People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
University Of Ferhat Abbas
Faculty of Sciences

Domain: Mathematics and Computer Sciences
Specialty: F3I

Presented by: Hiba DJARI and Malak LAMARA

Subject

Deep Learning Models for Brain Cancer Segmentation and Prognostic Analysis

The jury:

Pr. Abdelouahab MOUSSAOUI	President	UFAS1 SETIF
Dr. Skander HAMDI	Supervisor	UFAS1 SETIF
Dr. Imene ZENBOUT	Examiner	UFAS1 SETIF

Academic Year: 2024/2025

Abstract

Deep Learning (DL) has introduced transformative potential in medical diagnostics, particularly in oncology, where complex tumor structures present significant challenges for treatment planning. The accurate segmentation of brain tumors from Magnetic Resonance Imaging (MRI) is a critical first step, and recent advancements in DL have greatly enhanced the efficiency and reliability of this process. By integrating these precise segmentation outputs with clinical and genetic data, prognostic analysis can be substantially improved, paving the way for more accurate tumor prediction and robust clinical decision support.

This thesis aims to develop and evaluate a comprehensive pipeline for brain tumor analysis, from segmentation to survival prediction. It focuses on comparing advanced DL architectures for segmenting brain tumors and explores the adaptation of large-scale, general-purpose vision models for this specialized medical task. The ultimate goal is to fuse imaging features with clinical and genetic data to accurately estimate patient survival probability and stratify risk levels, thereby enhancing personalized treatment strategies.

This work offers a thorough exploration of both established and cutting-edge DL models for brain tumor segmentation and prognosis. The first contribution is a comparative analysis of segmentation models, including a standard U-Net, a DeepResUNet, and a VGG19-based U-Net, alongside foundational models like a fine-tuned Segment Anything Model (SAM) and the specialized Medical Segment Anything Model (MedSAM). The results underscored the effectiveness of MedSAM, which demonstrated superior performance . The second contribution introduces a prognostic analysis pipeline that leverages features extracted from the segmentation masks. This pipeline proved highly effective, with the Support Vector Machine (SVM) model showing notable success in accurately identifying high-risk patients, while the XGBoost model demonstrated strong predictive power for classifying low-risk patients.

Overall, these contributions validate a powerful, integrated approach for brain cancer analysis. The findings confirm that leveraging both CNN-based architectures and fine-tuned foundational models can yield highly accurate tumor delineations. Furthermore, the successful integration of these segmentation results into a prognostic pipeline demonstrates the critical role of automated image analysis in fostering personalized and more effective clinical management of brain tumors.

ملخص

لقد أحدثت التعلم العميق إمكانات تحويلية في مجال التشخيص الطبي، وبشكل خاص في علم الأورام، حيث تفرض الهياكل الورمية المعقدة تحديات كبيرة على تخطيط العلاج. ويعد التقسيم الدقيق لأورام الدماغ من صور التصوير بالرنين المغناطيسي خطوة أولى حاسمة، وقد عززت التطورات الحديثة في التعلم العميق بشكل كبير من كفاءة وموثوقية هذه العملية. ومن خلال دمج مخرجات التقسيم الدقيقة هذه مع البيانات السريرية والجينية، يمكن تحسين التحليل التكهني بشكل كبير ، مما يمهد الطريق لتنبؤ أكثر دقة بمسار الورم وتوفير دعم قوي للقرارات السريرية.

تهدف هذه المذكرة إلى تطوير وتقييم مسار عمل شامل لتحليل أورام الدماغ، بدءًا من التقسيم وصولاً إلى التنبؤ بفرص البقاء على قيد الحياة. وهي تركز على مقارنة بنيات التعلم العميق المتقدمة لتقسيم أورام الدماغ، وتستكشف تكيف نماذج الرؤية الحاسوبية الكبيرة وذات الأغراض العامة لهذه المهمة الطبية المتخصصة. ويتمثل الهدف النهائي في دمج السمات المستخلصة من الصور مع البيانات السريرية والجينية لتقدير احتمالية بقاء المريض على قيد الحياة وتصنيف مستويات الخطورة بدقة، وبالتالي تعزيز استراتيجيات العلاج المخصصة.

يقدم هذا العمل استكشافًا شاملاً لنماذج التعلم العميق الراسخة والمتطورة في مجال تقسيم أورام الدماغ والتكهن بمسارها. تتمثل المساهمة الأولى في تحليل مقارن لنماذج التقسيم، بما في ذلك نموذج U-Net القياسي، ونموذجDeepResUNet ، ونموذج U-Net القائم علىVGG19 ، إلى جانب النماذج التأسيسية مثل نموذج SAM المحسّن والنموذج الطبي المتخصص MedSAM. وقد أبرزت النتائج فعالية نموذج MedSAM الذي أظهر أداءً متفوقًا. أما المساهمة الثانية فتتمثل في تقديم مسار عمل للتحليل التكهني يستفيد من السمات المستخرجة من أقنعة التقسيم. لقد أثبت مسار العمل هذا فعالية عالية، حيث أظهر نموذج الآلات المتجهات الداعمة نجاحًا ملحوظًا في التحديد الدقيق للمرضى ذوي الخطورة العالية، بينما أظهر نموذج XGBoost قدرة تنبؤية قوية لتصنيف المرضى ذوي الخطورة المنخفضة.

بشكل عام، تُبرهن هذه المساهمات على فعالية وقوة النهج المتكامل لتحليل سرطان الدماغ. وتؤكد النتائج أن الاستفادة من كل من المعماريات القائمة على الشبكات العصبونية الالتفافية والنماذج التأسيسية المحسنة يمكن أن تؤدي إلى تحديد دقيق للغاية للأورام. علاوة على ذلك، يوضح التكامل الناجح لنتائج التقسيم هذه في مسار عمل التكهن بالمرض الدور الحاسم للتحليل الآلي للصور في تعزيز الإدارة السريرية المخصصة والأكثر فعالية لأورام الدماغ.

Résumé

Résumé L'apprentissage profond a révélé un potentiel transformateur pour le diagnostic médical, en particulier en oncologie, où les structures tumorales complexes posent des défis importants pour la planification du traitement. La segmentation précise des tumeurs cérébrales à partir de l'imagerie par résonance magnétique constitue une première étape essentielle, et les avancées récentes en apprentissage profond ont considérablement amélioré l'efficacité et la fiabilité de ce processus. En intégrant ces sorties de segmentation précises avec des données cliniques et génétiques, l'analyse pronostique peut être substantiellement améliorée, ouvrant la voie à une prédiction plus exacte des tumeurs et à un support à la décision clinique plus robuste.

Cette thèse vise à développer et évaluer un pipeline complet pour l'analyse des tumeurs cérébrales, de la segmentation à la prédiction de la survie. Elle se concentre sur la comparaison d'architectures d'apprentissage profond avancées pour la segmentation des tumeurs cérébrales et explore l'adaptation de modèles de vision généralistes à grande échelle pour cette tâche médicale spécialisée. L'objectif final est de fusionner les caractéristiques d'imagerie avec les données cliniques et génétiques afin d'estimer avec précision la probabilité de survie des patients et de stratifier les niveaux de risque, améliorant ainsi les stratégies de traitement personnalisées.

Ce travail propose une exploration approfondie des modèles d'apprentissage profond établis et de pointe pour la segmentation et le pronostic des tumeurs cérébrales. La première contribution est une analyse comparative de modèles de segmentation, incluant un U-Net standard, un DeepResUNet, et un U-Net basé sur VGG19, ainsi que des modèles fondamentaux comme un Segment Anything Model affiné et le modèle spécialisé MedSAM. Les résultats ont souligné l'efficacité de MedSAM, qui a démontré des performances supérieures. La deuxième contribution introduit un pipeline d'analyse pronostique qui exploite les caractéristiques extraites des masques de segmentation. Ce pipeline s'est avéré très efficace, le modèle de machine à vecteurs de support montrant un succès notable dans l'identification précise des patients à haut risque, tandis que le modèle XGBoost a démontré une forte capacité prédictive pour la classification des patients à faible risque.

Globalement, ces contributions valident une approche puissante et intégrée pour l'analyse du cancer du cerveau. Les résultats confirment que l'exploitation à la fois des architectures basées sur les réseaux de neurones convolutifs et des modèles fondamentaux affinés peut produire des délinéations tumorales de haute précision. De plus, l'intégration réussie de ces résultats de segmentation dans un pipeline pronostique démontre le rôle essentiel de l'analyse d'images automatisée dans la promotion d'une gestion clinique personnalisée et plus efficace des tumeurs cérébrales.

Acknowledgment

We would like to express our sincere gratitude to our supervisor Dr. Skander HAMDI , for their invaluable guidance and support throughout this research. Their profound expertise, critical insights, and meticulous feedback were instrumental in shaping this thesis. We are deeply appreciative of their commitment to academic excellence and the intellectual freedom they provided, which allowed us to explore this topic in depth. This work would not have been possible without their steady direction and scholarly advice.

Dedication

All praise to Allah, whose strength, blessings, and guidance have led me through this long and trying path.

I would also like to express my gratitude to my own resolve and strength in enduring the many obstacles, pressures, and responsibilities that came with this journey.

I'm truly grateful to my family especially my mother Soraya , my sister Dounia, my brother Youcef, my cat Minou and my friends for always being there for me. My mom's care and strength carried me through more than she knows and Dounia's words always lifted me up when I needed it most. Even on quiet days, Minou's little presence brought me unexpected comfort. To my friends thank you for your patience, encouragement, and for always reminding me I could do this.

A special thank you to my project partner and friend, Malak. We've been through so much together over these past five years, having you by my side throughout this whole university journey made everything easier, more meaningful, and a lot more fun. I couldn't have asked for a better person to share this chapter with, and I'm so thankful for your constant support, your good energy, and just... being you.

My genuine appreciation to my supervisor Hamdi Skander for his constant guidance, support, and encouragement moral and material throughout the period of my final year project.

Lastly, I would like to thank the committee president and examiners for sparing their time, effort, and consideration in reading my thesis and participating in my defense.

Hiba Djari

Dedication

First, my deepest gratitude to my dad Lamara Brahim, your unshakable confidence and silent fortitude walked right alongside me every step of the way. You gave me everything I needed to survive, yet so much more.

My gorgeous mum, Lamara Hayat, your love, prayers, and selfless heart led me on the path. You were my anchor during every stormy tempest.

To my brothers Abderahim, Charaf Eddine and Akram thank you for being the source of my strength at all times. The love you have for me gave me strength to never lose hope.

To my grandmother, your prayers and your presence in my heart will continue to console and give me peace.

To my professor, Skander Hamdi thank you for being there all these years. Your constant encouragement, warm guidance, and faith in me from the very start have been greatly valued. Your advice guided me to grow, learn, and become what I am today.

To my soul friend Hiba Djari thank you so much for the beautiful memories, for allowing me to share life with you, and for your genuine effort and support towards this thesis. You made it less difficult and deserving.

And lastly, thank you to myself for not giving up when the times were rough, for standing every time that I fell, and for having the courage to stand here. I am proud of the person that I have become to be.

It's not just a triumph of mine; it is that of all of you.

Malak Lamara

Contents

List of Figures

List of Tables

List of Acronyms

ARU-GD Attention Res-UNet with Guided Decoder

CNN Convolutional Neural Networks

CNS central nervous system

CT Computed Tomography

DL Deep Learning

DSC Dice Similarity Coefficient

DWI Diffusion-Weighted Imaging

ET Enhancing Tumor

FLAIR Fluid-Attenuated Inversion Recovery

fMRI functional Magnetic Resonance Imaging

GBM glioblastoma multiforme

HD95 Hausdorff Distance

HGG Higher-Grade Gliomas

IoU Intersection over Union

LGG Lower-Grade Gliomas

MedSAM Medical Segment Anything Model

ML Machine Learning

MRI Magnetic Resonance Imaging

MRS Magnetic Resonance Spectroscopy

PET Positron Emission Tomography

PWI Perfusion-Weighted Imaging

RBF radial basis function

ResNet Residual Networks

SAM Segment Anything Model

SVM Support Vector Machine

TC Tumor Core

TCGA The Cancer Genome Atlas

TCIA The Cancer Imaging Archive

ViT Vision Transformer

WFL Weighted Focal Loss

WT Whole Tumor

Chapter 1

Introduction

1.1 Research Context and Motivation

Medical imaging has revolutionized diagnosis and treatment planning in clinical practices since the discovery of X-ray imaging in the twentieth century. As technology proceeds unregulated, several imaging modalities have been developed, each for addressing some of the clinical requirements, such as Computed Tomography (CT), MRI, Positron Emission Tomography (PET), etc. Of these, MRI is the gold standard for imaging soft tissues, particularly neurology. It is a significant role in detecting, investigating, and monitoring brain abnormalities such as tumors.

One of the most malignant and deadly cancers is brain cancer, or gliomas. One of the brain tumors is the Lower-Grade Gliomas (LGG) that, while they start slowly in contrast to Higher-Grade Gliomas (HGG), tend to become more malignant if not diagnosed and treated early. Due to heterogeneity in behavior and appearance, correctly identifying and segmenting LGG on MRI images is exceedingly crucial for early prognosis and diagnosis. Visual inspection is slow, strongly expert-labor-intensive, and prone to inter-observer variations.

In such a scenario, DL has proven to be revolutionary in the analysis of medical images. DL, particularly Convolutional Neural Networks (CNN) and their variants, can learn hierarchical features from raw image data automatically such that one is able to attain accurate classification, segmentation, and prediction without having to design features manually. Developments in attention mechanisms, transformer models, and hybrid models further enabled DL models to be even more accurate in detecting subtle patterns in complex medical images and are best suited for challenging operations like brain tumor segmentation.

Despite all of these advances, there are still a few difficulties. The complex nature of brain tissue, the variable shape and location of tumors, and the lack of annotated medical data limit automatic brain tumor analysis to an extremely difficult task. Besides, for LGG, tumors are diffuse and visually indeterminate, and hence, high sensitivity in their early detection when the condition is non-life-threatening. There is an urgent need for stable, explainable, and clinically feasible DL models that can assist radiologists and oncologists in accurate segmentation and prognostic assessment of brain tumors.

These critical needs drive this thesis, and our primary contribution is to address them. We build and prove new DL architectures that can correctly segment LGG from MRI images and provide prognostic information, resulting in early diagnosis and customized treatment protocols. Using advanced DL architectures and domain-specific augmentations, this work contributes directly to closing the gap between computational advancement and clinical relevance in neuro-oncology.

1.2 Research Contributions

This thesis focuses on the application of DL techniques for brain tumor segmentation and prognostic analysis, specifically targeting LGG. The main contributions of the research are as follows:

- A concise review of existing DL approaches applied to brain tumor analysis, particularly in the context of segmentation and prognosis using MRI scans, to establish the foundation for the current work.

- Design and implementation of several DL-based segmentation models for LGG analysis, based on popular architectures such as U-Net,Deep ResUNet, VGG19-based U-Net , MedSAM and the SAM. These architectures were selected and adapted to be maximally appropriate for segmenting brain tumor regions in MRI scans.
- Exploring DL as a tool for prognostic analysis, aiming to predict how the disease might progress by using features learned from MRI images. This approach helps support doctors in making better and more informed clinical decisions.
- Testing and validating the models on a public LGG MRI Segmentation dataset. This included all necessary steps, such as model preprocessing, training and evaluation using common performance metrics like precision, recall, F1-score, Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and average prediction time.

1.3 Structure of thesis

The thesis has a structure designed to facilitate a smooth, coherent, and logical progression of ideas, processes, and outcomes related to the use of DL in brain cancer segmentation and prognostic analysis. Each chapter builds systematically on the previous one, providing both a strong theoretical foundation and practical insight. This allows the reader to follow the research from inspiration to methodological formulation and implementation in real life. The thesis has four main chapters, each focusing on a key aspect of the research. It begins with a review of the clinical issues of brain tumor diagnosis and the importance of proper segmentation, especially for LGG on MRI. This is then followed by a chapter giving the theoretical background and addressing DL models like U-Nets and Deep ResUNet that form the basis for the solutions proposed. Next, a comprehensive review of literature critically compares recent DL-based techniques both on the same dataset applied in this study as well as other publicly available datasets. The methodology chapter subsequently explains the design, implementation, and performance evaluation of the DL models used in segmentation and prognosis analysis, including model architecture and comparison of performance. The thesis concludes with a summary of the key conclusions, their broader implications, and suggestions for further research, particularly towards the establishment of the clinical application of DL in neuro-oncology.

Chapter 2: Background
This chapter gives the background required for understanding the advanced computational techniques utilized in this thesis for brain tumor analysis. It begins with describing the primary role of Machine Learning (ML) and DL in medical image analysis compared to the traditional approach. The chapter subsequently delves into the theoretical foundation, architectures, function, and principal applications of some of the state-of-the-art DL models such as various U-Net architectures and Transformer-based models such as SAM and MedSAM. Furthermore, it touches on simple traditional ML algorithms, gradient boosting algorithms, and ensemble methods applicable to the classification and prognostic tasks. The strengths and limitations and applicability of these diverse models in addressing the sophistications of medical imaging and brain tumor analysis are also considered, providing a theoretic framework that is comprehensive for the methods later employed.

Chapter 3: Literature Review
Recent works in brain tumor segmentation using MRI are carefully reviewed, covering both the same dataset employed in this work and other publicly available datasets. This includes a description of the state of the art in brain tumor segmentation from MRI and DL in medical image analysis. The recent methods and their limitations are explained to find out what gaps in research this thesis will aim to address.

Chapter 4: Methodology
This chapter describes the methodological approach taken in our work for the segmentation and prognosis analysis of brain tumors using DL. It begins by discussing the dataset that was utilized in the study, including how it is composed, its quality, and how appropriate it is. Since the data were clean and well structured to begin with, minimal preprocessing was required, with missing values being imputed using the

most frequent value for each feature and one could thus concentrate more directly on model development and training. The chapter also describes the design and implementation of the proposed DL models, including the reasons for which the selected modeling techniques were used, parameterization, and training procedures. Step-by-step model training and validation procedure for determining the reproducibility and reliability of results is described. The evaluation procedure is also described, along with the most important performance measures that were employed in evaluating the performance of the models.

