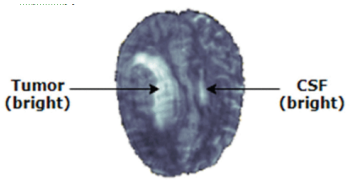

Figure 3.5: The T2 weighted modality

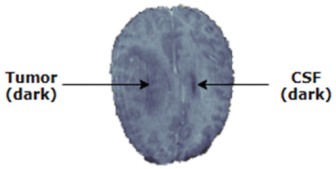

Figure 3.6: The T1ce modality

Sequence	TR	TE	Tumor	CSF
FLAIR	Very long	Very long	Bright	Dark
T1	Short	Short	Dark	Dark
T2	Long	Long	Bright	Bright
T1ce	Long	Long	Dark	Dark

Table 3.1: Brain MRI sequences based on features and graphical appearance.

3.3.3 Data exploration

As an exploring step, we tried to visualise the data and scatter the data points to check the possibility of classification.

The figure 3.7 indicates that the classes are considered clusters, which the CNN can classify.
On the other hand, the number of voxels within each class is counted to check the data imbalance. The voxel counter applied to the raw data shows the following ratios:

- Background : 98 %
- Edema : 0.25 %
- Non-enhancing Tumor : 0.65 %
- Enhancing tumor : 0.22 %

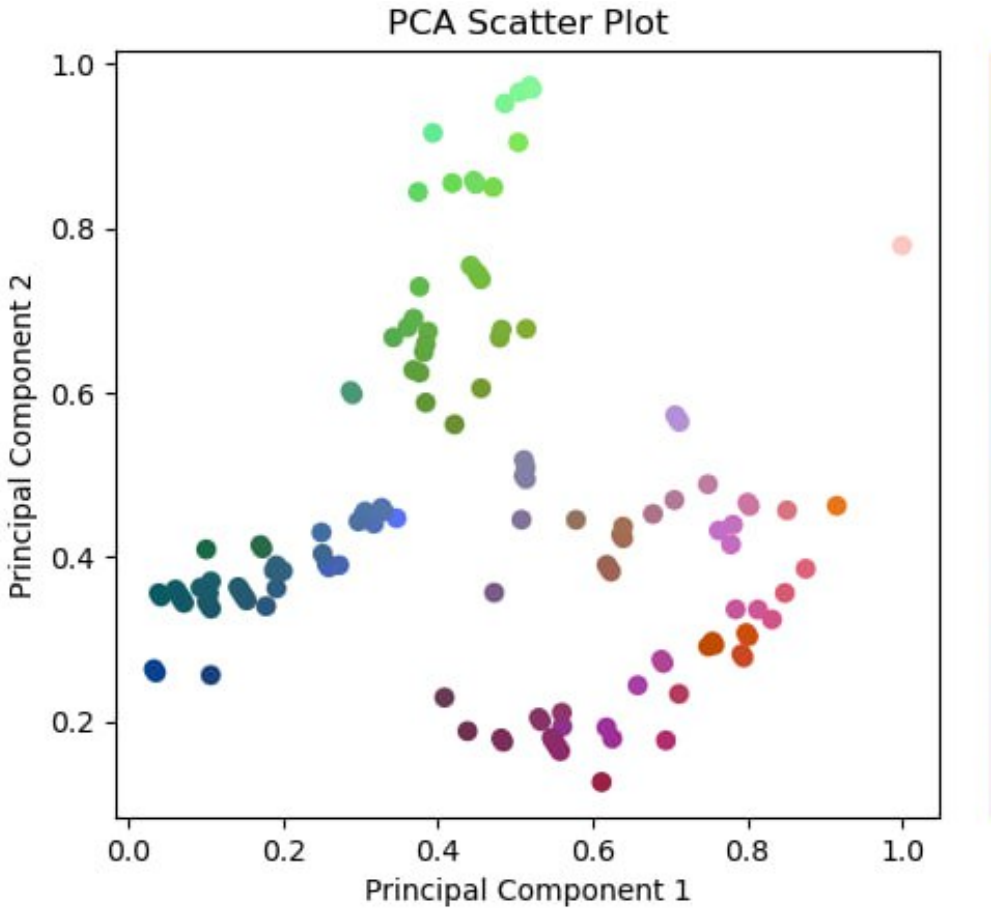

Figure 3.7: Visualisation of the data voxele's intensity

The ratios indicate the imbalance of the data, and as an action to reduce the dominance of the background over the other classes, cropping is applied to the images of data where the black area is cropped as in 3.8.

after the cropping, the ratios become :

- Background : 75 %
- Edema : 7.91 %
- Non-enhancing Tumor : 9.204 %
- Enhancing tumor : 7.889 %

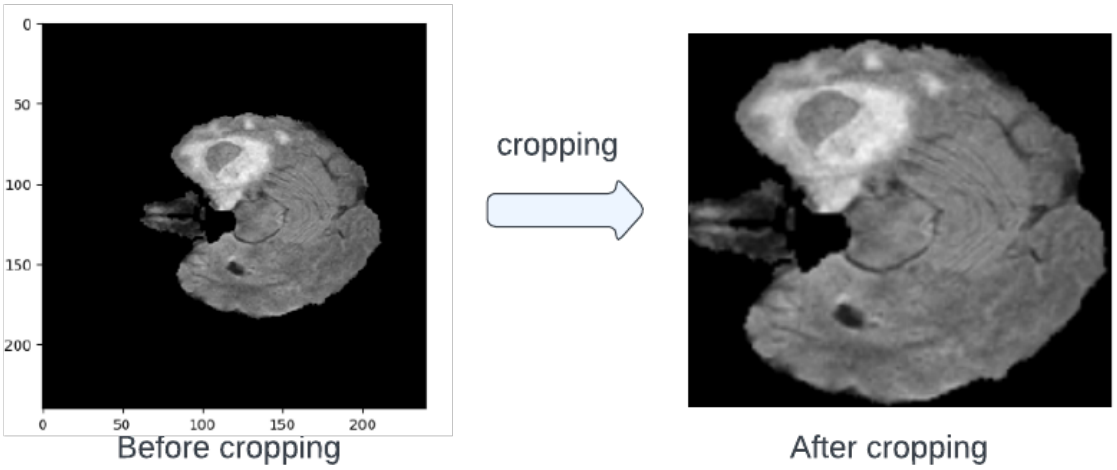

Figure 3.8: Cropping of the black area.