General Conclusion This chapter summarizes the key findings on model performance for tumor segmentation and prognosis, highlighting their worth in creating more objective and efficient clinical assessments. It also considers the study's broader contribution to neuro-oncology and makes specific recommendations for future research focused on model validation, interpretability, and eventual clinical application.

Chapter 2

Background

2.1 Introduction

Brain cancer remains one of the most dangerous and complex forms of cancer with difficult early diagnosis, treatment planning, and prediction. Its aggressiveness and heterogeneity necessitate extremely accurate and timely diagnosis to guide clinical management. Among imaging modalities, MRI remains the gold standard due to its lack of invasiveness and ability to reveal detailed anatomy accurately. However, brain tumor segmentation on MRI scans by hand is time-consuming, susceptible to inter-observer variability, and time-consuming, making automated procedures increasingly important in modern clinical workflows.

Over the recent years, ML and DL have emerged as revolutionary technologies in medical image processing. Due to their data-driven nature, they enable the discovery of subtle patterns and correlations in imaging data to support tasks such as tumor segmentation, classification, and patient outcome prediction. DL models, particularly CNN, have achieved unrivaled success in image segmentation due to their ability to learn hierarchical features directly from unprocessed pixel values.

This work focuses on state-of-the-art DL models such as U-Net,VGG19-UNet, Deep ResUNet, and the SAM to segment brain tumors and predict prognosis. These models are trained to capture intricate tumor morphologies with high-resolution feature representation, often augmented with skip connections or pre-trained backbones. At the same time, we apply classical ML algorithms like SVM, Logistic Regression, Random Forests, XGBoost, and CatBoost which are especially effective for classifying derived features and helping clinical predictions, particularly in cases of limited or heterogeneous data.

In addition to improving decision support, we also explore ensemble techniques such as majority voting that leverage the strengths of many models to improve robustness and accuracy. Through the integration of deep and machine learning methods, this current work strives to offer an end-to-end methodology for brain tumor analysis with improved segmentation accuracy, reliable classification, and improved prognostic value. The ultimate goal is to aid clinicians with standardized, evidence-based information for effective treatment planning and improved patient outcomes.

2.2 Deep Learning Approaches

2.2.1 CNN

CNNs are a category of deep architecture aimed at the processing of structured information such as images and video[32]. CNNs apply convolutional layers to extract local features employing filters that travel across input data. Convolutional layers apply filters to scan for low-level features such as edges and texture in lower layers and more complicated ones towards deeper layers[32]. These filters are represented as small matrices (kernels) that perform a discrete convolution operation defined as:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n) \tag{2.1}$$

where I is the input image, K is the kernel, and S is the resulting feature map. Over time, hierarchical representations emerge as deeper layers combine simpler patterns into more complex structures. Pooling layers are eventually used to reduce the spatial sizes and introduce invariance to small shifts or distortions[32]. Common types include max pooling and average pooling, which help control overfitting and reduce computational cost.

There are fully connected layers that carry out higher-level reasoning for operations like classification and regression[18]. These layers flatten the output of previous convolutional and pooling operations and pass them through a dense neural network to output prediction scores. Dropout regularization is often applied here to prevent overfitting by randomly deactivating neurons during training.

Also,CNNs are highly parameter-sharing effective and are now the building block for most computer vision models, including medical image analysis, object detection, and image segmentation[18]. Architectures such as AlexNet, VGGNet, ResNet, and Inception have demonstrated the power of CNNs in large-scale image recognition tasks. For instance, ResNet introduced residual connections that help mitigate vanishing gradients and allow training of much deeper networks effectively [21]. In medical imaging, CNNs have shown superior performance in tumor detection, organ segmentation, and disease classification due to their ability to learn hierarchical spatial features [34].

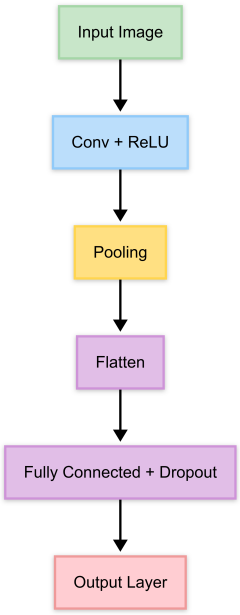

Figure 2.1: CNN Architecture Overview

2.2.2 U-Net

U-Net is a CNN architecture particularly designed for image segmentation [**?**]. U-Net was created in 2015 by Olaf Ronneberger, Philipp Fischer, and Thomas Brox and was originally employed for biomedical image segmentation but has since been employed in other fields that necessitate object boundary delineation from images accurately [66]. The model consists of symmetric U-shaped architecture with an expansive path (decoder) and a contracting path (encoder) . Context features are represented by the encoder with cascaded convolutional and pooling layers, reducing spatial sizes but increasing feature depth [3].

In practice, the encoder applies repeated blocks of two 3×3 convolutional layers followed by ReLU activations, then a 2×2 max pooling operation that reduces the spatial resolution. Each downsampling step typically doubles the number of feature maps, capturing increasingly abstract features. The decoder

path reverses this process using transposed convolutions to upsample the feature maps. At each stage, the decoder integrates high-resolution features from the encoder via skip connections, enabling the model to recover precise spatial information lost during pooling [50, 28].

The convolution operation used in each layer can be mathematically defined as:

$$Y_{i,j}^{(k)} = \sigma\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} W_{m,n}^{(k)} \cdot X_{i+m,j+n} + b^{(k)}\right) \tag{2.2}$$

where X is the input feature, $W^{(k)}$ is the filter for the k-th feature map, $b^{(k)}$ is the bias term, and σ represents the non-linear activation (typically ReLU) [28].

The feature fusion from skip connections can be represented as:

$$F_{\text{concat}} = \text{Concat}(F_{\text{encoder}}, F_{\text{decoder}}) \tag{2.3}$$

This ensures preservation of fine-grained detail that is vital in tasks like tumor segmentation and edge detection.

Spatial sizes are reversed back through up-convolutions in the decoder, and features are mixed together using skip connections to preserve localized representations accurately [**?**]. One of the greatest advantages of U-Net is that it can produce high-accuracy predictions with relatively little training data and hence is of inestimable worth in medical imaging, for example, where datasets are typically highly annotated and small in size [66]. Aside from biomedical applications, U-Net has also been utilized in self-driving cars for road detection and obstacle detection, as well as in satellite image analysis for uses such as land cover mapping and environmental monitoring [3].

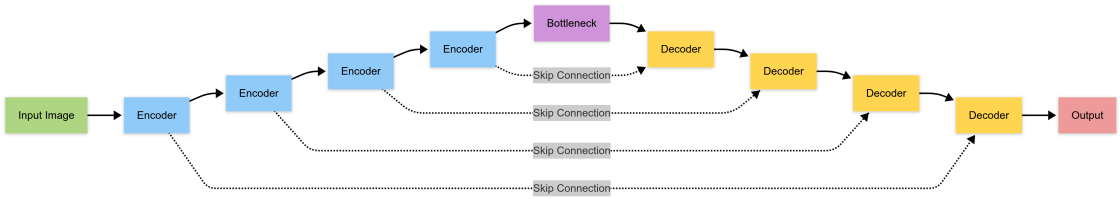

Figure 2.2: The U-Net model architecture

2.2.3 Deep ResUNet

Deep ResUNet is an enhanced deep learning architecture that combines the strengths of U-Net and Residual Networks (ResNet) to improve semantic segmentation, especially in medical and remote sensing applications. Its architecture maintains the encoder-decoder structure of U-Net but integrates residual blocks at each stage to mitigate the vanishing gradient problem and enable deeper, more expressive models. Each residual block applies a series of convolutional operations while preserving input features via identity skip connections, enhancing gradient flow and network stability. Formally, a residual block output can be expressed as:

$$y = F(x, \{W_i\}) + x \tag{2.4}$$

where x is the input, $F(x, \{W_i\})$ denotes the residual mapping learned by stacked convolutional layers, and y is the output of the block [22]. Deep ResUNet typically includes deeper encoders with downsampling through strided convolutions, followed by symmetric upsampling in the decoder, using transposed convolutions or upsampling layers. Skip connections between encoder and decoder preserve spatial resolution and contextual information. To enhance feature representation, some variants employ dilated convolutions or attention mechanisms [65]. Loss functions such as the Generalized Dice Loss are often used to address class imbalance in segmentation tasks [54]. Deep ResUNet has shown state-of-the-art results in segmenting tumors, anatomical structures, and geospatial features, outperforming traditional U-Net in both accuracy and generalization [59].

2.2.4 VGG19-UNet

The VGG19-UNet model combines a VGG19 encoder with a U-Net decoder, demonstrating versatility in domains of application. In medical imaging, it is applied for brain tumor segmentation to aid survival prediction models [41]. In cultural heritage application, it accomplishes precise semantic segmentation of Balinese dance movement [47]. Environmental science applications include waterbody segmentation from satellite imagery when combined with Otsu's preprocessing [57]. Medical uses include liver CT segmentation using multi-section fusion techniques [17]. The architecture is conceptually similar to superior models like DoubleU-Net, which enhances medical segmentation using deep convolutional networks [29].

The VGG19 network, originally proposed by Simonyan and Zisserman, is composed of 19 layers: 16 convolutional layers using 3×3 filters and 3 fully connected layers. It is known for its uniform architecture, where each convolutional layer is followed by a ReLU activation and every few layers are downsampled by max pooling. This structure enables the capture of hierarchical visual features, from low-level edges to high-level semantics, essential in segmentation tasks [52].

In the VGG19-UNet configuration, VGG19 acts as the encoder. At each stage of the encoder, feature maps are passed through two convolutional layers followed by ReLU activation and then downsampled using a 2×2 max pooling layer with stride 2 [52]. The decoder mirrors the encoder by upsampling the feature maps using transposed convolutions and concatenating them with the corresponding encoder outputs via skip connections [50]. This fusion preserves spatial context, which is crucial for precise segmentation [50].

The convolutional operation in each layer can be expressed as:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n) \tag{2.5}$$

where I is the input image, K is the convolutional kernel, and $S(i,j)$ is the output feature map [50].

2.2.5 SAM

SAM is a breakthrough in image segmentation foundation models that indicates potential and shortcomings in medical applications. As a Vision Transformer (ViT) model prompted [62], SAM possesses very good zero-shot generalization capability which can be applied across various medical imaging modalities [63]. The model is particularly robust in reducing annotation burden without compromising segmentation accuracy [26], though its performance significantly varies as a function of different anatomical structures and imaging modalities [26, 1]. Though SAM's clinician-controlled interactive prompting mechanism facilitates flexible segmentation [63], issues remain concerning dealing with fine pathological details and 3D volumetric consistency [1]. Current studies highlight SAM's versatility as a generic computer vision-bridging technology capable of being used for medical image analysis [62], yet specialized adaptations usually need to be made to address the unique requirements of clinical imaging tasks [26, 1]. The model's fine-tunability and compatibility with semi-supervised approaches reflect growing promise for medical applications [1], despite current limitations in handling complex medical imaging cases [26, 63].

Original SAM Architecture

SAM's architecture comprises three main components:

- **Image Encoder:** A ViT that processes the input image to generate latent embeddings
- **Prompt Encoder:** Translates user inputs (points/boxes/text) into embedding vectors
- **Mask Decoder:** Predicts segmentation masks from combined image and prompt embeddings [30]

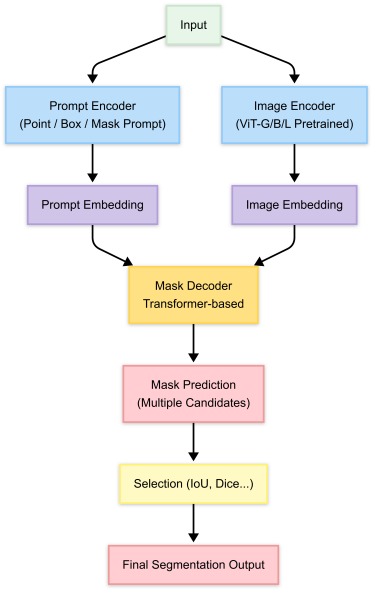

Figure 2.3: Original SAM architecture

Key Mathematical Formulations:

$$M = D(E_I(I), E_P(P)) \tag{2.6}$$

where I is the input image, P represents prompts, E_I/E_P are encoders, D is decoder, and M is output mask [30].

Loss Functions:

$$\text{BCE} = -[y \log(\hat{y}) + (1-y)\log(1-\hat{y})] \tag{2.7}$$

$$\text{Dice} = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{2.8}$$

$$\text{Final Loss} = \text{BCE} + \text{Dice} \tag{2.9}$$

Fine-Tuned SAM

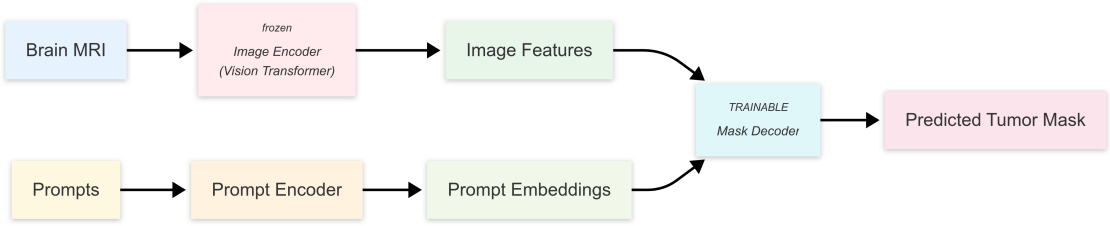

Figure 2.4: Fine-tuned SAM architecture

Medical Adaptations:

In medical imaging, adaptations of SAM, such as MedSAM, have been proposed to address domain-specific challenges. MedSAM incorporates medical domain knowledge to improve performance on tasks like organ and tumor segmentation [37]. Furthermore, studies have explored the integration of SAM with semi-supervised learning techniques to leverage limited annotated data effectively, enhancing its applicability in clinical settings [64].

2.2.6 MedSAM

Medical Segment Anything Model (MedSAM) represents a significant advancement in medical image segmentation, leveraging the capabilities of general-purpose foundation models for specialized medical tasks [37]. It is an adaptation of the Segment Anything Model (SAM), a promptable segmentation system developed by Kirillov et al. (2023) from Meta AI, designed to provide high-quality segmentation masks for a wide variety of objects in general images given various input prompts [30]. MedSAM aims to transfer the powerful zero-shot generalization capabilities of SAM to the medical domain, addressing the challenge of segmenting diverse anatomical structures and pathologies across different imaging modalities [37, 36].

The architecture of MedSAM, inherited from SAM, consists of three main components: an image encoder, a prompt encoder, and a lightweight mask decoder [30].

- The image encoder is typically a Vision Transformer (ViT) (e.g., ViT-H) pre-trained on extensive datasets. It processes the input image into high-dimensional image embeddings that capture rich visual information. The ViT divides an image into patches, linearly embeds them, adds positional embeddings, and then processes them through a series of Transformer blocks.

- The prompt encoder handles various types of input prompts that specify the object to be segmented. These prompts can include points (foreground/background), bounding boxes, segmentation masks (coarse), or free-form text. The encoder converts these sparse or dense prompts into prompt embeddings.

- The mask decoder then efficiently maps the image embedding and prompt embeddings to output segmentation masks. It typically uses a Transformer decoder architecture, employing mechanisms like self-attention on prompt embeddings and cross-attention between prompt embeddings and image embeddings to interpret the prompts in the context of the image, ultimately predicting the object mask(s) [30]. This decoder is designed to be lightweight, allowing for near real-time interaction.

A core component of the Transformer blocks used in both the image encoder and the mask decoder is the attention mechanism, particularly scaled dot-product attention. This can be mathematically represented as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.10}$$

where Q (Query), K (Key), and V (Value) are matrices derived from the input embeddings, and d k

is the dimension of the key vectors. This mechanism allows the model to weigh the importance of different parts of the input when producing an output .

The adaptation of SAM to MedSAM often involves fine-tuning the SAM model on large-scale medical image datasets encompassing diverse modalities (e.g., Computed Tomography (CT), Magnetic Resonance Imaging (MRI), ultrasound, X-ray) and tasks (e.g., organ, tumor, lesion segmentation) [37, 36]. This specialization enhances its performance on medical-specific structures while retaining strong generalization.

One of the key advantages of MedSAM is its ability to perform zero-shot or few-shot segmentation for many medical tasks by leveraging appropriate prompting, significantly reducing the need for extensive task-specific annotations and model training [37]. This makes it highly valuable for a wide range of medical imaging applications where data annotation is a bottleneck. MedSAM has demonstrated promising results in segmenting various anatomical structures, tumors, and other pathologies, showcasing its potential to serve as a versatile tool for medical image analysis and support clinical workflows [37, 36].

2.3 Standard Classifiers

2.3.1 SVM

SVM is a supervised learning algorithm primarily used for classification tasks, but it can also be adapted for regression. At its core, SVM aims to find the optimal hyperplane that maximally separates data points belonging to different classes in a high-dimensional space. This is achieved by maximizing the margin between the closest data points from each class, known as support vectors, through the optimization problem:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to} \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \forall i \tag{2.11}$$

In cases where the data is not linearly separable, SVM employs kernel functions such as radial basis function (RBF), polynomial or sigmoid kernels to project the data into a higher-dimensional space where a separating hyperplane can be established. Architecturally, SVM does not follow a layered structure like neural networks but operates through mathematical optimization, often formulated as a convex quadratic programming problem. The effectiveness of SVM has been demonstrated in various domains, including bioinformatics, image classification, and text categorization, particularly when the dataset is high-dimensional but relatively small in size [9]. Its robust generalization capability, especially with limited training data, makes it a competitive baseline model in many machine learning pipelines [25].