3.3.4 Data combining

Combining various MRI modalities into a single 3D image volume is a preprocessing step for the BraTS 2020 data set. According to the results of the investigation [**47**], it is feasible to avoid using T1 in combination, but combining FLAIR, T2, and T1ce pictures is regarded as the best strategy for brain tumor segmentation since it offers the most thorough information about the tumor and its surrounding tissue. FLAIR imaging focuses on the edema (swelling) around the tumor, whereas T2 imaging focuses on the tumor's necrotic (dead) tissue. T1CE imaging improves tumor contrast and helps identify it from surrounding healthy tissue. The combination of FLAIR, T2, and T1ce images gives the most complete and accurate picture of brain tumors in the Brats dataset. The combining process is done by the following steps:

1. Load the MRI images for each modality (T2, FLAIR, and T1ce).
2. Stack the images along the 3th dimension to create a 3D volume where each channel corresponds to a different modality.
3. Resample the image volume to a common voxel size to ensure consistency across modalities.
4. Use the combined 3D volume as input to segmentation models.

3.4 What is the U-Net Model

The UNet model is a type of convolutional neural network that we refer to as "deeply connected. Both an encoder network and a decoder network are part of it. The decoder network reconstructs the output layer segmentation map, while the encoder network is in charge of extracting features from the incoming picture. Convolutional processes with weights that indicate multi-channel feature maps make up the encoder and decoders [**48**].

3.5 U-Net Architecture

The network's general structure is obvious from the term alone. It's a U-shaped network made up of an expansive route and a contracting path. The network itself resembles a skateboard ramp, and the fundamental idea is that while moving uphill (an expansive path), the network learns to locate the object, while moving downhill (a contracting path), the network learns to classify the object[**49**].

3.5.1 Skip connections

Convolution is a matrix multiplication, while transpose convolution is a reverse-order multiplication. U-shape (U-net) is an encoder-decoder scheme design with long skip connections used for tasks with the same spatial dimension as the input, such as picture segmentation[**50**], optical flow estimation, and video prediction. Skip connections can be used to recover fine-grained features in prediction, and symmetrical long skip connections perform well in dense prediction tasks (medical picture segmentation).

3.5.2 Contracting path

This path is similar to a convolutional network, with layers made up of two 3x3 convolutions and a rectified linear unit. Each layer is followed by a 2x2 max pooling operation with a stride of two and a doubling of the number of feature channels. Convolutions begin with 64 feature channels and progress until 1024 channels are present, at which point the output transfers to the expansive path[**50**].

3.5.3 Expansive path

A U-Net differs from other CNNs in that the last pooled output is not delivered to a fully connected layer, but instead goes through a 2x2 "up-convolution". This is followed by layers in which the up-convoluted output is concatenated with its contracted path pair to re-include localization information and treated in two 3x3 convolutions each followed by a ReLU. Finally, an 1x1 convolution transfers the resulting 64-component feature vector to classes[**50**].

3.5.4 3D U-Net architecture

The 3D U-Net architecture is a deep neural network designed for volumetric medical image segmentation, using 3D convolutional layers to process volumetric data for MRI and CT scans[**51**].

The 3D U-Net architecture is composed of an encoder and decoder network connected by skip links. The encoder network increases feature channels while decreasing spatial resolution, while the decoder network uses 3D transposed convolutional layers to improve feature spatial resolution while decreasing the number of feature channels. Skip connections are used to preserve spatial information.

3D U-Net is a powerful and effective method for volumetric medical picture segmentation, used in a variety of applications such as brain tumor, liver, and cardiac segmentation.

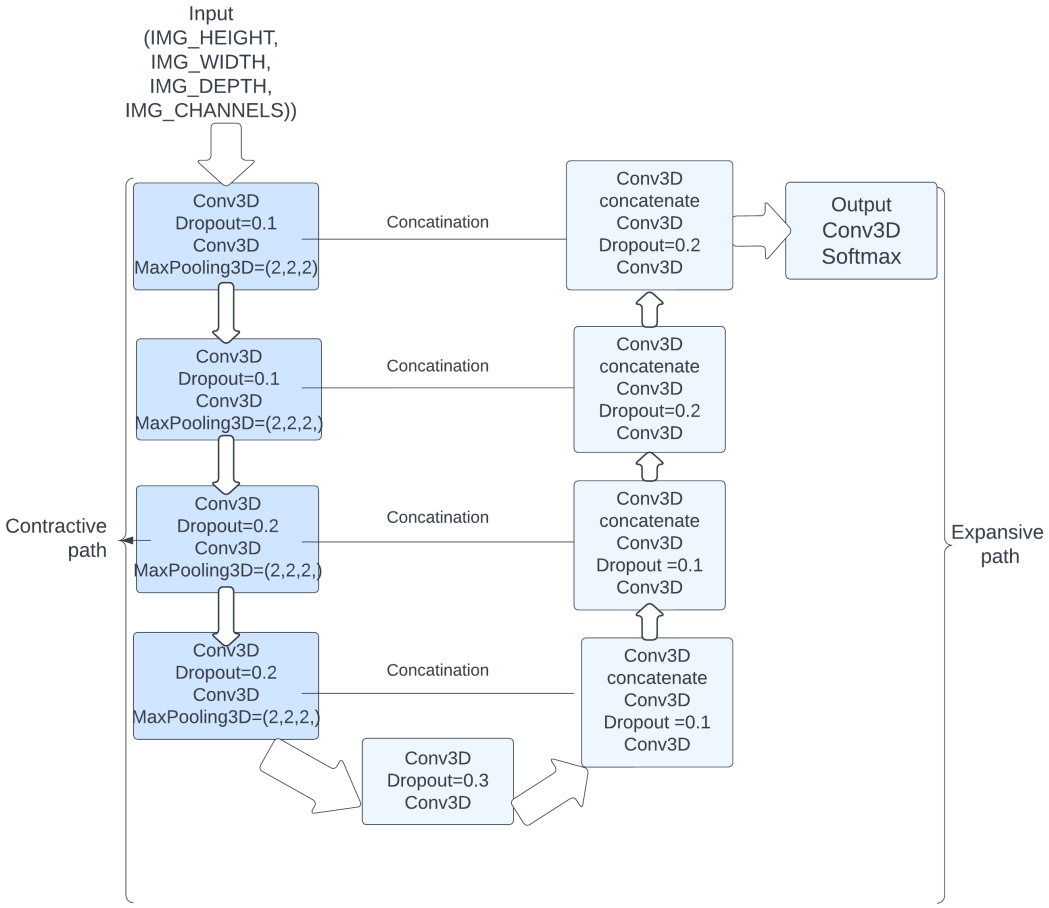

Figure 3.9: Our U-net like architecture

Advantages

The 3D U-Net architecture has been proven to be highly accurate in a number of medical picture segmentation tasks, including brain tumors, the liver, and the heart. It surpassed other cutting-edge topologies in terms of segmentation accuracy, sensitivity, and specificity for lung segmentation. Volumetric segmentation is also more reliable and accurate than 2D segmentation techniques, as it can account for the whole volume of the picture. This is essential for the segmentation of tiny items or structures. The 3D U-Net has demonstrated noise resistance against a variety of noise types, including impulse noise and Gaussian noise. Wang et al. (2021) found that the 3D U-Net outper-

formed other state-of-the-art designs in terms of accuracy and noise resistance. The deep learning architecture of the 3D U-Net enables automatic feature learning from input data, eliminating the need for manual feature extraction, which can take a long time and result in information loss[**52**].