2.3.2 Logistic Regression

Logistic Regression is a widely used statistical model that serves as a foundational method for binary classification problems. It estimates the probability that a given input belongs to a particular category by modeling the log-odds as a linear function of the input variables. Unlike linear regression, which predicts continuous outputs, logistic regression maps predicted values to probabilities using the logistic (sigmoid) function [24, 10]. This property makes it particularly suitable for tasks such as disease prediction, credit scoring, and document classification [24]. The core of the model is the transformation:

$$P(y = 1 \mid \mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}} \tag{2.12}$$

where $P(y = 1 \mid \mathbf{x})$ denotes the probability of the positive class given input features $\mathbf{x}$, and $\beta_0, \ldots, \beta_n$ are the model coefficients learned via maximum likelihood estimation [40]. One of the key strengths of logistic regression is its interpretability, as the coefficients directly relate to the log-odds of the outcome [24]. Furthermore, regularization techniques such as L1 (Lasso) and L2 (Ridge) can be incorporated to improve generalization and handle multicollinearity in high-dimensional data [56]. Despite its simplicity, logistic regression often performs competitively with more complex models when the relationship between predictors and the log-odds is approximately linear [24].

2.4 Ensemble Learning

2.4.1 Random Forest

Random Forest is an ensemble learning algorithm widely used for both classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the class that is the mode of the classes (for classification) or mean prediction (for regression) of the individual trees [20]. The architecture of a Random Forest involves randomly selecting subsets of the dataset and subsets of features for each decision tree, a method known as bootstrap aggregating or "bagging." This randomization helps in reducing variance and overfitting compared to individual decision trees [33].

Each tree in the forest is trained on a bootstrap sample drawn with replacement from the original data. At each node, a random subset of features is chosen, and the best split is made only among those features, which decorrelates the trees and enhances ensemble robustness [23]. The prediction for classification tasks is made using a majority voting scheme across all trees in the forest. This can be expressed mathematically as:

$$\hat{y} = \text{mode}\left\{h_1(x),\ h_2(x),\ \ldots,\ h_T(x)\right\} \tag{2.13}$$

where $\hat{y}$ is the final prediction, $h_t(x)$ is the prediction from the t^{th} tree, and T is the total number of trees. Random Forests are particularly effective when dealing with high-dimensional data, noisy features, and imbalanced datasets, and they have been successfully applied in domains such as bioinformatics, remote sensing, and medical diagnostics [11, 2].

2.4.2 XGBoost

Extreme Gradient Boosting (XGBoost) is a powerful and scalable implementation of gradient boosted decision trees, proposed by Chen and Guestrin in 2016. Designed with both efficiency and accuracy in mind, XGBoost has become one of the most successful algorithms in structured data machine learning competitions and applications. It operates by building an ensemble of decision trees in a sequential manner, where each new tree attempts to correct the residuals (errors) of the previous ensemble using a gradient descent approach on a specified loss function.

The objective function in XGBoost can be formally expressed as:

$$L(\phi) = \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2.14}$$

where l is a differentiable convex loss function (e.g., logistic loss), $\hat{y}_i$ is the prediction for instance i, f_k represents the k-th tree in the ensemble, and $\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2$ is a regularization term penalizing model complexity (with T being the number of leaves and w the leaf weights) [7].

XGBoost introduces innovations such as sparsity-aware learning, weighted quantile sketch for approximate tree split finding, and cache-aware block structures for memory efficiency. These features significantly improve computational speed and memory usage without compromising predictive performance [7, 4]. XGBoost has been widely adopted across domains such as bioinformatics, fraud detection, and medical diagnostics due to its robustness, scalability, and ability to handle missing data and class imbalance effectively [43].

2.4.3 CatBoost

CatBoost is a high-performance gradient boosting algorithm developed by researchers at Yandex, specifically optimized for handling categorical features without the need for extensive preprocessing such as one-hot encoding or label encoding [46]. It implements ordered boosting, a novel technique that mitigates the prediction shift problem caused by traditional target leakage in gradient boosting [15]. Unlike other gradient boosting libraries, CatBoost uses symmetric trees, meaning that the same split condition is applied at each level of the tree across all branches, which enhances both inference speed and regularization [46].

A key innovation of CatBoost is its approach to categorical variables. Instead of static encodings, it uses target statistics with a permutation-driven scheme to reduce overfitting. For a categorical feature x, the target statistic is calculated as:

$$\mathrm{TS}_i = \frac{\sum_{j=1}^{i-1} \mathbb{1}[x_j = x_i] \cdot y_j + a \cdot P}{\sum_{j=1}^{i-1} \mathbb{1}[x_j = x_i] + a} \tag{2.15}$$

where TS_i is the target statistic for the i^{th} data point, a is a smoothing parameter, and P is the prior [15].

CatBoost also introduces gradient-based feature combinations, which allow the model to automatically construct new categorical features during training by combining existing ones, further improving model expressiveness [15]. It has demonstrated strong empirical performance in fields such as recommendation systems, natural language processing, and medical data analysis due to its capability to handle heterogeneous and high-cardinality data effectively.

2.4.4 Majority voting

Majority voting is a widely used ensemble technique in machine learning where predictions from multiple classifiers are aggregated to make a final decision based on the most frequently predicted class label. It operates under the assumption that while individual classifiers may be weak, their combined decisions can enhance accuracy, particularly when the models are diverse and uncorrelated [31, 45]. Formally, majority voting is defined as:

$$\hat{y} = \arg\max_{c \in C} \sum_{t=1}^{T} \mathbb{1}(h_t(\mathbf{x}) = c) \tag{2.16}$$

where $\hat{y}$ is the final predicted class, C is the set of possible classes, $h_t(\mathbf{x})$ is the prediction of the t^{th} classifier, and $\mathbb{1}(\cdot)$ is the indicator function that returns 1 if the classifier predicts class c, with T being the total number of classifiers. Variants of majority voting include hard voting (based on discrete labels) and soft voting, where the predicted class probabilities are averaged and the class with the highest probability is selected [55, 49]. Majority voting plays a central role in ensemble methods such as Bagging and Random Forests, and has shown significant success in fields like bioinformatics, medical imaging, and document classification [14].

2.5 Conclusion

This chapter has established the foundational knowledge required to understand the complex landscape of automated brain tumor analysis. It began by outlining the critical clinical need for accurate and efficient segmentation and prognostic tools, highlighting the limitations of manual methods and the promise of computational approaches. The discussion then delved into the two core technological pillars that underpin this thesis: deep learning for image analysis and machine learning for classification.

The discussion covered a spectrum of state-of-the-art deep learning architectures, from the foundational U-Net and its powerful variants like Deep ResUNet, to the transformative potential of large-scale models like SAM and its medical adaptation, MedSAM. These models represent the cutting edge of semantic segmentation, capable of learning the intricate spatial hierarchies necessary to delineate tumors from complex anatomical backgrounds.

Complementing the deep learning models, the chapter also reviewed a suite of standard and advanced machine learning classifiers, including SVM, Logistic Regression, Random Forest, XGBoost, and CatBoost. These algorithms are instrumental for the second phase of the prognostic analysis where features derived from imaging and other data are used to predict clinical outcomes. The principles of ensemble learning, particularly majority voting, were also introduced as a strategy to enhance predictive robustness and accuracy.

In essence, this chapter has provided a comprehensive theoretical and technical toolkit. This foundational knowledge serves as the basis for the specific methodological choices, experimental designs, and novel contributions presented in the subsequent chapters of this work.

Chapter 3

Literature review

3.1 Overview on Brain Cancer

Brain cancer is a heterogeneous group of malignant and benign brain or central nervous system (CNS) tumors and is one of the most complex and life-altering forms of cancer from a clinical standpoint. Its diagnosis, prognosis, and treatment are very challenging due to the complex anatomy of the brain, the aggressiveness of most tumors, and the variability of patient responses to therapy. MRI is the gold standard for non-surgical assessment of tumors, with superior contrast resolution and the ability for imaging anatomical and pathologic processes in multiple planes of imaging. More recent MRI modalities—such as Diffusion-Weighted Imaging (DWI), Perfusion-Weighted Imaging (PWI), Magnetic Resonance Spectroscopy (MRS) and functional Magnetic Resonance Imaging (fMRI) have contributed additional diagnostic power by enabling measurements of cellular density, vascularity, metabolic profiles, and cortical function [13, 60]. These modalities are helpful pre-surgical information and aid in tumor grading, yet histopathological analysis is the gold standard for definitive diagnosis and classification [35]. Gliomas are among the most clinically influential brain tumors, arising from glial cells and possessing a wide degree of aggressiveness. LGG are less malignant and slow growing and more inclined to be linked with favorable genetic mutations such as IDH1/2 and 1p/19q co-deletion [42]. In contrast, however, HGG, like glioblastoma multiforme (GBM), are characterized by their aggressive nature, therapeutic resistance, and poor survival despite multimodality treatment. The standard treatment today consists of a combination of maximal safe surgery, radiotherapy, and chemotherapy, to which targeted therapy and immunotherapy are now being added in suitably selected cases [53]. The integration of imaging, molecular biomarkers, and clinical data has become increasingly important to personalized patient treatment plans and survival [42, 53].

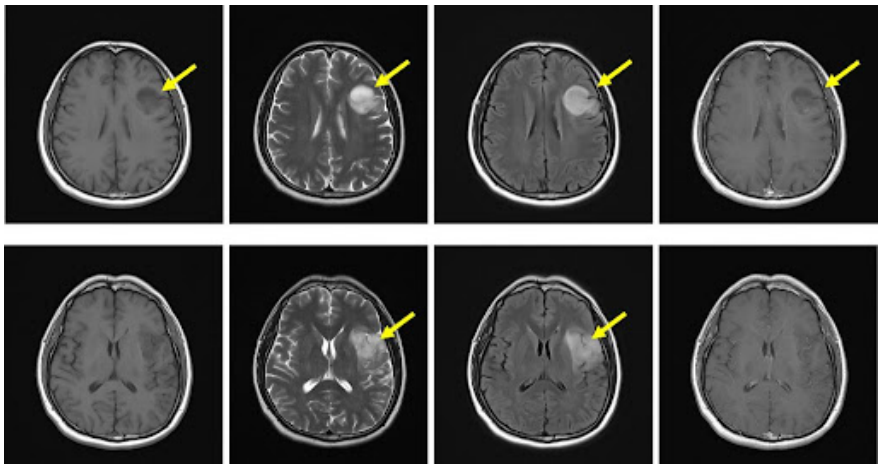

Figure 3.1: LGG on Multi-Sequence MRI

3.2 Related Works

In recent years, DL has transformed the field of medical image segmentation by offering automated, highly accurate, and efficient tumor detection methods. In contrast, CNNs and transformer-based models have demonstrated remarkable progress in achieving reliable and consistent segmentation results. This paper explores the latest advancements in DL for brain tumor segmentation, highlighting key architectures, methodologies, and their impact on clinical applications.

Recently, X. Han et al. [19] proposed that U-CCNet is an enhanced version of the U-Net architecture to improve the precision of brain tumor segmentation with the introduction of cross-attention mechanisms. Supervised training and model testing were performed on the LGG segmentation dataset (TCGA-LGG), which is the most widely used benchmark for low-grade glioma segmentation. Results show that the proposed model performs well in MRI segmentation tasks of brain tumors. Through the use of attention mechanisms, U-CCNet gets better attention on tumor regions across different MRI slices with the improved registration of their spatial interactions. Perpendicular feature extractions at an individual level by classical-U-Net-type architecture. U-CCNet further helps to improve communication at the encoding-decoding level, thereby elevating segmentation accuracy and edge detection. The U-CCNet model offers several key advantages, including discrimination between tumor and non-tumor areas with cross-attention, enhanced segmentation performance across various MRI sequences, increased generalizability, and lower computational cost compared to 3D models. Performance evaluations demonstrated that U-CCNet outperformed traditional U-Net and CNN-based approaches in terms of DSC and IoU scores.

S. Pereira et al. [44] proposed a CNN-based segmentation with tiny convolutional kernels 3×3. The model was created to overcome the difficulties related to brain tumor segmentation by allowing deeper networks with fewer parameters. This makes the model less prone to overfitting and keeps fine tumor structure details intact. The study also introduced pre-processing techniques to enhance model performance, such as intensity normalization, which normalizes MRI scans to diminish patient differences, and data augmentation, which artificially expands the data set to enhance model robustness. The experimental results show that the CNN-based method achieved a DSC of 78. 5% for the entire tumor, 65.1% for the core tumor, and 76.2% for the enhancing tumor, which are the best compared to previous work. The use of small convolutional kernels allowed the model to learn more localized spatial features, defining the tumor boundary more precisely. In addition, pre-processing techniques played an important role, in which intensity normalization reduced the heterogeneity of the MRI intensity between patients to achieve more consistent segmentation performance. Data augmentation also improved the robustness, in which the model would be usable in unseen MRI scans. Although good performance was achieved, some limitations, such as the misclassification of small tumor regions and the inability to segment very heterogeneous tumors, were acknowledged.

In another work [58], B. Wan et al. explore advanced methods for glioma segmentation and propose a new model that overcomes the problem of erratic results in the quality of segmentation of fully CNN based on DL. They do not account for the peculiar genomic and background data of the patients with respect to brain gliomas, potentially granting space for wrong diagnosis and treatment. The proposed model tackled this issue by enhancing the overall structure and incorporating an innovative loss function. In the beginning, DeepLabv3+ was used as the model's overall architecture with RegNet as an image encoder, and an attribute encoder module was proposed to integrate the patient's genomic background and basic attributes with image depth information which fused into a 2D CNN, this is complemented by the image encoder and atrous spatial pyramid pooling into the encoder module dealing with the multimodal batch fusion problem. Moreover, linear weighting a combined cross-entropy loss, and Dice loss were applied to address the sample imbalance problem. An innovative loss function is proposed to suppress certain sizes of regions to avoid segmentation errors of noise-like regions, hence, the resultant segmentations are more stable. Experiments were conducted on the Lower-Grade Glioma Segmentation Dataset, a widely used benchmark for brain tumor segmentation. The effectiveness of the proposed approach was confirmed with rigorous testing, where the method achieved a 94.36 DSC and IoU score of 91.83.

Drawing on previous work, a research about MRI Segmentation in Brain Low-Grade Gliomas Using SVM and CNN by Q. Yang et al. [61] presents the application of machine learning methods specifically SVM and CNN, to segment brain LGG from MRI images, a primary operation in diagnosis, classification, and treatment planning. The research compares the segmentation ability of these models between efficiency, computational cost, and accuracy. SVM technique employs training 109 models individually, with each

model being constructed for one patient's MRI scan to provide fast segmentation in seconds. While SVM achieves an average accuracy of 0.937 and an F1 score of 0.546, its precision is 0.456, showing that SVM is not able to distinguish tumor and non-tumor areas. On the other hand, the CNN model, trained for two hours with a very large dataset of 19,760 augmented images, demonstrates superior segmentations with almost perfect accuracy (0.998) and F1 score (0.999). Although CNN is far better than SVM, it requires a high-performance computing environment and significantly longer training durations.

A. Shomirov et al. [51] titled Brain Tumor Segmentation of HGG and LGG MRI Images Using Weighted Focal Loss (WFL)-Based 3D U-Net is dedicated to the urgent task of accurate brain tumor segmentation from MRI images as a pre-requisite for effective treatment planning. The authors propose an augmentation of the base 3D U-Net model by adding a WFL function. This novel loss function attempts to reverse the common issue of class imbalance in medical images, namely underrepresentation of the minority tumor areas.

According to their method, the authors assign greater weights to minority classes and lower weights to majority classes of the WFL model. Their weighting strategy can give high enough importance to less prominent regions of tumors by the model, and thus improve the accuracy of segmentation. They trained on the 2019 and 2020 Brain Tumor Segmentation Challenge (BraTS) datasets, i.e., HGG and LGG cases.

The outcome indicates that the performance values achieved by the developed WFL-based 3D U-Net model are of top class. On BraTS 2019 images, the mean Dice value is achieved by the model as 0.830 for Tumor Core (TC), 0.913 for Whole Tumor (WT), and 0.815 for Enhancing Tumor (ET) in the case of HGG tumors. For LGG tumors, mean Dice values are 0.731 for TC, 0.775 for WT, and 0.685 for ET. Similarly, BraTS 2020 results of data reported Dice values of 0.843 (TC), 0.892 (WT), and 0.871 (ET) for HGG and 0.7501 (TC), 0.7985 (WT), and 0.6103 (ET) for LGG.

These findings suggest that the inclusion of the WFL in the 3D U-Net model adequately corrects class imbalance, and the outcome is improved segmentation of high-grade and low-grade gliomas.

Another research, based on the U-Net++ Network with EfficientNet Encoder by Y. Chen et al. [8], proposes a DL approach to brain tumor segmentation through the expansion of the U-Net++ model by using EfficientNet as an encoder. EfficientNet, with enhanced feature extraction ability and cost of computation reduction, replaces the conventional encoder with more effective learning of MRI image fine details by the model. The authors then optimize U-Net++ by eliminating dense skip connections, reducing computational complexity without losing valuable spatial information from retained feature map connections on the same resolution level. This design adjustment fills the semantic gap between low-level and high-level features, yielding improved tumor segmentation.

In optimizing the model during training, the Adam optimizer was employed since it is efficient, has minimal memory usage, and employs first-order gradients. EfficientNet model was pre-trained on the ImageNet weights, which provided a good initialization without causing overfitting. The dataset was split into a training, validation, and test datasets in a ratio of 8:1:1, and the model was trained for 100 epochs with an initial learning rate of 0.001, which was updated dynamically. Cosine annealing was also applied to prevent the model from getting trapped in local minima.

For calculation, the DSC was utilized to quantify segmentation accuracy since it is a suitable metric of ground truth to predicted segmentation similarity. The experiment was conducted 10 times, and the mean DSC was calculated as 0.9180, which suggests that the model performed well for the segmentation of brain tumors. EfficientNetB4 encoder as feature extractor was employed in creating feature maps of varying resolutions (16×16, 32×32, 64×64, and 128×128) to feed the feature maps to the U-Net++ network. Binary cross-entropy loss and Dice loss were employed in the effort to enhance the accuracy of segmentation. Visualization of the segmentation output confirmed the model's accuracy in tracing tumor borders correctly, as witnessed from the comparisons between predicted masks and ground truth annotations.

As proposed by R. Ramin et al. [48] in the study of Brain Tumor Segmentation based on DL and attention mechanism using multimodality MRI to enhance brain tumor segmentation accuracy through a Distance-Wise Attention mechanism and multimodal MRI (T1, T1C, T2, and FLAIR). A two-path CNN model and attention mechanism, along with preprocessing method,s are proposed in the paper for improved segmentation. The experiments reveal that the simple two-path CNN model without preprocessing fails to segment tumor areas properly. Even though incorporating the attention mechanism separately improves segmentation for all tumor areas, incorporating both preprocessing and attention leads to phenomenal improvement in accuracy. Specifically, Dice scores in three tumor regions improve from 0.2531, 0.2796, and 0.2143 to 0.8756, 0.8550, and 0.8715 for enhancing tumor (End), entire tumor, and TC, respectively. The

study also comes to indicate that the quality of segmentation is more subject to preprocessing in comparison to the attention mechanism when dealing with small tumor areas. Additionally, local features are better than global features in providing a precise segmentation.

C. Sekaran et al. [16] proposed IC-Net (Inverted-C) network architecture is utilized in the research, capitalizing on emerging segmentation methods integrated to improve local and global contextual feature extraction to establish the tumor boundary with precision. Experimental result indicated IC-Net working significantly better compared to conventional U-Net models with enhanced correct segmentation and resistance in various locations of tumors. While enhanced, the issue of hyperparameter optimization, data restriction, and limitation of computational resource was identified.

A novel brain tumor segmentation from MRI images using the hybrid CNN is another attempt by D. Daimary et al. [12] to improve the segmentation of brain tumors by hybrid DL models. In this regard, three hybrid neural network architectures have been proposed: Seg-UNet, Res-SegNet, and U-SegNet. This is achieved by combining the features from SegNet, U-Net, and ResNet18. The Seg-UNet model combines SegNet and U-Net and is innovative in its approach regarding feature extraction because it uses an upsampling and downsampling method to ensure better extraction of features. Res-SegNet incorporates one element-wise addition layer, originated from ResNet18, and performs better on the target datasets than the baseline architectures. These models were trained and validated using the BraTS dataset with accessible preprocessing methods, such as 2D slice extraction, intensity normalization, and data augmentation.

The proposed hybrid architectures, however, were compared with some of the most widely known models in the same category: SegNet3, SegNet5, and U-Net, for the purpose of evaluation of segmentation capability. Every model has a neuronal number of 172,800 in the input layer with multiple hidden layers. Features were classified in the output layer into four separate categories: enhancing tumor (green), necrotic & non-enhancing tumor (red), peritumoral edema (yellow), and background (gray). The reported results were analyzed using ground truth comparison images. The segmented images were shown to be considerably superior to classical CNNs for brain tumor segmentation. Among the segmented models proposed, Seg-UNet excelled, achieving 99.1% overall accuracy and 73.4% mean IoU, outperforming both baseline SegNet and U-Net models. These architectures require significantly longer training time due to their higher complexity; however, they demonstrate a considerable improvement in the accuracy of the segmentation.

D. Maji et al. [38] developed Attention Res-UNet with Guided Decoder (ARU-GD), which proposes a guided decoder that explicitly supervises every decoder layer, helping the feature learning process. It also integrates attention gates inside the Res-UNet backbone, which helps to refine the activation of relevant areas while filtering out unimportant information. The authors trained and evaluated their model using the BraTS 2019 HGG dataset and preprocessed MRI scans to standardize them. A weighted guided loss function was leveraged to ensure better prediction accuracy at each of the decoder layers while training the network. Experimental results revealed the supremacy of ARU-GD over baseline architectures such as UNet, Res-UNet, and Res-UNet added with attention gates, as well as DL models like VGG-Net, MobileNet, QuickNAT, DenseNet, and XceptionNet. Worth noting is that the proposed model achieves the Dice score of 0.911, 0.876, and 0.801 for WT, TC, and enhancing tumor segmentation, respectively, with mean IoU scores of 0.838, 0.781, and 0.668.

F. Isensee et al. [27] applied nnU-Net to the BraTS 2020 challenge and demonstrated its flexibility and strong performance. Their effort started with the baseline nnU-Net configuration, which already showed competitive performance. To further improve segmentation accuracy, they incorporated BraTS-specific enhancements, such as postprocessing tweaks, region-based training, and increased data augmentation, along with some other small adjustments. Additionally, they re-implemented the BraTS ranking scheme to formally compare the best-performing nnU-Net variant using competition standards. Their method won BraTS 2020 with Dice scores of 88.95, 85.06, and 82.03 for WT, TC, and enhancing tumor, respectively, and Hausdorff Distance (HD95) measurements of 8.498, 17.337, and 17.805. These results emphasize the potential of nnU-Net for precise brain tumor segmentation, especially when adapted to characteristics in specific datasets.

Ref	Architecture	Dataset	Score
Xiaowei Han et al. [19]	U-CCNet	TCGA-LGG	Improved DSC and IoU
Sérgio Pereira et al. [44]	CNN	BraTS	DSC: 78.5% (WT) 65.1% (TC) 76.2% (ET)
Bing Wan et al. [58]	DeepLabv3+ with RegNet	LGG Segmentation Dataset	DSC: 94.36 IoU: 91.83
Qifan Yang et al. [61]	SVM and CNN	LGG MRI Dataset	SVM: Accuracy 0.937 , F1 0.546 CNN: Accuracy 0.998, F1 0.999
Shomirov et al. [51]	WFL-Based 3D U-Net	BraTS 2019/2020	BraTS 2019: DSC 0.913 (WT), 0.830 (TC), 0.815 (ET)
Yunyi Chen et al. [8]	U-Net++ with EfficientNet Encoder	BraTS	DSC: 0.9180
Ranjbarzadeh et al.[48]	CNN with Distance-Wise Attention	Multimodal MRI (T1, T1C, T2, FLAIR)	DSC: 0.8756 (ET) 0.8550 (WT) 0.8715 (TC)
Chandra Sekaran D S et al.[16]	IC-Net	BraTS2020	Better segmentation and tumor localization
Dinthisrang Daimary et al. [12]	Hybrid CNNs (Seg-UNet, Res-SegNet, U-SegNet)	BraTS	Seg-UNet: Accuracy 99.1% , IoU 73.4%
Dhiraj Maji et al. [38]	Attention Res-UNet with ARU-GD	BraTS2019 (HGG)	DSC: 0.911 (WT); 0.876 (TC); 0.801 (ET) IoU:0.838(WT); 0.781(TC); 0.668(ET)
Fabian Isensee et al [27]	nnU-Ne	BraTS2020	DSC: 88.95 (WT); 85.06 (TC); 82.0 (ET)

Table 3.1: Summary of related works on brain tumor segmentation using DL approaches.

3.3 Conclusion

The literature surveyed in this chapter confirms that the field of automated brain tumor segmentation is a dynamic and rapidly advancing area of research. A comprehensive review of recent studies reveals several key trends and persistent challenges.

A predominant theme is the widespread adoption and evolution of the U-Net architecture. Many success-

ful approaches, such as U-CCNet, 3D U-Net, and U-Net++, build upon its foundational encoder-decoder structure. To enhance performance, researchers have integrated advanced techniques, including attention mechanisms to improve feature localization, specialized loss functions like WFL to combat class imbalance, and the use of powerful pre-trained backbones like VGG19, RegNet, and EfficientNet to serve as more effective feature extractors.

The studies consistently demonstrate the superiority of DL models, particularly CNNs and their variants, over traditional ML methods like SVM in terms of segmentation accuracy, although often at a higher computational cost. Furthermore, the literature highlights innovative strategies such as incorporating non-imaging patient data and developing hybrid models to push the boundaries of segmentation precision.

Despite these significant advancements, common challenges remain. These include accurately segmenting heterogeneous tumor sub-regions, managing the high computational demands of complex models, and ensuring robust generalization across different datasets and imaging protocols. This review of the state-of-the-art clarifies the landscape of existing solutions and identifies the remaining gaps. The work presented in this thesis is positioned to address these challenges, thereby contributing to the ongoing effort to develop more accurate and clinically viable tools for brain tumor analysis.

Chapter 4

Methodology

4.1 Overview

This chapter outlines the methodological framework adopted in this study, which is divided into two main components: Segmentation model development and implementation and Prognostic analysis based on extracted features. The process involves deploying multiple deep learning models for tumor segmentation, followed by feature-based classification to assess patient outcomes.

4.2 Dataset Description

The dataset used in this study is the LGG Segmentation Dataset [6], comprising brain MR images accompanied by manual FLAIR abnormality segmentation masks. The images were obtained from The Cancer Imaging Archive (TCIA) and correspond to 110 patients included in the The Cancer Genome Atlas (TCGA) LGG collection. Figure 4.1 shows a representative sample from the dataset.

This dataset has been used in the following studies:

- Mateusz Buda et al., Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm.[5]
- Maciej A. Mazurowski et al., Radiogenomics of lower-grade glioma: algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with The Cancer Genome Atlas data.[39]

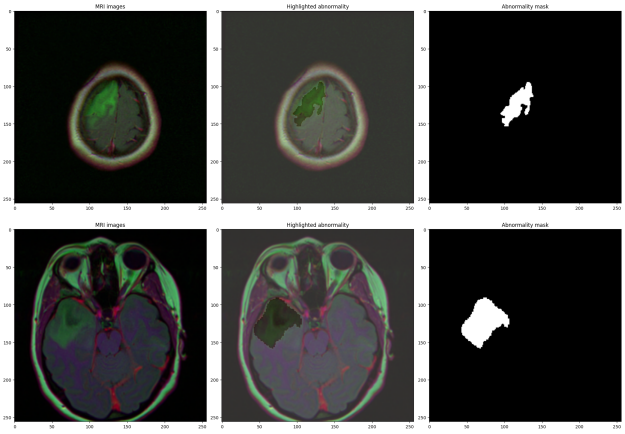

Figure 4.1: Representative Sample from the Dataset

The images include at least one Fluid-Attenuated Inversion Recovery (FLAIR) sequence per patient and are paired with genomic cluster labels and survival data. A file named `data.csv` provides associated tumor genomic cluster information and clinical variables.

4.3 Models Implementation

This study applies and contrasts four of the latest neural architectures for brain tumor segmentation, selected on the basis of their demonstrated performance in medical image tasks. The pipeline advances stepwise from data acquisition to prognosis analysis, with rigorous testing at each step.

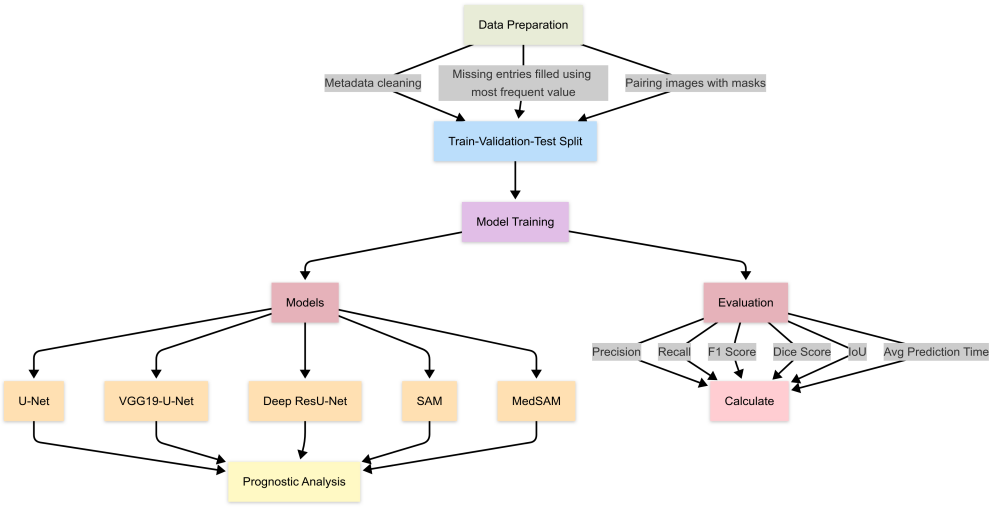

Figure 4.2: General Overview of our Methodology

4.3.1 Data Preparation

The MRI dataset used in this study included a structured CSV metadata file alongside the raw brain scan images. Prior to model training, a preprocessing step was carried out to ensure data integrity and consistency across all samples.

An initial inspection of the metadata file revealed the presence of missing values. To address this, a missing value imputation strategy was applied. Specifically, missing entries were filled using the most frequent value in each column. This approach preserves the distribution of categorical and ordinal features without introducing artificial noise. The imputation process was carried out using a frequency-based statistical strategy, ensuring that no column remained incomplete before further processing.

Following the metadata cleaning, a randomized sampling procedure was applied to the available MRI scan files to select a representative subset for segmentation tasks. Each MRI image was paired with its corresponding segmentation mask, and both were loaded into a custom data structure for downstream processing.

This cleaned and structured dataset was subsequently passed into a preprocessing pipeline for image and mask transformation, followed by a stratified train-validation-test split to maintain representative distributions across all subsets.

4.3.2 U-Net

1.3.2.1 Implementation

A custom U-Net architecture was implemented from scratch in PyTorch, tailored for binary segmentation of brain tumors in MRI images. The network follows an encoder-decoder structure with skip connections

and includes a central bridge module to enhance context preservation between compression and expansion paths. The network is composed of the following:

- Double convolution blocks are employed to enhance feature learning.
- An encoder (downsampling path) is included to extract contextual information.
- A bridge module connects the encoder and decoder layers.
- A decoder (upsampling path) reconstructs spatial details.
- It takes 3-channel RGB MRI slices as input.
- It outputs a 1-channel mask for binary tumor segmentation.

1.3.2.2 Training and validation loss

Training and validation losses were recorded after each epoch to track learning progress. The training loss consistently decreased, while the validation loss also decreased and then largely stabilized, indicating the model's convergence.

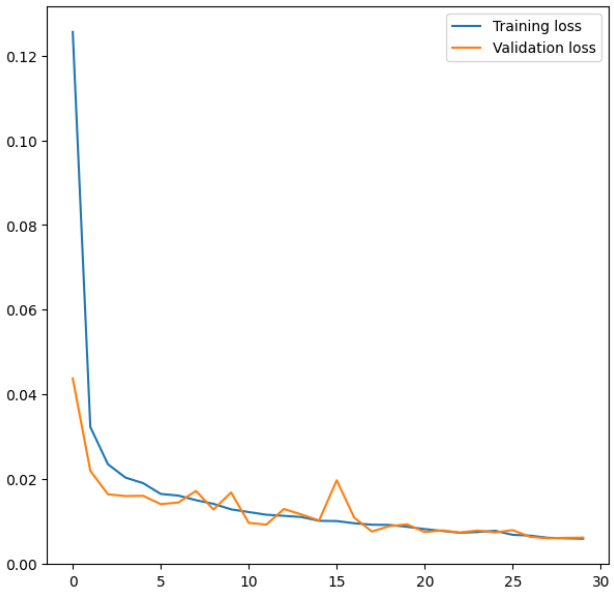

Figure 4.3: Training and Validation Loss Across Epochs

1.3.2.3 Results

As shown in Figure 4.12, the U-Net model has successfully segmented the brain tumor indicating that the it has learned to accurately identify and delineate the tumor region within the input MRI image.

Figure 4.4: Sample U-Net Brain Tumor Segmentation Result

The final model was evaluated on the held-out test set using a suite of semantic segmentation metrics, all relevant to binary classification tasks at the pixel level:

Metric	**Score**
Precision	0.930
Recall	0.791
F1 Score	0.855
Dice Score	0.855
IoU	0.746
Avg Prediction Time (s)	0.004

Table 4.1: U-Net Evaluation Metrics and Results

These results demonstrate excellent precision. The F1 score and Dice coefficient validate that the predicted tumor regions closely match the ground truth annotations. Moreover, the extremely low inference time suggests the model is viable for near-real-time applications in clinical or research workflows.

4.3.3 Deep ResUNet

1.3.3.1 Implementation

The network architecture used in this study is characterized by the following components:

- **Encoder (Downsampling Path):**
  - Composed of a series of blocks, each typically containing convolutional layers followed by residual blocks.
  - These blocks are designed to learn hierarchical features and important tumor patterns from the input MRI images.
  - With each subsequent block, the spatial dimensions of the feature maps are reduced (e.g., via max-pooling or strided convolutions), while the feature depth (number of channels) increases.
  - Skip connections are established at each level of the encoder to pass feature maps to the corresponding levels in the decoder.
- **Bridge:**
  - This central part of the network connects the encoder and decoder paths.
  - It represents the deepest layer of the network, where feature maps have the smallest spatial resolution and highest feature complexity, capturing high-level contextual information about the tumor.
- **Decoder (Upsampling Path):**
  - Consists of a series of blocks that include upsampling operations (e.g., transposed convolutions or bilinear upsampling followed by a convolution) and residual blocks.
  - The upsampling operations progressively increase the spatial resolution of the feature maps.
  - At each stage, the upsampled feature maps are combined (typically concatenated) with the corresponding feature maps received from the encoder via skip connections. This allows the network to integrate low-level detail with high-level semantic information.
  - The residual blocks in the decoder help refine the features and improve the reconstruction of fine details for accurate segmentation.
- **Output Layer:**
  - The final layer of the decoder maps the learned high-resolution features to a 1-channel output.

- This output is a binary segmentation mask, where pixel values of 1 indicate the predicted tumor region, and 0 represent the background.
- A sigmoid activation function is typically applied to the output layer to produce pixel-wise probabilities for the binary classification (tumor vs. background).

The model takes MRI slices as input.

1.3.3.2 Training and validation loss

Figure 4.5 displays the Deep ResUNet's learning curves over 50 epochs. Training and validation losses decreased substantially, indicating effective learning and convergence, despite some in-training fluctuations from which the model recovered. Key segmentation metrics, typically including IoU and Dice coefficient, also showed consistent improvement and stabilization at strong performance levels during training. The overall learning dynamics suggest good model generalization, though a slight divergence between training and validation losses in later epochs may hint at early signs of mild overfitting that could be monitored.