Disadvantages

The 3D U-Net has a high cost of computing and data requirements, which can be limiting for some medical image analysis applications. To work at its best, the 3D U-Net needs a substantial amount of annotated training data, which can be expensive and time-consuming. Additionally, the 3D U-Net design is susceptible to over-fitting, which is when a model grows too complicated and begins to remember the training set of data rather than understanding the underlying patterns[**52**].

3.6 Conclusion

We have proposed a U-Net-like architecture in this chapter as a unique method for segmenting 3D brain tumors in order to overcome the difficulties of precise and effective segmentation in volumetric brain imaging. Through modifications tailored to 3D segmentation tasks, our research aims to make use of the U-Net architecture's benefits while also enhancing performance and clinical application.

For precise 3D brain tumor segmentation, we determined the necessity of a specialized architecture that efficiently captures spatial context and includes multi-scale data. We expanded U-Net's functionality to handle volumetric data in response to the success of U-Net in 2D picture segmentation, offering a full solution for the segmentation of brain tumors in numerous slices of the 3D domain.By combining 3D convolutional layers and volumetric pooling operations to capture spatial relationships and maintain spatial resolution throughout the network, our suggested U-Net-like architecture improves on the original U-Net design. In order to simplify the integration of multi-scale information and enable accurate localization of tumor boundaries, we also implemented skip connections.The outcome, experimental results, and challenges for our U-net-like architecture will be presented in the

next chapter, as will the evaluation of the whole deep learning model.

Chapter 4

Evaluation and experimental results

4.1 Introduction

In this chapter we present the evaluation phase for our approach, first we will tackle the existing metrics that evaluate the volumetric segmentation models. Then we show the experimental results of the approach in addition to the applied technologies and environment.

4.2 Metrics

Before passing on to the tests and experiments, we should tackle different metrics to evaluate the performance of volumetric approaches. On the other hand, ensuring that the model is making accurate matches is not the only problem we have in the process of building an end-to-end system; there are several hurdles to overcome when dealing with medical volume segmentation, such as metric selection and the inefficiency of the metric evaluation.the use in the literature of multiple definitions, leading to difficulties with large volumes and last but not least, a lack of support for fuzzy segmentation by existing metrics. In spite of that, the following metrics for volume segmentation are derived from the literature review. There are six types of metrics: overlap-based, volume-based, pair-counting-based, information-theoretic-based, probabilistic-based, and spatial distance-based.

4.2.1 Medical volume segmentation metrics

Medical volume can be denoted as a point set $X = \{x_1, x_2, ..., x_n\}$ where $|X| = w * h * d = n$. The ground truth segmentation set is $S_g = \{S_g^1, ..., S_g^k\}$ where k is the number of labels / classes. $f(x)_g$ is the assignment that indicates where x belongs to S. In the case of crisp segmentation or what known as a voxels membership $f(x)_g = 1$ so $x \in S_g, f(x)_g = 0 sox \notin S_g$ (When the classification is binary). In case of fuzzy classification or a probability degree $f(x)_g \in [0, 1]$, where the crisp segmentation is just a special case of the fuzzy segmentation. $f(x)_t$ is the assignment that indicates the predicted segmentation.

4.2.2 Confusion matrix

A confusion matrix is a common table used for supervised classification that indicates how many pixels or data samples are correctly segmented or classified and how many are not[**53**], it is well illustrated in the figure 4.3. In addition to several performance metrics that are calculated from the confusion matrix, like in equations 4.1 and 4.2, these values [**54**] can also evaluate the effectiveness of the model and at what state the algorithm can give the highest possible accuracy.

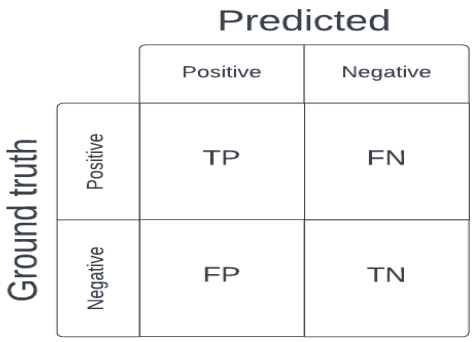

Figure 4.1: Confusion matrix in case of binary classification.

1. *TP* : The model correctly predicts a positive sample.
2. *TN* : The model correctly predicts a negative sample.

3. FP : The model incorrectly predicts a positive sample when the actual label is negative.

4. FN : The model incorrectly predicts a negative sample when the actual label is positive.

There are other values that can be extracted from the confusion matrix as follows :

- $$Specificity = Recall = TPR = \frac{TN}{TN + FP} \tag{4.1}$$

- $$Sensitivity = TNR = \frac{TP}{TP + FN} \tag{4.2}$$

4.2.3 Dice similarity coefficient *DSC*

DSC is a spatial overlap metric that is used as a statistical validation metric. DSC measures the similarity and reproducibility between the predicted and ground truth masks. It ranges from 0 to 1, where 1 indicates a perfect match. Formula :

$$DCS = \frac{2TP}{2TP + FP + FN} = \frac{2|S_g^1 \cap S_t^1|}{|S_g^1| + |S_t^1|} \tag{4.3}$$

4.2.4 Intersection over Union *IOU*

Known also as the Jaccard Index, it is widely used in the evaluation of semantic segmentation tasks. Comparable to DSC, the IOU metric measures the overlap between the predicted and ground truth masks and also ranges from 0 to 1, whereas DSC 1 expresses a perfect match. Formula :

$$IOU = \frac{TP}{TP + FP + FN} = \frac{|S_g^1 \cap S_t^1|}{|S_g^1 \cup S_t^1|} = \frac{DCS}{2 - DCS} \tag{4.4}$$

4.2.5 The Mutual Information *MI*

The mutual information between 2 variables is a measure of the amount of information one variable has about the other i,e the reduction is uncertainty of one variable, given that the other is known, as in [**55**] used the MI as a similarity between image segmentation, where the MI of regions (segments) is calculated instead on individual pixels.

$$MI(S_g, S_t) = H(S_g) + H(St) - H(S_g, S_t) \tag{4.5}$$

where $H(S)$ is the marginal entropy and $H(S_1, S_2)$ is the joint entropy

4.3 Technologies and environment

Our U-net model is implemented in Python and uses the TensorFlow library, as well as other machine learning and matrix manipulation libraries. The model is trained on a local CPU, the i5 9300H, and a compact GPU, the Intel 630 HD, with 16 GB of RAM.