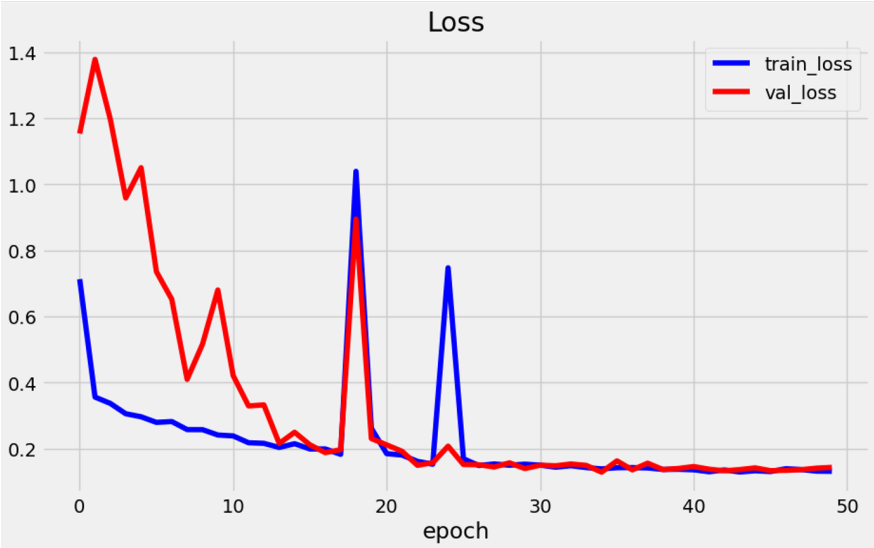

Figure 4.5: Deep ResUNet Training and Validation curves

1.3.3.3 Results

After the training phase, the Deep ResUNet model was evaluated on unseen MRI images. Figure 4.6 showcases that the model successfully locates the main tumor but smooths over its complex, infiltrative edges and incorrectly generates a small false positive region. It's a strong but imperfect result, highlighting the common trade-off between localization and boundary detail.

Figure 4.6: Sample Deep ResUNet Brain Tumor Segmentation Result

The final model was also evaluated quantitatively on the held-out test set using a suite of semantic segmentation metrics.

Metric	**Score**
Precision	0.883
Recall	0.763
F1 Score	0.819
Dice Score	0.819
IoU	0.693
Avg Prediction Time (s)	0.058

Table 4.2: Deep ResUNet valuation Metrics and Results

4.3.4 VGG19-UNet

1.3.4.1 Implementation

A U-Net-like segmentation model was constructed, leveraging a pre-trained VGG19 encoder combined with a custom-designed decoder for the task of brain tumor segmentation in MRI images. The architecture integrates the feature extraction power of VGG19 with the localization capabilities of the U-Net structure. The network is composed of the following key sections:

- **Encoder (Downsampling Path):**
  - Utilizes the convolutional layers from a VGG19 model, pre-trained on the ImageNet dataset, to serve as the feature extraction backbone.
  - Processes input MRI images (assumed to be 3-channel, compatible with VGG19) to capture hierarchical features.
  - Gradually reduces spatial dimensions while increasing feature depth.
- **Bridge:**
  - Forms the connection between the encoder and decoder paths.
  - Represents the deepest and most compressed feature map from the encoder.
- **Decoder (Upsampling Path):**
  - Employs a series of upsampling layers (e.g., transposed convolutions) and standard convolution layers to progressively increase spatial resolution.

- At each upsampling stage, it concatenates the upsampled features with the corresponding high-resolution features received from the encoder via skip connections.
- This fusion of features helps in recovering fine-grained details and precise localization of tumor boundaries.

- **Output Layer:**
  - The final layer consists of a convolution (typically 1x1) that maps the feature channels to a single channel.
  - Produces a binary segmentation mask, where pixel values of 1 indicate the predicted tumor region and 0 represent the background.

1.3.4.2 Training and Validation Loss

Monitoring training and validation losses throughout the epochs is crucial for assessing the learning progress and identifying potential issues like overfitting. Typically, the training loss should consistently decrease, while the validation loss should also decrease and then stabilize, indicating good generalization.

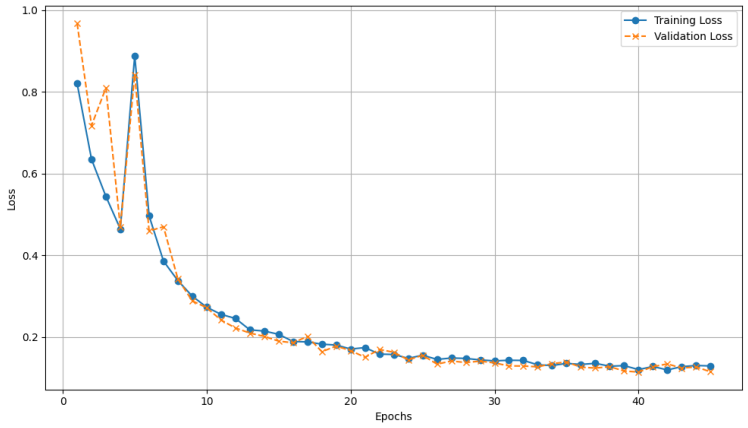

Figure 4.7: Training and Validation Loss for VGG19-UNet Across Epochs

1.3.4.3 Results

As shown in Figure 4.8,the model has done an excellent job of capturing the tumor.

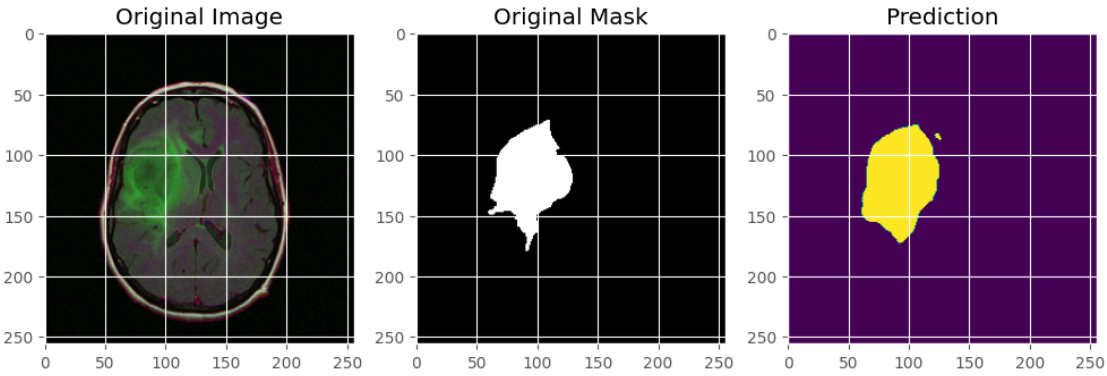

Figure 4.8: Sample VGG19 U-Net Brain Tumor Segmentation Result

The quantitative metrics indicate that the VGG19 U-Net model effectively segments brain tumors. The scores suggest a good ability to capture actual tumor areas while maintaining a reasonable level of precision

in its predictions. Overall, the results reflect a solid agreement between the model's segmentations and the ground truth, demonstrating competent delineation of tumor regions. The model's efficiency is also notable, with a quick average prediction time.

Metric	Score
Precision	0.712
Recall	0.858
F1 Score	0.778
Dice Score	0.778
IoU	0.637
Avg Prediction Time (s)	0.072

Table 4.3: VGG19 U-Net Evaluation Metrics and Results

4.3.5 SAM

1.3.5.1 Without data augmentation

- As shown in Figure 4.9, the dataset contains approximately 65% tumorous cases, providing imbalanced representation for evaluation

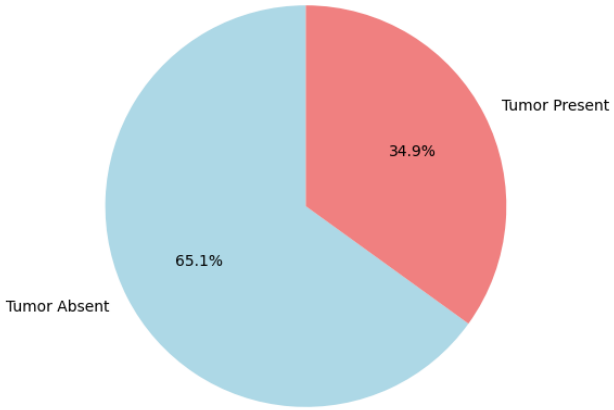

Figure 4.9: Distribution of images with and without tumor before data augmentation

1.3.4.1.1 Original SAM

Implementation:

- Image Encoder: A ViT that processes the input image to generate latent embeddings
- Prompt Encoder: Translates user inputs (points/boxes/text) into embedding vectors
- Mask Decoder: Predicts segmentation masks from combined image and prompt embeddings

Results:

Figure 4.10: Example predictions showing SAM's limited medical segmentation capability

- As shown in Figure 4.10, the predictions from the original SAM model fail to differentiate the tumor from healthy brain tissue, often producing masks that cover extensive or entire brain regions. This excessive over-segmentation indicates a lack of spatial and contextual understanding required for medical imaging tasks. The results visually align with the poor quantitative metrics, reinforcing the need for specialized fine-tuning or hybrid adaptations when applying general-purpose segmentation models like SAM to clinical use cases such as brain tumor localization.

1.3.4.1.2 Fine-tuned SAM

Implementation:

- ViT-B encoder frozen with original SAM weights
- Custom decoder trained on medical data (35 epochs)
- Combined loss (50% BCE + 50% Dice)
- No data augmentation applied

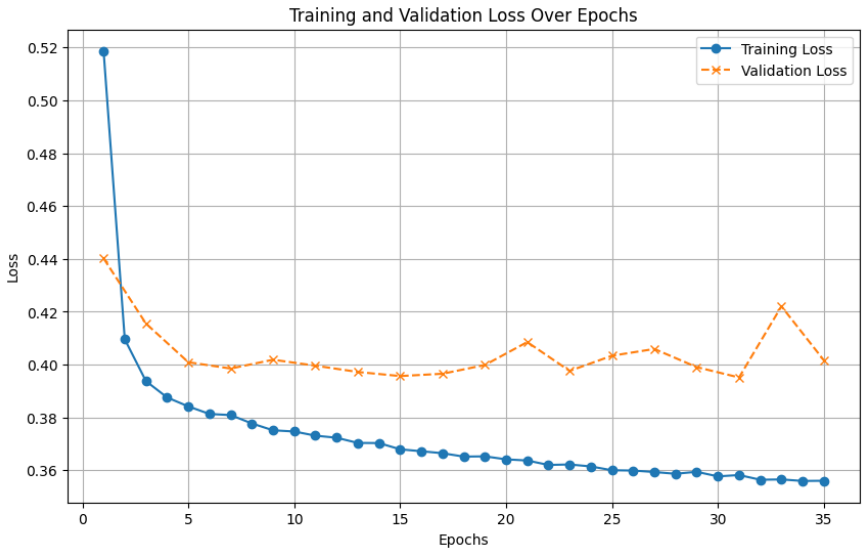

Figure 4.11: Training & Validation Loss of fine-tuned SAM (No Data Augmentation)

Results: The fine-tuned model shows improvements in detecting tumor regions, as seen in Figure 4.12where it captures the tumor area reasonably well.

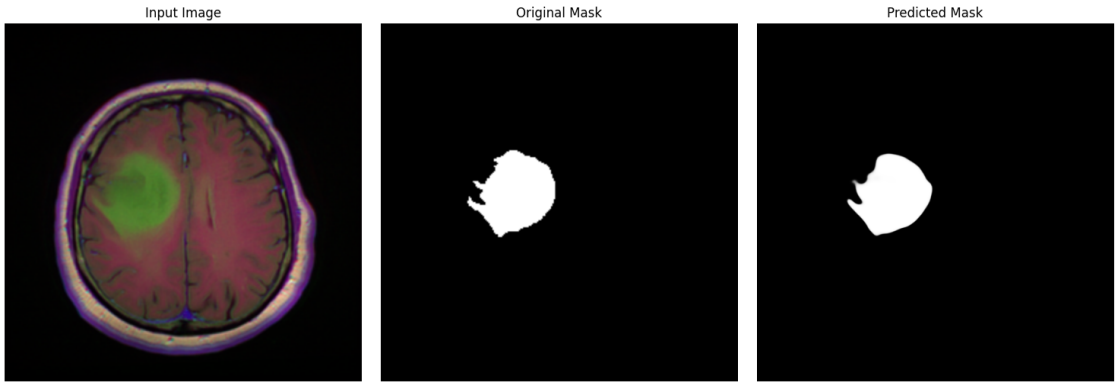

Figure 4.12: Example Predictions of Fine-tuned SAM for MRI Scans With Tumors (without data augmentation)

However, the model still makes mistakes. As illustrated in Figure4.13, when presented with an input image that has an empty original mask (indicating no tumor), the fine-tuned SAM incorrectly predicts a non-empty mask, indicating a false positive.

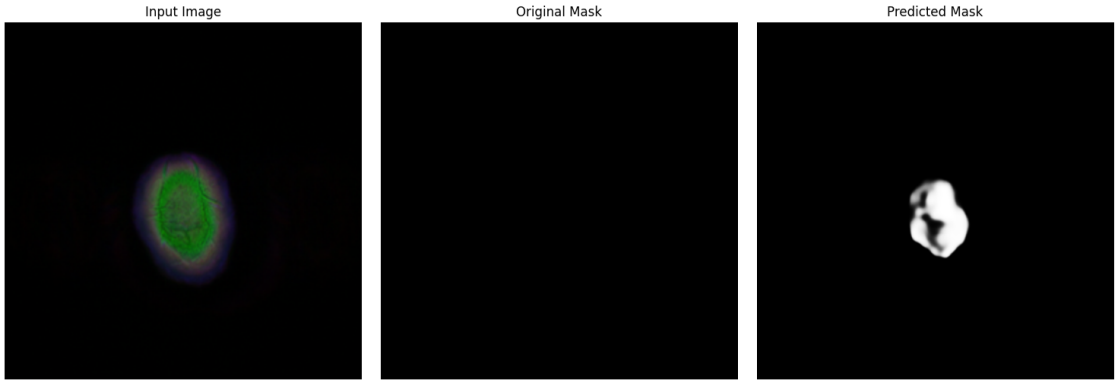

Figure 4.13: Example Predictions of Fine-tuned SAM for MRI Scans Without Tumors (without data augmentation)

Metric	**Original SAM**	**Fine-Tuned SAM**
Precision	0.009	0.757
Recall	0.326	0.688
F1 Score	0.018	0.721
Dice Score	0.018	0.721
IoU	0.009	0.564
Avg Prediction Time (s)	0.250	0.205

Table 4.4: Comparative Evaluation Metrics Original SAM vs. Fine-Tuned SAM (Without Data Augmentation)

1.3.5.2 With data augmentation

- As shown in Figure 4.14, the dataset contains approximately 51.8% tumorous cases, providing balanced representation for evaluation

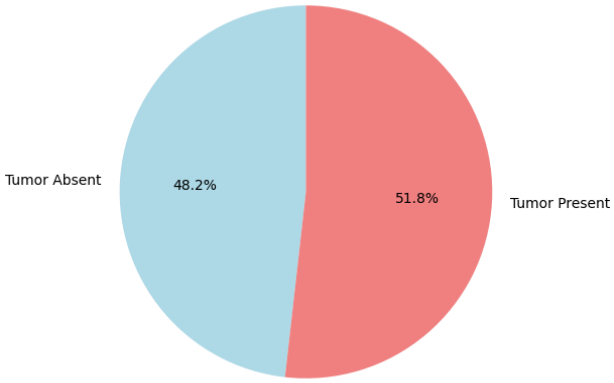

Figure 4.14: Distribution of images with and without tumor after data augmentation

1.3.5.2.1 Original SAM

Implementation:

- Applied spatial transformations (flips, rotations)
- Added intensity variations (brightness/contrast)
- Balanced tumor presence from 34.9% to 51.8%

Results:

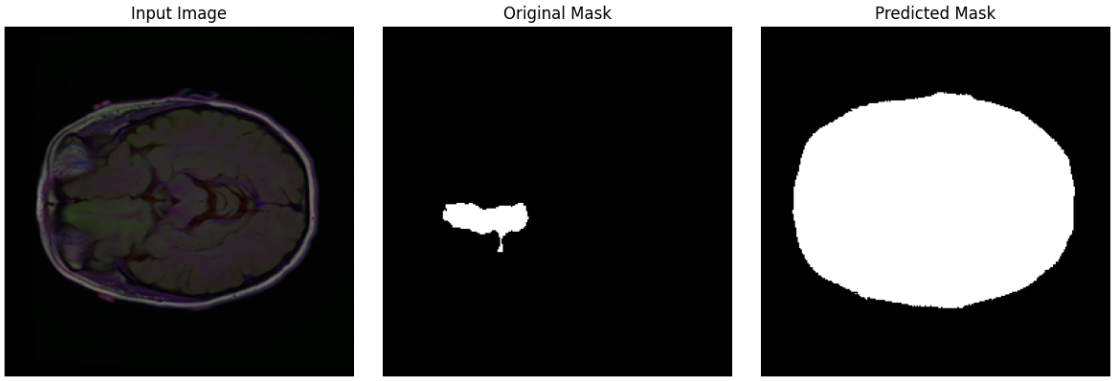

Figure 4.15: Example predictions showing augmented SAM's limited medical segmentation capability

T̈he data augmentation approach failed to improve segmentation accuracy, with the model still predicting the entire brain region as tumor tissue (gross over-segmentation). Most critically, no meaningful boundary detection was achieved, as evidenced by the predicted masks' inability to distinguish between healthy and pathological tissue.

1.3.5.2.2 Fine-tuned SAM

Implementation:

- Frozen encoder preserved pre-trained feature quality while enabling decoder specialization.
- Consistent segmentation quality across diverse MRI scans.
- The training loss decreases consistently over epochs, showing effective learning on the training data.

- Introducing variability in training data accelerates learning and enhances generalization, as evidenced by lower validation loss.