4.4 Training process

We start by creating a training set (split the dataset in to train=0.80 ,(test=0.20) before developing our brain tumor segmentation model. We train the model to precisely segment the tumors using this set, which includes 3D slices of several types of brain tumors and matching masks. The weights of our model will be optimized using the Adam optimizer during training with a learning rate of 0.0001.

A mix of dice loss and focal loss will make up the loss function. The weighting of the dice loss will be [wt0, wt1, wt2, wt3] = [0.25, 0.25, 0.25, 0.25], giving each of the four classes equal weight. The model will be able to focus on cases that are challenging to categorize since the focused loss will be weighted with 1.

We utilize the accuracy metric, which measures the proportion of properly categorized cases, and the IOU-Score metric with a threshold of 0.5, which measures the intersection over union of the predicted and ground truth masks, to assess how well the model performed during training.

4.4.1 Focal loss *FL*

Focal loss adapts the standard CE to deal with extreme foreground-background class imbalance, where the loss assigned to well-classified examples is reduced [**56**].
focal loss formula:

$$L(y, p^{)} = -y(1-p^{)}log(p^{)}-(1-y)plog(1-p^{)} \tag{4.6}$$

4.4.2 Dice loss *DC*

Dice loss directly optimize the Dice coefficient which is the most commonly used segmentation evaluation metric[**57**, **58**].
Dice loss formula:

$$DL = 1—DiceCoefficient \tag{4.7}$$

4.4.3 total loss *TL*

We chose total loss, which is a combination of two losses, dice and focal, both of which are suitable for our objective. focal loss formula:

$$totalloss = Diceloss + focalloss$$

4.5 Results

In this section, we unveil the model outcome with real brain tumor cases . Also by applying different scenarios to the brain image as noise to simulate the real world artifacts, for the purpose of testing how far is the tolerance to variation of the deep learning model.

4.5.1 Test with the original data:

Before moving to the test results and the model performance, we tested the model before training to see how the model is far from giving random results as illustrated in the figure 4.2 the model outcome is completely confusing between the classes even when the whole tumor region have the least proportion area compared to the background.

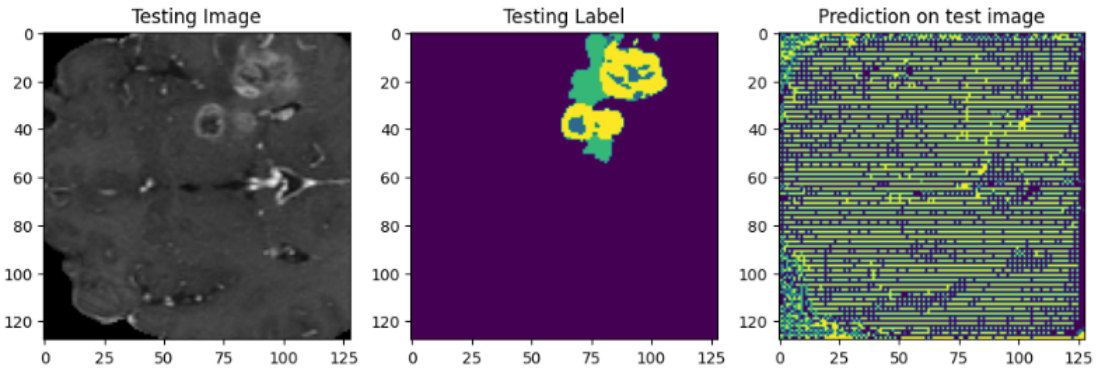

Figure 4.2: Randomness test

In this step, we predict from the test data with no changes applied to the input. As a result, we have a mean IOU of 0.24, which is

apparently a low value for segmentation for the whole validation data; this is why we applied a threshold.(threshold = 0.5), dividing where images with a high IOU score are separated from the low score image to find an explanation for the low accuracy of the model,results in the table 4.2,and in the figure4.3 of the confusion matrix for the model.

metrics	**Background**	**Edema**	**Enhancing**	**Non-enhancing**
Recall=0.024	0.96	00	0.025	0
Sensitivity=0.96	0.96	00	0.025	00

Table 4.1: Recall and sensitivity values for each class

Quality	number	Mean IOU
good	43	0.64
bad	26	0.38

Table 4.2: Result analysis

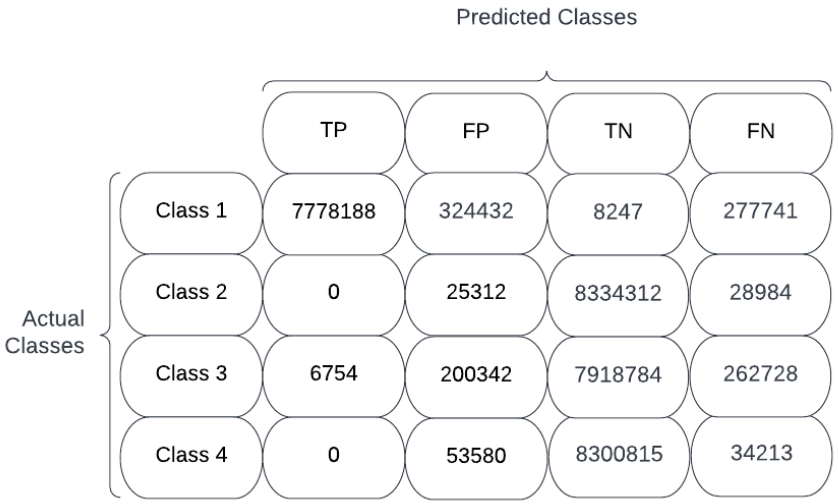

Figure 4.3: confusion matrix

Because the model's purpose is typically to minimize total training loss, unbalanced data can result in a biased model that favors the majority class. This bias might hamper the model's capacity to generalize to new data, resulting in poor performance[**59**].

Figure	IOU score	DSC
image 1	0.740	0.85
image2	0.781	0.87
image3	0.783	0.87

Table 4.3: images with good effect of figure4.4.

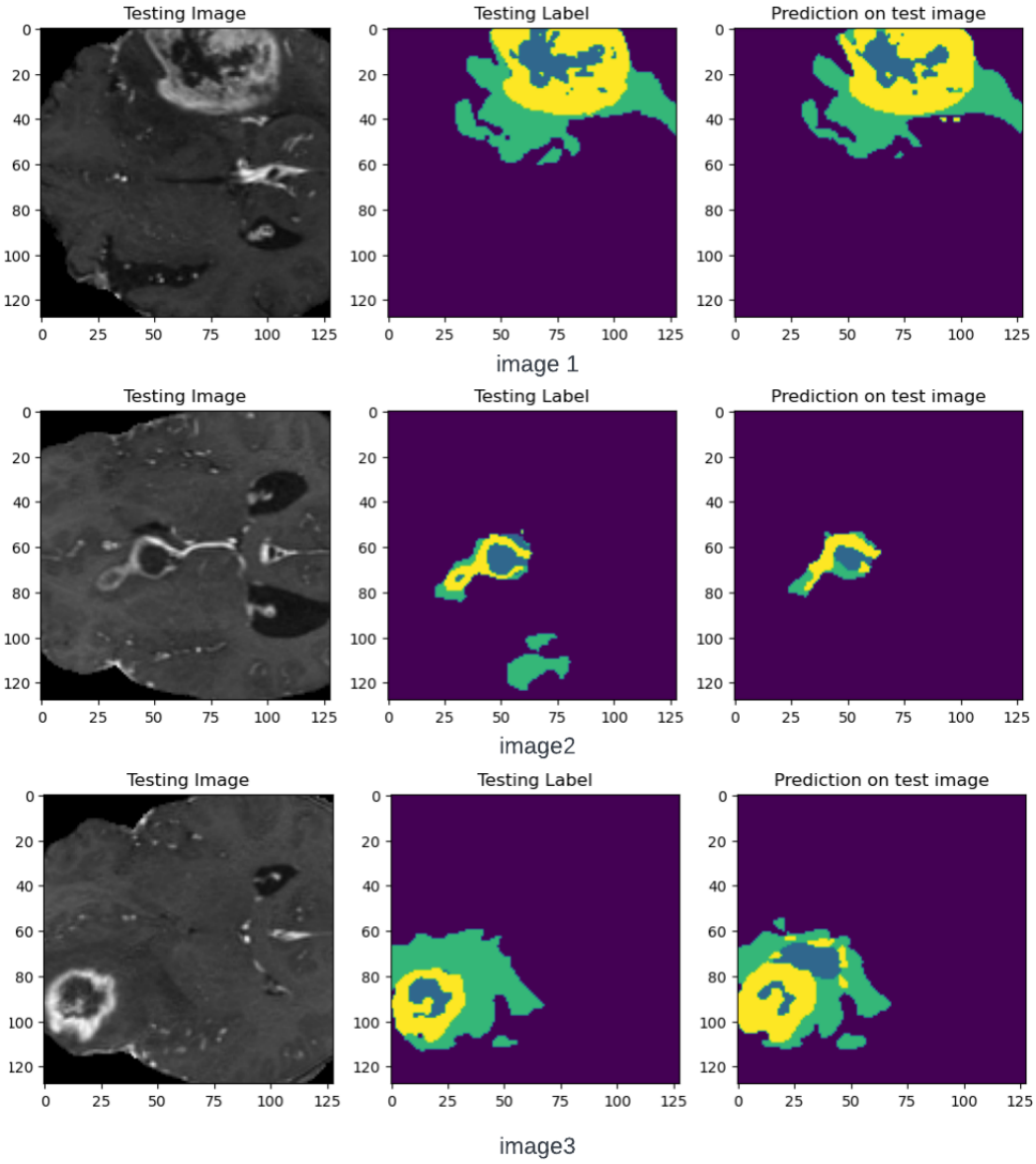

Figure 4.4: good image effect1

Figure	IOU score	DSC
image 1	0.23	0.37
image 2	0.30	0.46

Table 4.4: bad images effect of figure 4.5.

Figure	IOU score	DSC
image 1	0.67	0.80
image 2	0.46	0.63

Table 4.5: model predicte tumor, reflected that the testing label did not 4.6.

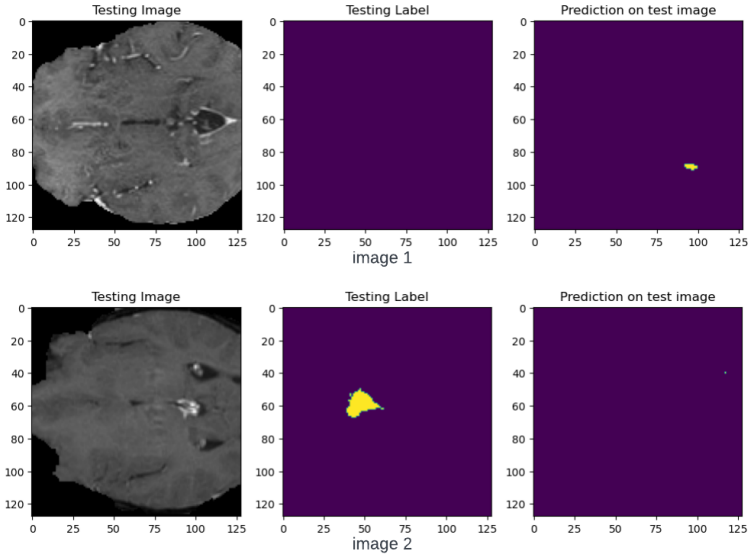

Figure 4.5: bad image effect.

We also noticed that the model predicte tumor as illustrated in the figure 4.6, reflected that the testing label did not contain any trace of the tumor:

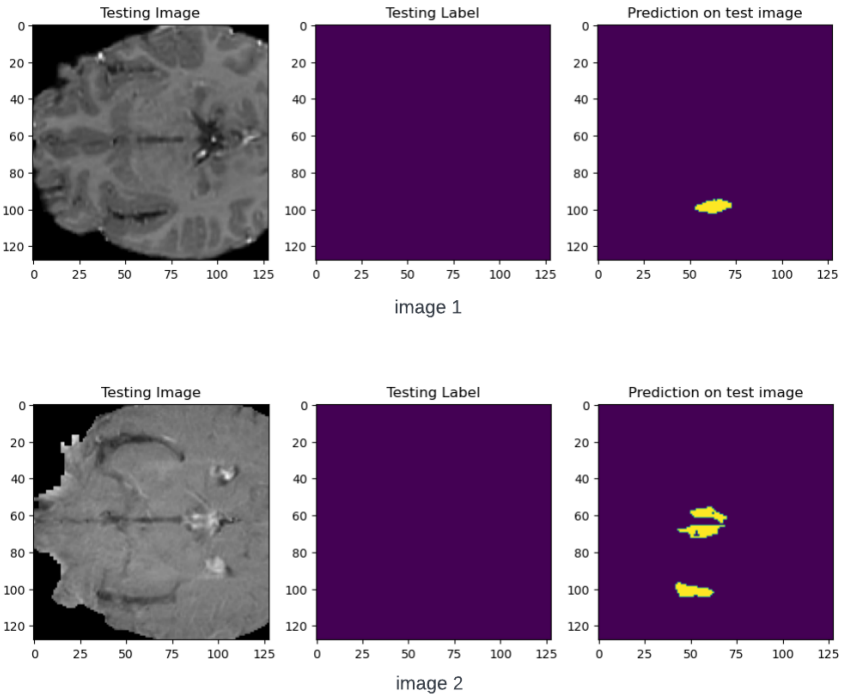

Figure 4.6: model predicte tumor, reflected that the testing label did not

4.5.2 Test with noisy data:

To see if the model is tolerant to variation,a noise was applied on some testing images a normally distributed noise. The following figures and table 4.6 show the model performance on noisy data. Based of the figure 4.7 the model performed well segmentation on light noise, even when the image is near to be corrupted as in images were the noise standard deviation is 0.3 and 0.4 the model is still capable of locating the whole tumor region although the segmentation of the tumor regions is low, it is worth to mention that the model has been not trained of the noisy data; which means that these are definitely new cases the model had to see for the first time. This can conclude that our model is not memorizing the training data even when the imbalanced data problem still exists.

Noise std	**IOU score**	**DCS**
noise=0.0	0.70	0.82
noise=0.1	0.6382	0.77
noise=0.2	0.5112	0.76
noise=0.3	0.3829	0.55
noise=0.4	0.21	0.34

Table 4.6: Noise results.

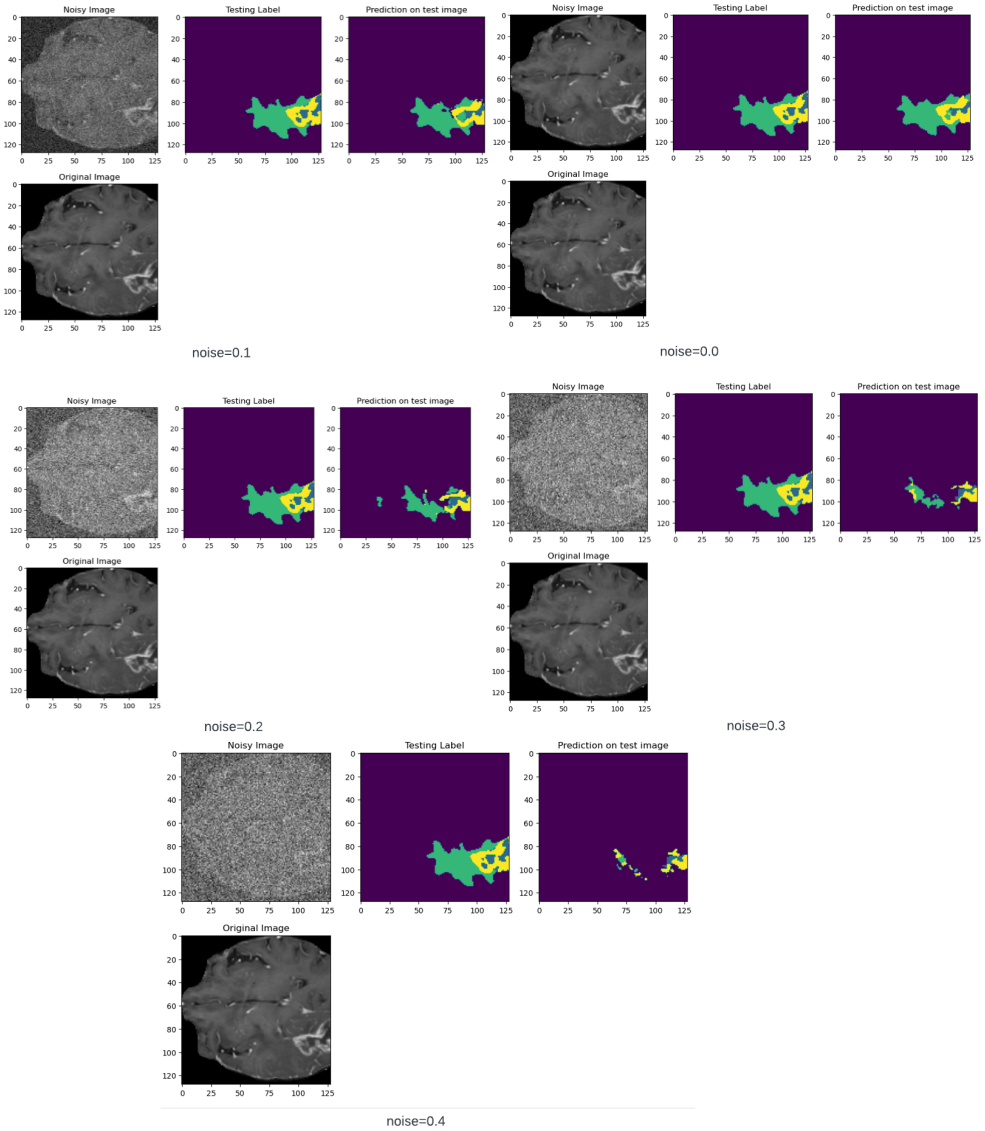

Figure 4.7: noise image effect.

4.6 discussion

Based on the experimental results, the model performed well in segmentation and tolerance to variation through the noise test on limited computation resources compared to the state of the art, even when the data is corrupted the model still recognise the existence of a tumor; this result that cannot be done by the naked eye, when the observation is a preliminary step in the diagnosis.
Despite of that, a handful of cases are missed by the model and can be considered exceptions. Also, we cannot deny the imbalance of data; even after the cropping, the background is still dominant over the whole tumor region. This concludes to the need for more refinement and improvement to handle the imbalance condition. Another important point that should be mentioned is that learning from multi-modality is a great way to deal with different angles of the problem and have a good representation of the data, specifically for a crucial problem such as brain tumor segmentation. However, other information from other data sources could be merged with MRI modalities to offer an obvious representation of the data to the deep learning model and avoid the exception cases that may occur only with the MRI data.

4.7 Conclusion

In this chapter, we illustrate the training information in addition to the development environment. The metrics used for the targeted task were also tackled.
passing by the results and outcomes of the U-net model. The test phase is applied under different scenarios. The first scenario is to test the data and analyze the results by figuring out what percentage of the good and bad results were performed by the model. This will lead to inferences about the model's learning and possible improvements to reduce the false-positive rate. The second scenario is to test the model on noisy data to see if it is tolerant to variation. The test result showed an encouraging outcome and proved that the deep learning model can outperform the human eye even when dealing with corrupted data, without forgetting that the model is trained on very limited computa-

tion results. With all the findings in the result section, the model still needs serious improvements to handle the imbalanced conditions and techniques to increase its IOU score.