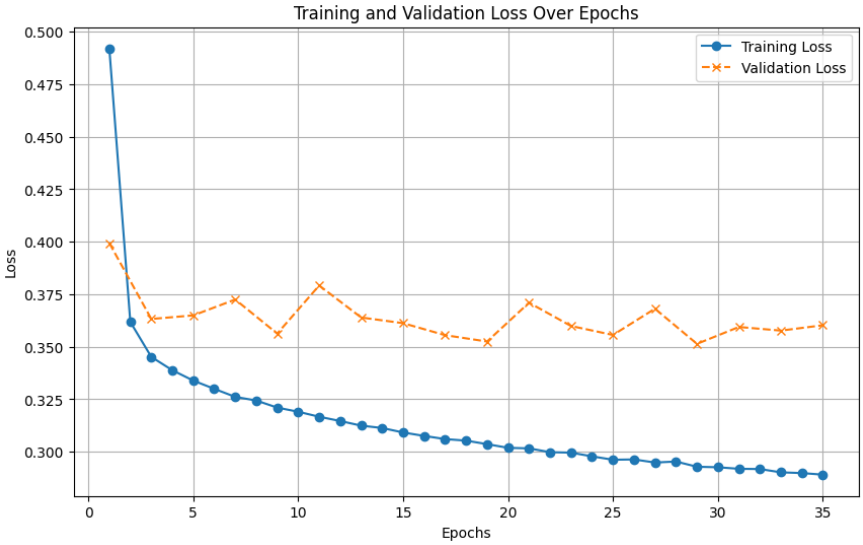

Figure 4.16: Training & Validation Loss of fine-tuned SAM (With Data Augmentation)

Results:

The fine-tuned SAM model with data augmentation demonstrates superior segmentation performance compared to the original model. Data augmentation enables better generalization to diverse cases, while fine-tuning optimizes the model for precise tumor delineation. Crucially, this enhanced robustness includes the ability to correctly predict empty masks for tumor-free images, avoiding the false positives seen in earlier iterations.

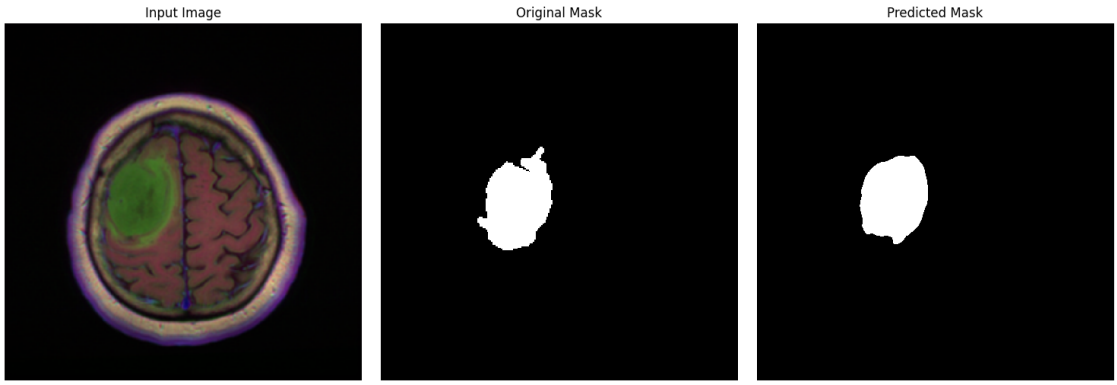

Figure 4.17: Example Predictions of Fine-tuned SAM in case of tumor (with data augmentation)

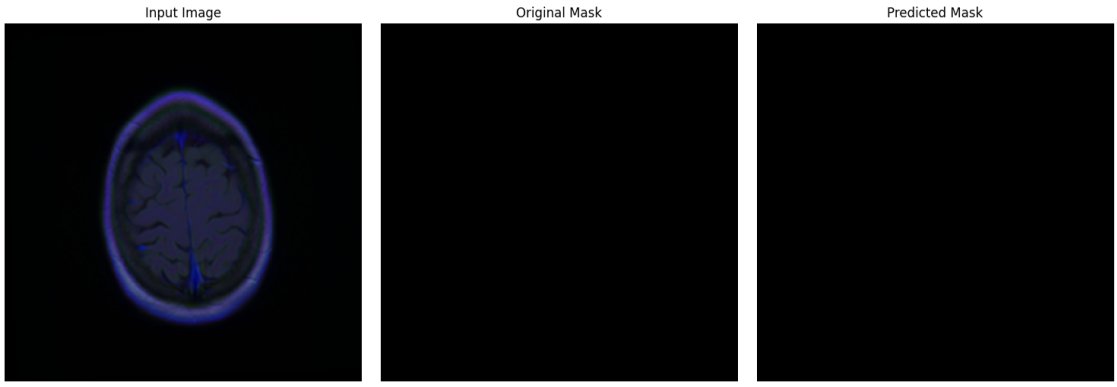

Figure 4.18: Example Predictions of Fine-tuned SAM in case of no tumor (with data augmentation)

Metric	Original SAM	Fine-Tuned SAM
Precision	0.012	0.703
Recall	0.302	0.749
F1 Score	0.023	0.725
Dice Score	0.023	0.725
IoU	0.012	0.569
Avg Prediction Time (s)	0.248	0.204

Table 4.5: Comparative Evaluation Metrics Original SAM vs. Fine-Tuned SAM (With Data Augmentation)

4.3.6 MedSAM

1.3.6.1 Overview and Setup

This approach utilizes the MedSAM, specifically employing a pre-trained model checkpoint (medsam_vit_b.pth), which is based on a Vision Transformer (ViT-B) architecture. The initial setup involves:

- Installation of the MedSAM library and all necessary software dependencies.
- Downloading the specified pre-trained MedSAM model weights.

Unlike training a model from scratch, this method leverages the knowledge already encapsulated within the pre-trained MedSAM to perform segmentation tasks.

1.3.6.2 Segmentation Process

The core segmentation process is orchestrated through a custom inference pipeline, including a dedicated function and a step-by-step application to test images:

- **Inference Function (medsam_inference(...)):**
  - A specialized function, medsam_inference(...), was defined to manage the interaction with the MedSAM model for a given region of interest.
  - Inputs: This function takes an image embedding (generated by MedSAM's image encoder) and a bounding box (defining the area to segment) as primary inputs.
  - Prediction: It prompts the MedSAM model to predict the segmentation mask specifically within the provided bounding box.
  - Post-processing: The predicted low-resolution mask from MedSAM is then upsampled to match the original image's dimensions.

- Output: It returns a binary mask where pixels valued 1 represent the predicted tumor (or object of interest) and 0 represent the background.
- Handling Empty Regions: A special case is implemented: if an empty bounding box is provided (indicating no tumor or object expected), the function returns an empty (all zeros) mask.

- **Application to Test Images:**
  - The inference process is applied to each image in the test set:
  - The test MRI image and its corresponding ground truth mask are read.
  - A bounding box prompt is determined:
    * If the ground truth mask contains a tumor, a bounding box around this annotated region is computed.
    * If the ground truth mask is empty (no tumor), an empty bounding box is used, signaling MedSAM (via the medsam_inference logic) to ideally return an empty prediction.
  - Image Preprocessing: The input MRI image is preprocessed to meet MedSAM's requirements, which includes resizing to a fixed resolution (e.g., 1024x1024 pixels) and normalization of pixel values.
  - Embedding Generation: The preprocessed image is passed through MedSAM's image encoder to obtain a high-level image embedding.
  - Mask Generation: The medsam_inference(...) function is then called with the image embedding and the determined bounding box to yield the final predicted segmentation mask.

1.3.6.3 Results

Following the setup and definition of the inference pipeline, the pre-trained MedSAM model's performance was qualitatively evaluated on new MRI images. The model predicts the tumor area based on the input image and the derived bounding box prompt. The visualizations demonstrate its performance in distinct scenarios:

- Figure 4.19 illustrates a case where the input MRI contains no tumor. Consistent with the empty ground truth mask and an appropriate empty bounding box prompt, MedSAM correctly produces an empty predicted mask.

Figure 4.19: MedSAM segmentation result for an MRI slice with no tumor

- Figure 4.20 shows an example where a tumor is present. MedSAM, guided by a bounding box derived from the ground truth, successfully segments the tumor region, with its predicted mask closely aligning with the ground truth annotation.

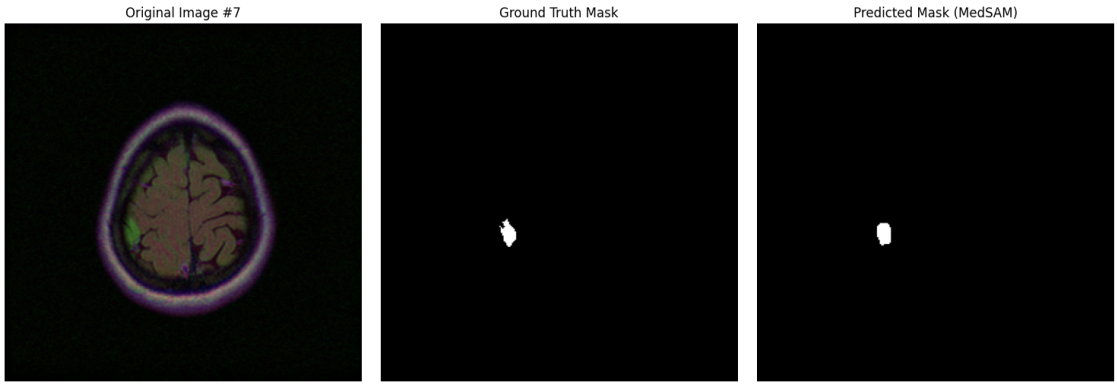

Figure 4.20: MedSAM segmentation result for an MRI slice with a tumor

The final model was also evaluated on the held-out test set using a suite of semantic segmentation metrics relevant to binary classification tasks at the pixel level:

Metric	**Score**
Precision	0.860
Recall	0.889
F1 Score	0.875
Dice Score	0.875
IoU	0.777
Avg Prediction Time (s)	0.235

Table 4.6: MedSAM Evaluation Metrics and Results

The quantitative metrics further validate the MedSAM model's effectiveness. The scores indicate a strong performance in accurately delineating tumor regions, characterized by a good balance between minimizing false detections and effectively identifying actual tumor areas. Overall, the metrics confirm a significant overlap between the predicted segmentations and the ground truth annotations when the MedSAM model is appropriately prompted.

4.4 Prognostic Analysis

The prognostic analysis pipeline integrates multi-modal data (genomic, clinical, and imaging features) to predict patient outcomes. It involves initializing models and data, extracting image-based features from MRIs, combining these with patient records for classification, and then generating risk scores and visualizations. The overall workflow of this prognostic analysis is depicted in Figure 4.21, while the Algorithm 1 presents our proposed prognostic analysis pipeline, step-by-step.

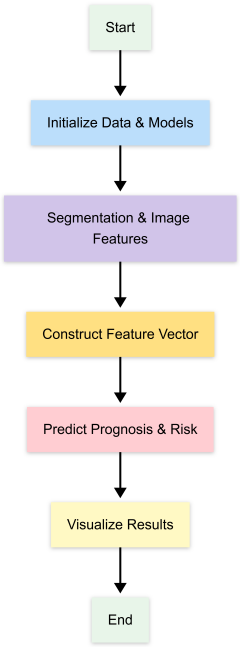

Figure 4.21: Overall flowchart of the prognostic analysis pipeline

4.4.1 Model and Dataset Initialization

A pretrained classifier trained on a combination of genomic, clinical, and anatomical features was loaded for outcome prediction.

A cleaned and preprocessed CSV dataset containing patient-level information was used as input, with missing values removed to maintain consistency with the training setup. Selected features included:

- **Multi-omics clusters:** RNASeqCluster, MethylationCluster, miRNACluster, CNCluster, RPPACluster, OncosignCluster, COCCluster
- **Clinical factors:** Age, Tumor grade
- **Anatomical indicators:** Tumor tissue site, Laterality, Tumor location

A separate segmentation model (e.g., U-Net, SAM or DeepRe-UNet) was loaded with pretrained weights to detect and segment tumor regions from MRI images.

4.4.2 Tumor Segmentation and Feature Extraction

The MRI image and corresponding ground truth mask were retrieved for each test sample. The segmentation model generated a predicted tumor mask, binarized using a fixed threshold. Morphological analysis was performed to extract key tumor features:

- **Tumor size:** Computed as the area of the largest connected region in the binary mask.
- **Tumor location:** Calculated from the centroid of the segmented region.

4.4.3 Feature Vector Construction

Extracted tumor features were injected into the patient record as follows:

- The tumor centroid's X-coordinate replaced the `tumor_location` feature.
- The tumor size replaced the `neoplasm_histologic_grade` feature to reflect actual tumor burden.

4.4.4 Prognostic Prediction and Thresholding

The final feature vector was passed through the classifier to produce a prognostic probability score p, indicating the likelihood of a poor outcome (e.g., mortality).

A dual-threshold strategy was applied for risk categorization:

- **Case 1:** $p < 0.35 \rightarrow$ Low-Risk, $p > 0.35 \rightarrow$ High-Risk
- **Case 2:** $p > 0.70 \rightarrow$ High-Risk, $p < 0.70 \rightarrow$ Low-Risk

Threshold Rationale: A lower threshold (0.35) emphasizes sensitivity, minimizing false negatives. The higher threshold (0.70) ensures specificity, prioritizing those at greatest risk.

Clinical Interpretation Example: A patient with $p < 0.35$ may be scheduled for routine monitoring, while one with $p > 0.70$ could be considered for early intervention or aggressive treatment.

4.4.5 Visualization

A visual comparison between the predicted and ground truth tumor masks was performed:

- Six MRI images were displayed with overlays.
- Red: Ground Truth Green: Predicted Mask
- Each image includes:
  - Predicted Risk (High/Low)
  - Probability Score
  - Actual Outcome (`death01`)

4.4.6 Feature Extraction from Segmentation

For each patient, tumor regions were segmented using the U-Net ,VGG19 U-Net, Deep ResUNet , SAM ,MedSAM model. Morphological features were then extracted from the predicted tumor masks generated by one of the models . These features included shape descriptors, area, perimeter, and other region-based properties. The extracted features were used as input for downstream classification to predict genomic subtypes and assess patient prognosis.

4.4.7 Classifier Models

The following machine learning models were employed to classify patients based on the extracted features:

SVM

Focusing on High Risk The sample results show the model consistently predicting "Low Risk" for all displayed patients. Notably, one actual high risk case was missed and classified as low risk, indicating a challenge in fully capturing all high risk individuals.

Algorithm 1 Risk Prediction Algorithm

Variables:
  Patient: string, *Image*: image
  GroundTruthMask, *PredictedMask*: image
  BinaryMask: image, *TumorSize*: float
  TumorCenter: tuple (X, Y), *RiskScore*: float
  RiskLevel: string, *PatientsDataset*: list
  FeatureVector: list

Start: Initialization
Load the tumor classification model ("tumor_classifier.xgb.pkl")
Load the segmentation model ("segmentation_model.pth")
Load and clean patient data from "patient_data.csv"
Define selected features:
  RNASeqCluster, MethylationCluster, miRNACluster, CNCluster,
  RPPACluster, OncosignCluster, COCCluster,
  age_at_initial_pathologic, neoplasm_histologic_grade,
  tumor_tissue_site, laterality, tumor_location
for each *patient* **in** *PatientsDataset* **do**

  Step 1: Tumor Segmentation
  Load MRI image for *Patient*
  PredictedMask ← Predict tumor mask using segmentation model
  BinaryMask ← Binarize the mask (*PredictedMask*, threshold = 0.5)

  Step 2: Feature Extraction
  TumorSize, *TumorCenter* ← Compute size and center from *BinaryMask*
  Update *Patient* record:
    Patient.tumor_location ← *TumorCenter.X*
    Patient.neoplasm_histologic_grade ← *TumorSize*
  FeatureVector ← ExtractFeatures(*Patient*, SelectedFeatures)

  Step 3: Risk Prediction
  RiskScore ← Predict(Classifier, *FeatureVector*)
  if *RiskScore* ¡ 0.35 **then**
    RiskLevel ← "Low Risk"
  else
    RiskLevel ← "High Risk"
  end if

  Step 4: Visualization
  RedEdges ← ExtractEdges(*GroundTruthMask*)
  GreenEdges ← ExtractEdges(*PredictedMask*)
  Image ← OverlayEdges(*Image*, *RedEdges*, color = "red")
  Image ← OverlayEdges(*Image*, *GreenEdges*, color = "green")
  TitleString ← "Risk: " + *RiskLevel* + " (Prob: " + Format(*RiskScore*) + ") - Actual: " + *Patient.GroundTruthLabel*
  ShowImage(*Image*, title = *TitleString*)

  Step 5: Clinical Decision (Simplified) *RiskLevel* is "Low Risk"
    Doctor action: Schedule Monitoring
    Doctor action: Prioritize Intervention
end for
End

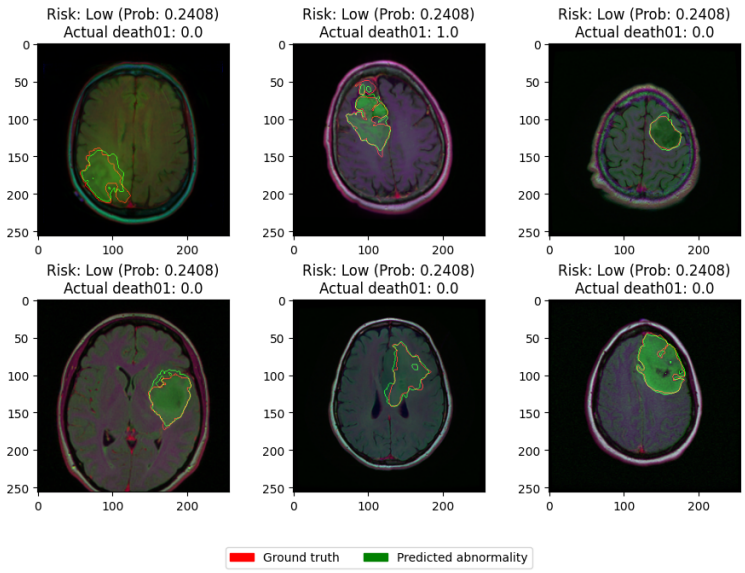

Figure 4.22: Results of the SVM classifier focusing on high risk

Focusing on Low Risk The sample results, when analyzed with a focus on low-risk prediction (using a higher probability threshold), show the model consistently classifying all displayed patients as "Low Risk" (with a relatively low probability). While this strategy aims to minimize false positives for high-risk, a critical observation is that a single actual high-risk case was still misidentified as low-risk. This highlights that, despite focusing on confident low-risk predictions, the model may still miss crucial high risk individuals, leading to potential false negatives.