General conclusion

In this thesis, we addressed deep learning-based multi-modal brain tumor segmentation. The approach developed is a 3D U-net designed for volumetric multi-modal MRI scans, aimed at extracting valuable diagnostic information contained within the medical image data. The automation of tumor segmentation using deep learning algorithms holds significant potential in the healthcare field, particularly in treatment planning and monitoring. It enables the identification of optimal treatment strategies, improves efficiency, saves time, and allows for quantitative assessment of different anatomical structures, tissue properties, and disease characteristics. The model development is based on various deep learning techniques aimed at achieving high accuracy and reliable segmentation. The input to the model is a multi-modal MRI tensor, which undergoes processing by the network to extract relevant features, ultimately yielding promising results.

By leveraging advanced deep learning techniques, the model demonstrates its capability to accurately segment the multi-modal MRI data, showcasing its potential for effective analysis and diagnosis in the medical field. The incorporation of multiple modalities allows for a comprehensive understanding of the underlying structures and characteristics, enhancing the overall performance and reliability of the segmentation process.

In this thesis, an extensive study was conducted utilizing the BRATS-2020 data-set. The developed model has showcased its effectiveness in accurately segmenting brain tumor regions within medical imaging data that encompasses multiple modalities. Through meticulous experimentation and evaluation, the obtained results consistently demonstrate improved segmentation performance in comparison to existing methods.

The thorough analysis of the BRATS2020 dataset has provided valuable insights and validation for the developed model's capability in managing the complex task as brain tumor segmentation. The model's correctness and robustness have been demonstrated by comprehensive evaluation of its performance against recognized benchmarks. These discoveries help advance medical imaging analysis by offering a dependable and effective tool for tumor localisation.
Despite the findings and outcomes of the model, several limitations and conditions were present during the research. One such limitation is the existence of imbalanced classes, where the background dominates compared to the tumor regions. Consequently, the model tends to prioritize background learning, classifying any misclassified voxel as background and disregarding the existence of other classes.

Additionally, deep learning algorithms require substantial amounts of data and computational resources to effectively generalize to different and new samples and cases. In future work, it is crucial to address the issue of data imbalance through dedicated techniques. Furthermore, incorporating auxiliary data from multiple medical sources and fusing it with multi-modal MRI scans can provide a comprehensive understanding of the tumor's overall state and its interaction with the brain.
In conclusion, the development of an accurate segmentation model represents a significant and encouraging advancement in medical analysis research, despite the limitations and areas for improvement. Each result and outcome contributes to the understanding of the training and learning processes employed by deep learning algorithms in addressing medical data and image analysis.

Bibliography

[1] *cancer,about-cancer,understanding,what-is-cancer,NATIONAL CANCER INSTITUE*

[2] *Overview of Types of Glioma,Aaron Cohen-Gadol, MD.*

[3] *mayoclinic,diseases-conditions,glioma,diagnosis-treatment.*

[4] *D*eep Learning for Medical Image Analysis, edited by S. Kevin Zhou, Hayit Greenspan, Dinggang Shen.

[5] *S*hen D, Wu G, Suk HI. Deep learning in medical image analysis. Annual review of biomedical engineering. 2017 Jul 11;19:221-48.

[6] *W*ang G, Li W, Zuluaga MA, Pratt R, Patel PA, Aertsen M, Doel T, David AL, Deprest J, Ourselin S, Vercauteren T. Interactive medical image segmentation using deep learning with image-specific fine-tuning.

[7] *L*itjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JA, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Medical image analysis.

[8] *D*.H. Hubel, T.N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, J. Physiol. 160 (1) (1962) 106.

[9] *A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, in: Proc. ICML, vol. 30, 2013.*

[10] *K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on ImageNet classification, arXiv:1502.01852, 2015.*

[11] *R*.K. Srivastava, J. Masci, S. Kazerounian, F. Gomez, J. Schmidhuber, Compete to compute, in: Advances in Neural Information Processing Systems, 2013, pp. 2310–2318.

[12] *M*.D. Zeiler, R. Fergus, Stochastic pooling for regularization of deep convolutional neural networks,arXiv:1301.3557, 2013.

[13] *B*aid U, Ghodasara S, Mohan S, Bilello M, Calabrese E, Colak E, Farahani K, Kalpathy-Cramer J, Kitamura FC, Pati S, et al. (2021) The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314.

[14] *L*iu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: Single shot multibox detector. In: European conference on computer vision, pp. 21–37. Springer.

[15] *Pereira S, Pinto A, Alves V, Silva CA (2016) Brain tumor segmentation using convolutional neural networks in mri images. IEEE Trans Comput Imag 35(5):1240–1251.*

[16] *Akil M, Saouli R, Kachouri R et al (2020) Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. Med Image Anal 63:101692*

[17] *Havaei M, Davy A,Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H (2017) Brain tumor segmentation with deep neural networks. Med Imag Anal 35:18–31.*

[18] *Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4700-4708).*

[19] *D*eep learning based brain tumor segmentation: a survey Zhihua Liu1 · Lei Tong1 · Long Chen1 · Zheheng Jiang2 · Feixiang Zhou1 · Qianni Zhang3 · Xiangrong Zhang4 · Yaochu Jin5 · Huiyu Zhou1, 17 June 2022.

[20] *Wang G, Li W, Ourselin S, Vercauteren T (2017) Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: International MICCAI brainlesion workshop, pp. 178–190. Springer*

[21] *Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. Springer.*

[22] *Intuitive Explanation of Skip Connections in Deep Learning,Nikolas Adaloglouon2020-03-23,theaisummer.*

[23] *Dong H, Yang G, Liu F, Mo Y, Guo Y (2017) Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In: annual conference on medical image understanding and analysis, pp. 506–517. Springer*

[24] *g G, Li W, Ourselin S, Vercauteren T (2017) Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: International MICCAI brainlesion workshop, pp. 178–190. Springer*

[25] *Shen H, Wang R, Zhang J, McKenna S (2017) Multi-task fully convolutional network for brain tumour segmentation. In: Annual Conference on Medical Image Understanding and Analysis, pp. 239–248. Springer*

[26] *Zhou C, Chen S, Ding C, Tao D (2018) Learning contextual and attentive information for brain tumor segmentation. In: International MICCAI brainlesion workshop, pp. 497–507. Springer*

[27] *Shen H, Wang R, Zhang J, McKenna S (2017) Multi-task fully convolutional network for brain tumour segmentation. In: Annual Conference on Medical Image Understanding and Analysis, pp. 239–248. Springer*

[28] *Nguyen HT, Le TT, Nguyen TV, Nguyen NT (2020) Enhancing mri brain tumor segmentation with an additional classification network. arXiv preprint arXiv:2009.12111.*

[29] *Iwasawa J, Hirano Y, Sugawara Y (2020) Label-efficient multitask segmentation using contrastive learning. arXiv preprint arXiv:2009.11160.*

[30] *Zhang D, Huang G, Zhang Q, Han J, Han J, Wang Y, Yu Y (2020) Exploring task structure for brain tumor segmentation from multimodality mr images. IEEE Trans Comput Imag Process 29:9032–9043*

[31] *Li X, Luo G, Wang K (2019) Multi-step cascaded networks for brain tumor segmentation. In: International MICCAI Brainlesion Workshop, pp. 163–173. Springer*

[32] *Islam M, Vibashan V, Jose VJM, Wijethilake N, Utkarsh U, Ren H (2019) Brain tumor segmentation and survival prediction using 3d attention unet. In: International MICCAI Brainlesion Workshop, pp. 262–272. Springer*

[33] *Zhou T, Ruan S, Guo Y, Canu S (2020) A multi-modality fusion network based on attention mechanism for brain tumor segmentation. In: 2020 IEEE 17th international symposium on biomedical imaging (ISBI), pp. 377–380. IEEE*

[34] *Zhou T, Canu S, Vera P, Ruan S (2020) Brain tumor segmentation with missing modalities via latent multi-source correlation representation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 533–541. Springer*

[35] *Yu B, Zhou L, Wang L, Yang W, Yang M, Bourgeat P, Fripp J (2021) Sa-lut-nets: Learning sample-adaptive intensity lookup tables for brain tumor segmentation. IEEE Trans Comput Imag 40(5):1417–1427*

[36] *Ghaffari M, Sowmya A, Oliver R (2020) Brain tumour segmentation using cascaded 3d densely-connected u-net.*

[37] *Yuan Y (2020) Automatic brain tumor segmentation with scale attention network. In: BrainLes MICCAI*

[38] *Henry T, Carre A, Lerousseau M, Estienne T, Robert C, Paragios N, Deutsch E (2020) Brain tumor segmentation with selfensembled,*

deeply-supervised 3d u-net neural networks: a brats 2020 challenge solution. arXiv preprint arXiv:2011.01045

[39] *dResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI. Authors : Rehan Raza, Usama Ijaz Bajwa, Yasar Mehmood, Muhammad Waqas Anwar , M. Hassan Jamal. Department of Computer Science,Lahore, Pakistan.Available online 14 June 2022, Version of Record 14 October 2022.*

[40] *Yuan Y (2022) Automatic brain tumor segmentation with scale attention network. In: BrainLes@MICCAI*

[41] *Jia H, Cai W, Huang H, Xia Y (2020) H2nf-net for brain tumor segmentation using multimodal mr imaging: 2nd place solution to brats challenge 2020 segmentation task. In: BrainLes@ MICCAI (2)*

[42] *Cirillo MD, Abramian D, Eklund A (2020) Vox2vox: 3d-gan for brain tumour segmentation. arXiv preprint arXiv:2003.13653*

[43] *Weninger L, Liu Q, Merhof D (2019) Multi-task learning for brain tumor segmentation. In: International MICCAI brainlesion workshop, pp. 327–337. Springer*

[44] *Liu C, Ding W, Li L, Zhang Z, Pei C, Huang L, Zhuang X (2020) Brain tumor segmentation network using attentionbased fusion and spatial relationship constraint. arXiv preprint arXiv:2010.15647.*

[45] *MICCAI 2020 BraTS ,Multimodal Brain Tumor Segmentation Challenge 2020: Data.*

[46] *BraTS,Multimodal Brain Tumor Segmentation Challenge 2017,v7labs*

[47] *TimeDistributed-CNN-LSTM: A Hybrid Approach Combining CNN and LSTM to Classify Brain Tumor on 3D MRI Scans Performing Ablation Study IEEE Access Published: 2020, Sidratul Montaha; Sami Azam; A. K. M. Rakibul Haque Rafid; Md. Zahid Hasan; Asif Karim; Ashraful Islam.*

[48] *What is UNet? How Does it Relate to Deep Learning?, Rachel StClair.*

[49] *An overview of Unet architectures for semantic segmentation and biomedical image segmentation,theaisummer.*

[50] *indiantechwarrio,rU-Net for Semantic Segmentation.*

[51] *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox Olaf Ronneberger .*

[52] *A 3D convolutional neural network with multi-branch attention for brain tumor segmentation from MRI images. Authors : Yuan Cao, Weifeng Zhou, Min Zang, Dianlong An, Yan Feng, Bin Yu. Available online 18 October 2022.*

[53] *Kelleher, J.D., Tierney, B. and Tierney, B., 2018. Data science an introduction. CRC Press.*

[54] *Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), pp.37-63.*

[55] *Russakoff DB, Tomasi C, Rohlfing T, Jr Maurer CR. Image similarity using mutual information of regions. In: 8th European Conference on Computer Vision, ECCV: 2004. p. 596–607.*

[56] *Focal Loss for Dense Object Detection. Authors : Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár Year: 2017*

[57] *Loss Functions for Medical Image Segmentation: A Taxonomy,JunMa.*

[58] *How To Evaluate Image Segmentation Models. Author : Seyma Tas.*

[59] *Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H. (2017). Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications, 73, 220-239.*

Multi-modal Brain Tumor Segmentation Using Stacked Denoising Autoencoders. Authors : Kiran Vaidhya, Subramaniam Thirunavukkarasu, Varghese Alex, and Ganapathy Krishnamurthi, Indian Institute of Technology Madras, Chennai, India.Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, First International Workshop, Brainles 2015 Held in Conjunction with MICCAI 2015 Munich, Germany, October 5, 2015, Revised Selected Papers.

Castillo LS, Daza LA, Rivera LC, Arbeláez P (2017) Volumetric multimodality neural network for brain tumor segmentation. In: 13th international conference on medical information processing and analysis, vol. 10572, p. 105720E. International Society for Optics and Photonics

Shaikh M, Anand G, Acharya G, Amrutkar A, Alex V, Krishnamurthi G (2017) Brain tumor segmentation using dense fully convolutional neural network. In: International MICCAI brainlesion workshop, pp. 309–319. Springer

Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH (2017) Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: International MICCAI Brainlesion Workshop, pp. 287–297. Springer

Lung and Infection CT-Scan-Based Segmentation with 3D UNet Architecture and Its Modification,Mohammad Hamid Asnawi,1 Anindya Apriliyanti Pravitasari,1, Gumgum Darmawan,1 Triyani Hendrawati,1 Intan Nurma Yulita,2 Jadi Suprijadi,1 and Farid Azhar Lutfi Nugraha1.*