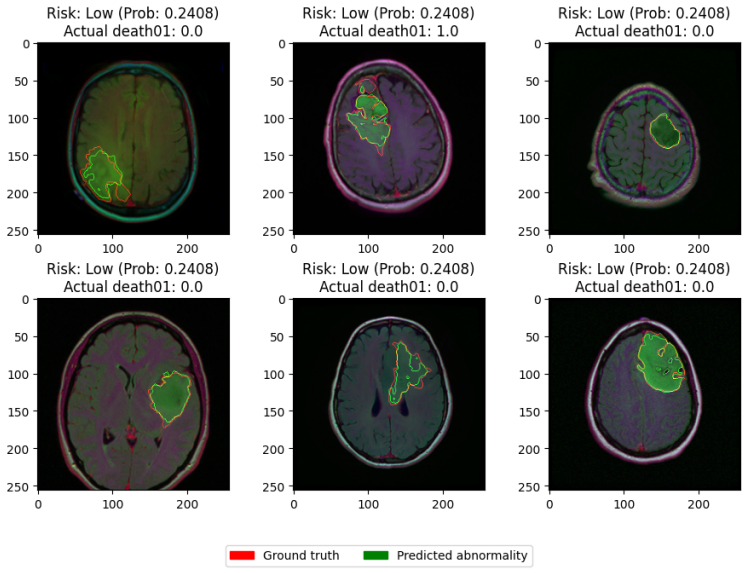

Figure 4.23: Results of the SVM classifier focusing on low risk

Random Forest

Focusing on High Risk The sample results show mixed predictions when focusing on high risk. It successfully identified the single actual high-risk case with a relatively high probability. This demonstrates a capacity to capture critical high-risk instances within this sample set.

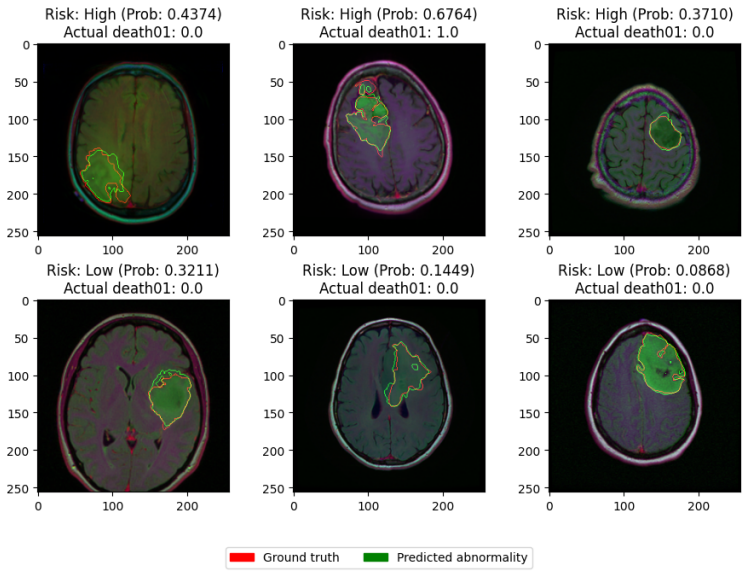

Figure 4.24: Results of the Random Forest classifier focusing on high risk

Focusing on Low Risk When focusing on low risk, the Random Forest sample results show accurate predictions. All instances predicted as low risk in this set were actual low-risk cases, and these identifications were made with appropriately low probabilities of being high risk. This suggests a strong ability of the model to correctly identify low risk instances within this particular sample.

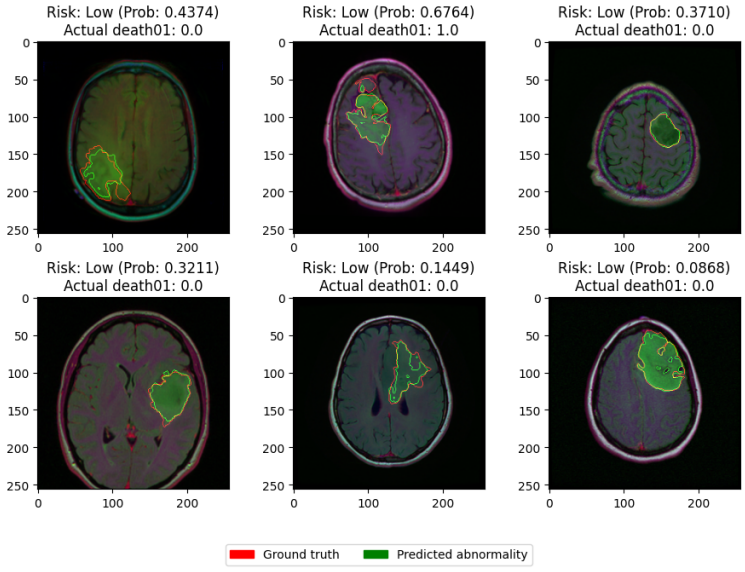

Figure 4.25: Results of the Random Forest classifier focusing on low risk

Logistic Regression

Focusing on High Risk The sample results indicate an aggressive tendency to predict high risk. While it successfully identified the one actual high-risk case with maximal reported confidence, it also incorrectly classified all other displayed low risk instances as high risk. This suggests that its approach in this sample, while capturing the critical case, leads to a very high rate of false positives.

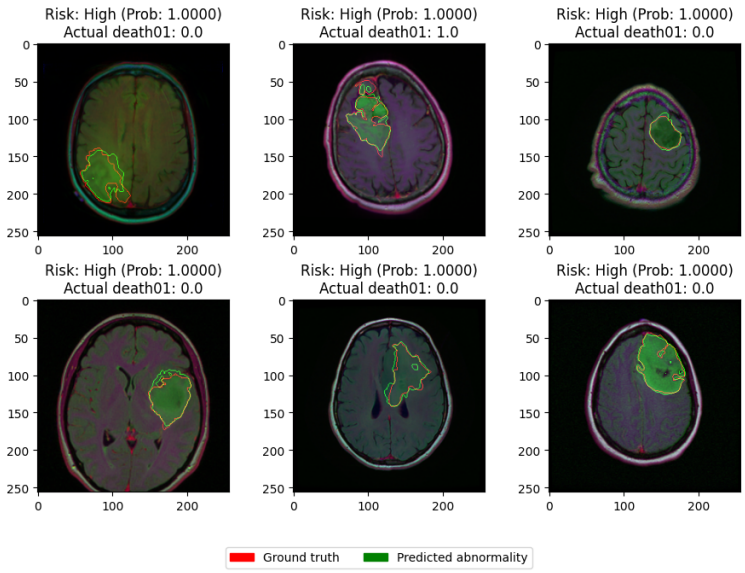

Figure 4.26: Results of the Logistic Regression classifier focusing on high risk

Focusing on Low Risk The provided sample shows no cases where the model predicted 'Low Risk.' All instances were instead classified as 'High Risk,' including those that were low risk. Consequently, the model did not label any actual low-risk cases as 'Low Risk' in this set, precluding a detailed assessment of its specific low-risk prediction performance from these examples.

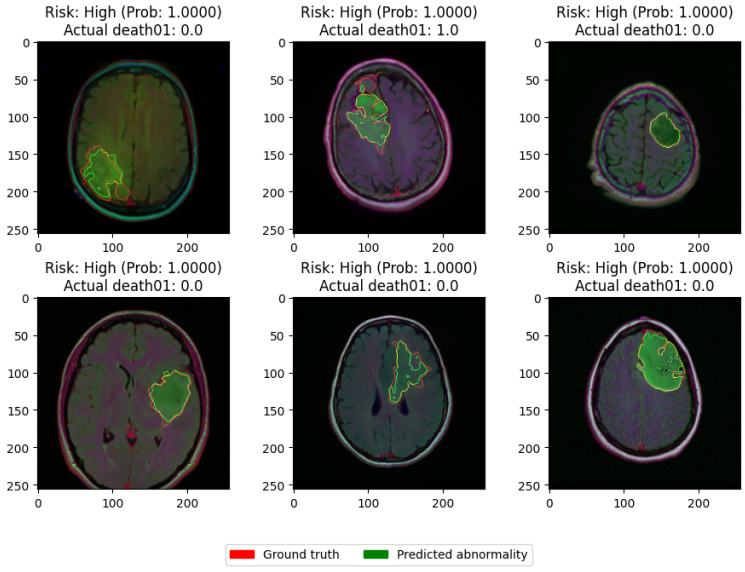

Figure 4.27: Results of the Logistic Regression classifier focusing on low risk

CatBoost

Focusing on High Risk The sample results show mixed predictions, successfully identifying the single actual high-risk case with a high probability. This indicates an improved ability to capture critical high risk instances in this sample.

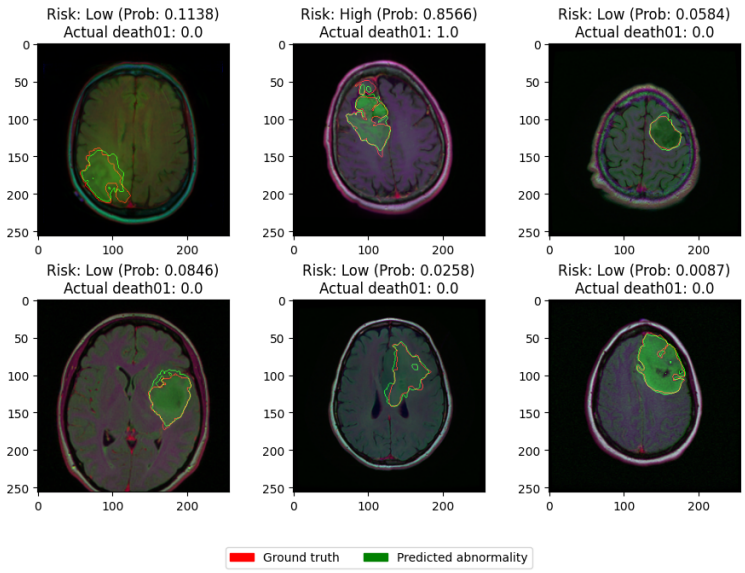

Figure 4.28: Results of the CatBoost classifier focusing on high risk

Focusing on Low Risk The sample results demonstrate strong performance in identifying true low risk cases, as all such instances were correctly classified. Importantly, the single actual high risk patient was also accurately predicted as "High Risk" with high confidence. This indicates the model effectively minimizes false positives for high risk predictions while maintaining accuracy for low-risk classifications.

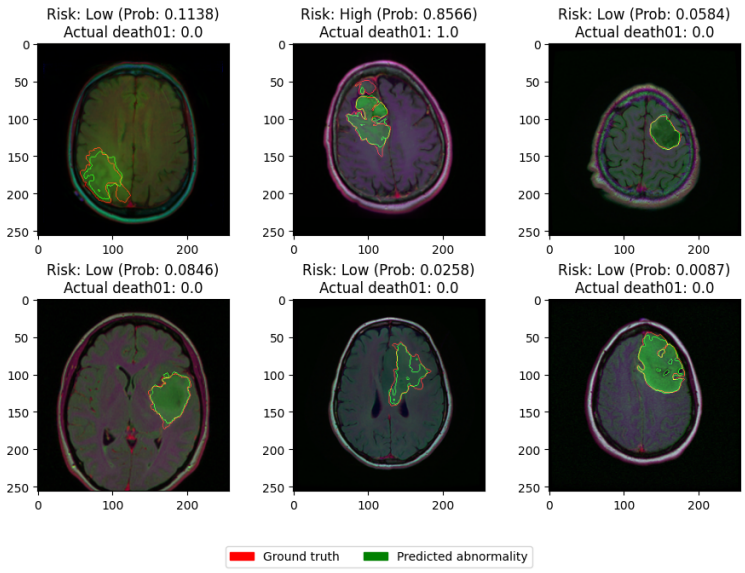

Figure 4.29: Results of the CatBoost classifier focusing on low risk

XGBoost

Focusing on High Risk For XGBoost with a high-risk focus, the sample results show the model successfully identifying the single actual high-risk patient. However, this strategy also led to several false positives, where low-risk patients were incorrectly flagged as high risk. This indicates a strong emphasis on recall for high risk cases, with a trade-off in precision, making it potentially suitable for initial screening where it's crucial not to miss critical cases.

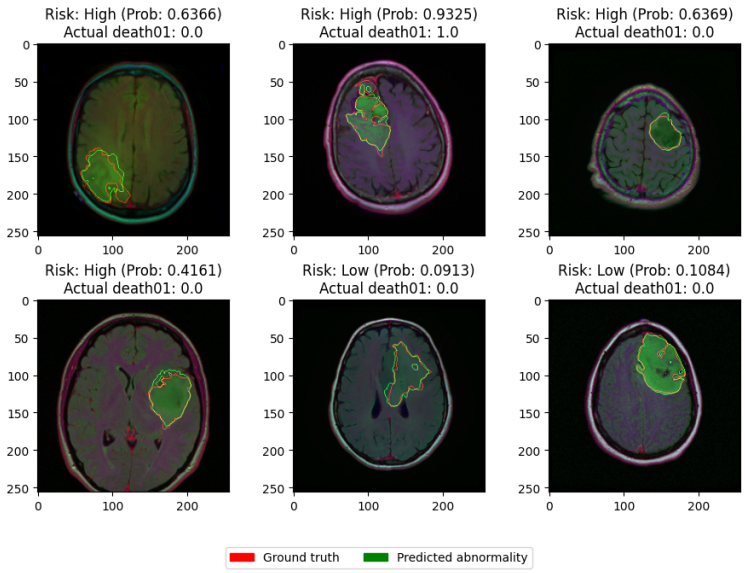

Figure 4.30: Results of the XGBoost classifier focusing on high risk

Focusing on Low Risk The sample results show strong performance in identifying true low risk cases. The model accurately classifies most low risk patients with high confidence (probabilities around 0.6). It also correctly identifies the single actual high risk patient with very high confidence. This indicates the model effectively minimizes false positives for high-risk predictions while maintaining accuracy for low-risk classifications.

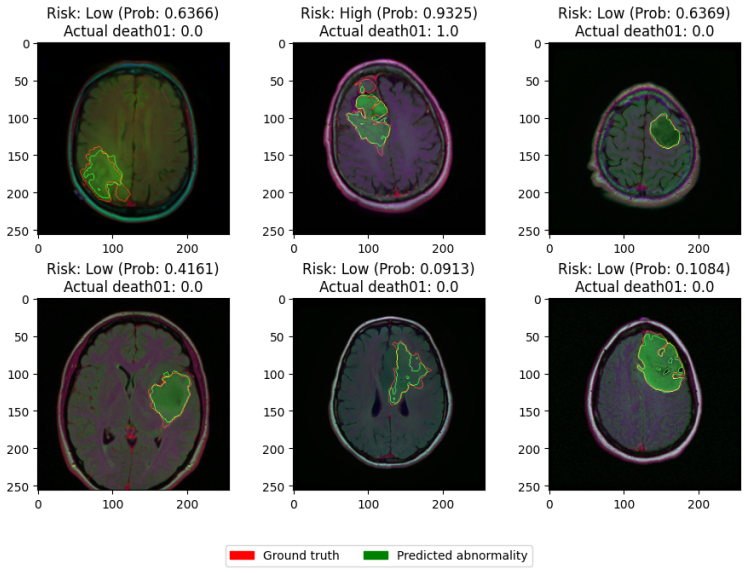

Figure 4.31: Results of the XGBoost classifier focusing on low risk

Majority Voting

Focusing on High Risk The Majority Voting sample results show mixed outcomes when predicting high risk. It successfully identified the single actual high risk case present in the sample, and this correct prediction was made with notably higher confidence compared to its incorrect high risk predictions. This suggests a capability to identify critical high-risk instances, though with some accompanying false positives in this set.

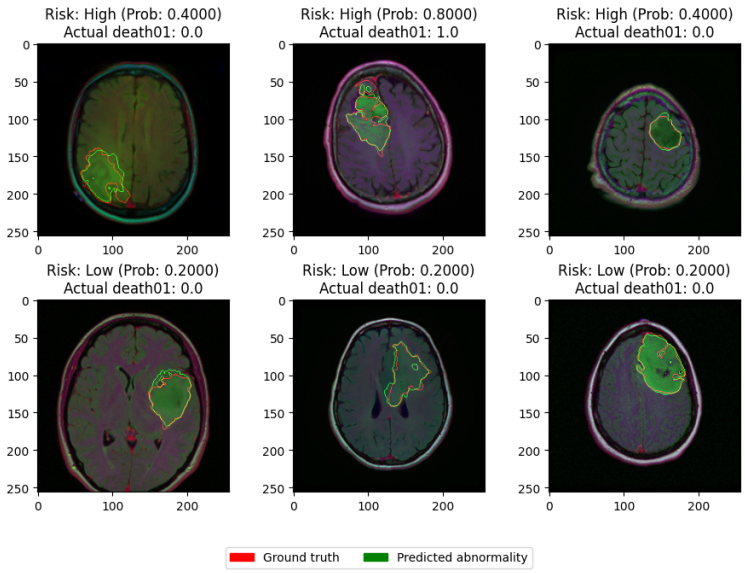

Figure 4.32: Results of the Majority voting classifier focusing on high risk

Focusing on Low Risk When focusing on low risk predictions, the Majority Voting sample results demonstrate accuracy. All instances that the model predicted as low risk were indeed actual low risk cases, and these were identified with a consistent and appropriately low assessed probability of being high risk. This indicates a reliable ability to correctly identify low risk instances within this sample set.

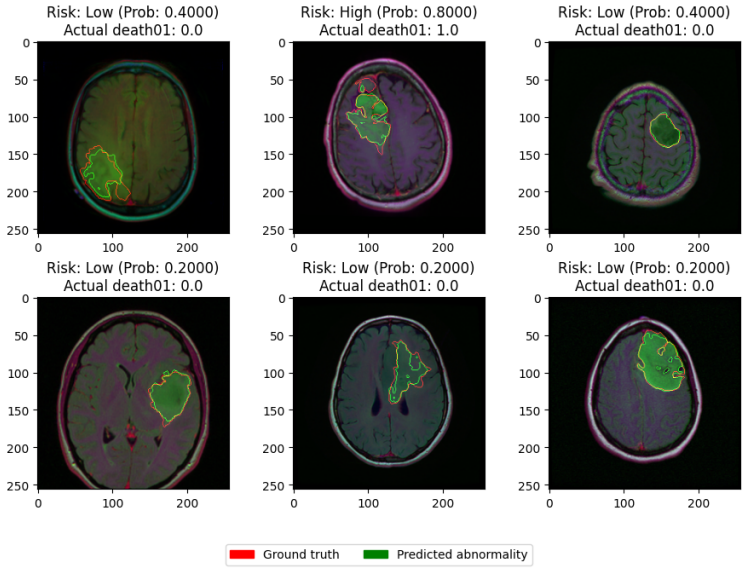

Figure 4.33: Results of the Majority voting classifier focusing on low risk

4.4.8 Confusion matrices

Confusion matrices were generated to analyze the pattern of misclassifications. All metrics were calculated on a held-out test set to ensure unbiased evaluation.

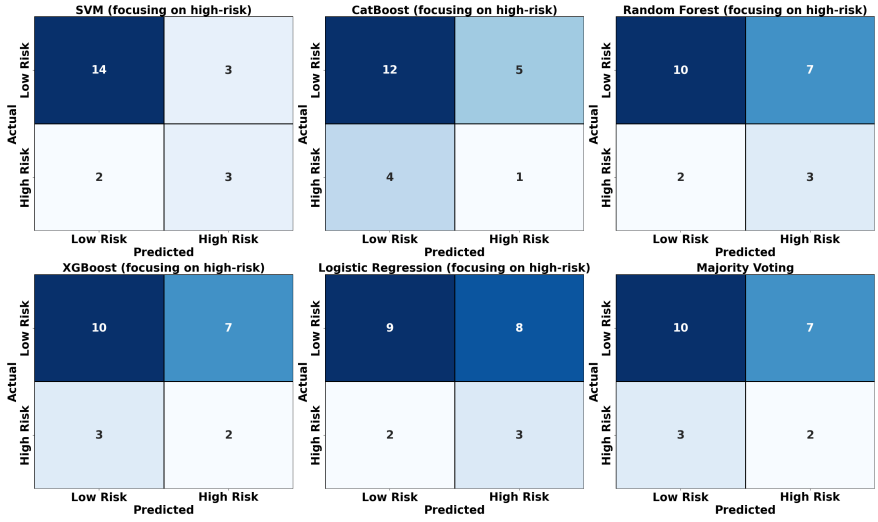

Figure 4.34: Confusion matrices for classifiers focusing on high-risk patient prediction

As shown in Figure 4.34, the SVM model stands out for its effectiveness. The matrix for SVM shows it successfully captures a notable number of actual high-risk cases while also committing relatively few errors by mislabeling low-risk individuals as high-risk. This balance appears more favorable in the SVM display compared to several other models presented; some identify a similar count of true high-risk patients but with a visibly greater spread of incorrect high-risk predictions among low-risk individuals, while other models seem to find fewer of the actual high-risk patients overall or show a generally higher tendency to misclassify.

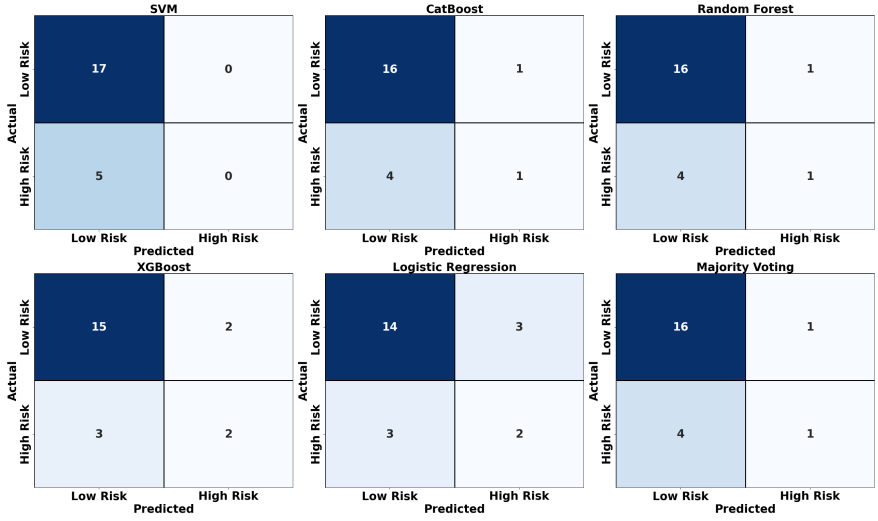

Figure 4.35: Confusion matrices for classifiers focusing on low-risk patient prediction

As shown in Figure 4.35,the SVM model again stood out. It uniquely managed to correctly classify all individuals who were actually low risk as "Low Risk," without incorrectly labeling any of these low-risk individuals as "High Risk." While other models also performed well, they typically had at least a few instances of either misclassifying low-risk individuals or not capturing all of them as accurately as the SVM did in this specific scenario.

4.5 Conclusion

This chapter has detailed the comprehensive methodological framework designed to investigate the use of deep learning for brain tumor analysis, moving from image segmentation to prognostic prediction.

Our approach began with the preparation of the public LGG Segmentation Dataset, which uniquely provided the necessary MRI scans, segmentation masks, and associated clinical and genomic data to support both phases of our research.

For the primary task of segmentation, a systematic and comparative evaluation was conducted. We implemented a diverse suite of architectures, from established CNN-based models like U-Net and Deep ResUNet to advanced foundation models like a fine-tuned SAM and the specialized MedSAM. This allowed for a rigorous assessment of different deep learning paradigms on this specific medical task.

Building on this, a novel prognostic analysis pipeline was established. This pipeline integrated the outputs from our segmentation models by extracting morphological tumor features and fusing them with the available patient data. A portfolio of machine learning classifiers, ranging from SVM to XGBoost, was then employed to stratify patient risk, with their performance analyzed in detail using confusion matrices.

The comprehensive and reliable approaches outlined in this chapter lay a strong groundwork for effectively presenting and critically examining the experimental results.

Chapter 5

General Conclusion

This thesis embarked on an investigation into the application of advanced DL methodologies for the critical tasks of brain tumor segmentation and prognostic analysis, with a specific focus on LGG. Motivated by the pressing need for more accurate, efficient, and objective tools to aid clinicians in diagnosing and managing this complex malignancy, the research systematically explored and validated a suite of contemporary DL architectures and ML classifiers.

The core of this work involved the design, implementation, and rigorous evaluation of diverse segmentation models, including the Recognized U-Net and its variants like Deep ResU-Net and VGG19-UNet, alongside advanced Transformer-based models such as SAM and its medical adaptation MedSAM. Through comprehensive experimentation on a public LGG MRI dataset, this research demonstrated the potential of these models to precisely delineate tumor boundaries, a crucial step for treatment planning and monitoring. The comparative analysis provided valuable insights into the relative strengths and weaknesses of each architecture. Notably, MedSAM achieved the highest Dice Score (87.46%) and IoU (77.71%), outperforming other models in brain tumor segmentation from MRI images and thereby emerging as a highly effective model for LGG segmentation, showcasing the benefits of domain-specific adaptation. The fine-tuning process for the SAM model proved crucial in adapting its general capabilities to the nuances of neuro-oncological imaging, highlighting the pathway towards leveraging foundational models for specialized medical tasks.

Beyond segmentation, this thesis extended its inquiry into the prognostic capabilities offered by DL. By extracting features from MRI scans, likely informed by the segmentation outputs or learned representations, a range of machine learning classifiers including SVM, Random Forest, Logistic Regression, CatBoost, XGBoost, and a Majority Voting ensemble were employed to predict aspects of disease progression. This exploration affirmed the potential of integrating imaging data with ML to furnish clinicians with prognostic indicators, thereby supporting more personalized and timely therapeutic interventions. Among the classifiers evaluated, SVM demonstrated notable efficacy, particularly in detecting high-risk patients with an F1-score of 69.70% and an accuracy of 77.27%, enabling early identification of severe cases. Furthermore, XGBoost performed best for identifying low-risk patients, achieving an F1-score of 65.08% and an accuracy of 77.27%, supporting reliable risk assessment and personalized treatment planning. The successful application of these techniques signifies a step towards realizing a more holistic computational approach to LGG management, moving beyond anatomical delineation to predictive analytics.

The findings of this research contribute significantly to the growing body of evidence supporting the transformative role of DL in neuro-oncology. By addressing the limitations of manual image interpretation – such as its time-consuming nature and inter-observer variability – the developed and evaluated models offer a pathway to more standardized, rapid, and potentially more accurate assessments. This work serves to connect computational advancements and their practical application in clinical settings, offering tools that can empower radiologists and oncologists in their decision-making processes.

However, the journey towards full clinical integration is ongoing. While promising results were achieved, this research acknowledges certain limitations inherent in such studies. The performance of DL models is often contingent on the size and diversity of training datasets; future work should aim to validate these models, including the strong performance observed with MedSAM (Dice: 87.46%, IoU: 77.71%) and the prognostic classifiers like SVM (High-Risk F1: 69.70%) and XGBoost (Low-Risk F1: 65.08%), on larger,

multi-institutional datasets to ensure broader generalizability. Furthermore, the "black-box" nature of some complex DL models continues to be a challenge; enhancing model explainability and interpretability will be crucial for building trust and facilitating clinical adoption. The prognostic analysis, while insightful, could be further enriched by incorporating longitudinal data and exploring a wider array of clinical and molecular markers.

In conclusion, this thesis has successfully demonstrated the significant utility of advanced DL techniques for both segmenting LGG from MRI scans and for providing valuable prognostic insights. Specifically, MedSAM excelled in segmentation with a Dice Score of 87.46% and IoU of 77.71%, while SVM proved most effective for high-risk patient detection (F1-score: 69.70%) and XGBoost for low-risk patient detection (F1-score: 65.08%) in prognostic tasks. The developed frameworks and comparative analyses offer a robust foundation for future innovations in the field. It is envisaged that the continued refinement and validation of such computational tools will play an increasingly pivotal role in advancing the precision and efficacy of brain cancer diagnosis and treatment, bringing us closer to a future where technology profoundly enhances patient care in neuro-oncology.

Bibliography

[1] Mudassar Ali, Tong Wu, Haoji Hu, Qiong Luo, Dong Xu, Weizeng Zheng, Neng Jin, Chen Yang, and Jincao Yao. A review of the segment anything model (sam) for medical image analysis: Accomplishments and perspectives. *Computerized Medical Imaging and Graphics*, 119:102473, 2025.

[2] Mariana Belgiu and Lucian Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114:24–31, 2016.

[3] Benlahmar. Architecture U-Net: Une explication détaillée, 2025.

[4] Jason Brownlee. *XGBoost With python: Gradient boosted trees with XGBoost and scikit-learn*. Machine Learning Mastery, 2016.

[5] Mateusz Buda, Ashirbani Saha, and Maciej A Mazurowski. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in biology and medicine*, 109:218–225, 2019.

[6] Mateusz Buda, Ashirbani Saha, and Maciej A. Mazurowski. Lgg mri segmentation dataset, 2019.

[7] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[8] Yunyi Chen, Lan Quan, Chao Long, Yuxuan Chen, Li Zu, and Chenxi Huang. Brain tumor segmentation based on the U-Net++ network with efficientnet encoder. *Technology and Health Care*, 32(1_suppl):183–195, 2024.

[9] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

[10] David R Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 20(2):215–232, 1958.

[11] D Richard Cutler, Thomas C Edwards Jr, Karen H Beard, Adele Cutler, Kyle T Hess, Jacob Gibson, and Joshua J Lawler. Random forests for classification in ecology. *Ecology*, 88(11):2783–2792, 2007.

[12] Dinthisrang Daimary, Mayur Bhargab Bora, Khwairakpam Amitab, and Debdatta Kandar. Brain tumor segmentation from mri images using hybrid convolutional neural networks. *Procedia Computer Science*, 167:2419–2428, 2020.

[13] Frederic G Dhermain, Peter Hau, Heinrich Lanfermann, Andreas H Jacobs, and Martin J van den Bent. Advanced mri and pet imaging for assessment of treatment response in patients with gliomas. *The Lancet Neurology*, 9(9):906–920, 2010.

[14] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.

[15] Anna Veronika Dorogush, Vasily Ershov, and Andrey Gulin. Catboost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*, 2018.

[16] Chandra Sekaran DS and J Christopher Clement. Enhancing brain tumor segmentation in mri images using the ic-net algorithm framework. *Scientific Reports*, 14(1):15660, 2024.

[17] Zhishuai Feng, Shanlin Liu, and Guizhi Feng. Multi-sections fusion method based on vggu-net for liver ct image segmentation. In *2024 IEEE 7th International Conference on Computer and Communication Engineering Technology (CCET)*, pages 51–55, 2024.

[18] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, and Jian Chen. Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377, 2018.

[19] Xiaowei Han, Jiwu Liu, and Jiayuan Zhao. U-ccnet: brain tumor mri image segmentation model with broader global context semantic information abstraction. In *2024 IEEE 7th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, volume 7, pages 1550–1554. IEEE, 2024.

[20] Trevor Hastie, Robert Tibshirani, Jerome Friedman, Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Random forests. *The elements of statistical learning: Data mining, inference, and prediction*, pages 587–604, 2009.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[23] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1, 1995.

[24] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*. John Wiley & Sons, 2013.

[25] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification, 2003.

[26] Yuhao Huang, Xin Yang, Lian Liu, Han Zhou, Ao Chang, Xinrui Zhou, Rusi Chen, Junxuan Yu, Jiongquan Chen, Chaoyu Chen, Sijing Liu, Haozhe Chi, Xindi Hu, Kejuan Yue, Lei Li, Vicente Grau, Deng-Ping Fan, Fajin Dong, and Dong Ni. Segment anything model for medical images? *Medical Image Analysis*, 92:103061, 2024.

[27] Fabian Isensee, Paul F Jäger, Peter M Full, Philipp Vollmuth, and Klaus H Maier-Hein. nnu-net for brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6*, pages 118–132. Springer, 2021.

[28] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation (2018). *arXiv preprint arXiv:1809.10486*, 1809.

[29] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas Lange, Dag Johansen, and Håvard D Johansen. Doubleu-net: A deep convolutional neural network for medical image segmentation. *arXiv preprint arXiv:2006.04868*, page 7, 2020.

[30] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.

[31] Ludmila I Kuncheva and Christopher J Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine learning*, 51:181–207, 2003.

[32] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[33] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[34] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[35] David N Louis, Arie Perry, Pieter Wesseling, Daniel J Brat, Ian A Cree, Dominique Figarella-Branger, Cynthia Hawkins, HK Ng, Stefan M Pfister, Guido Reifenberger, et al. The 2021 who classification of tumors of the central nervous system: a summary. *Neuro-oncology*, 23(8):1231–1251, 2021.

[36] Donghang Lyu, Ruochen Gao, and Marius Staring. Mcp-medsam: A powerful lightweight medical segment anything model trained with a single gpu in just one day. *arXiv preprint arXiv:2412.05888*, 2024.

[37] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.

[38] Dhiraj Maji, Prarthana Sigedar, and Munendra Singh. Attention res-unet with guided decoder for semantic segmentation of brain tumors. *Biomedical Signal Processing and Control*, 71:103077, 2022.

[39] Maciej A Mazurowski, Kal Clark, Nicholas M Czarnek, Parisa Shamsesfandabadi, Katherine B Peters, and Ashirbani Saha. Radiogenomics of lower-grade glioma: algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with the cancer genome atlas data. *Journal of neuro-oncology*, 133:27–35, 2017.

[40] Scott Menard. *Applied Logistic Regression Analysis*, volume 106. SAGE, 2002.

[41] Ali Nawaz, Usman Akram, Anum Abdul Salam, Amad Rizwan Ali, Attique Ur Rehman, and Jahan Zeb. Vgg-unet for brain tumor segmentation and ensemble model for survival prediction. In *2021 International Conference on Robotics and Automation in Industry (ICRAI)*, pages 1–6, 2021.

[42] Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *New England Journal of Medicine*, 372(26):2481–2498, 2015.

[43] Didrik Nielsen. Tree boosting with xgboost-why does xgboost win" every" machine learning competition? Master's thesis, NTNU, 2016.

[44] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE transactions on medical imaging*, 35(5):1240–1251, 2016.

[45] Robi Polikar. Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, 6(3):21–45, 2006.

[46] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[47] I Gede Adnyana Putra, Ni Luh Wiwik Sri Rahayu Ginantra, I Putu Dedy Sandana, I Gusti Agung Indrawan, and I Putu Agus Eka Darma Udayana. Semantic segmentation of balinese dance agem with vgg19-unet method. In *2024 Ninth International Conference on Informatics and Computing (ICIC)*, pages 1–6, 2024.

[48] Ramin Ranjbarzadeh, Abbas Bagherian Kasgari, Saeid Jafarzadeh Ghoushchi, Shokofeh Anari, Maryam Naseri, and Malika Bendechache. Brain tumor segmentation based on deep learning and an attention mechanism using mri multi-modalities brain images. *Scientific reports*, 11(1):10930, 2021.

[49] Lior Rokach. Ensemble methods for classifiers. In *Data Mining and Knowledge Discovery Handbook*, pages 959–979. Springer, 2010.

[50] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[51] Ahliddin Shomirov, Jing Zhang, and Mohammad Masum Billah. Brain tumor segmentation of hgg and lgg mri images using wfl-based 3d u-net. *Journal of Biomedical Science and Engineering*, 15(10):241–260, 2022.

[52] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[53] Roger Stupp, Warren P Mason, Martin J Van Den Bent, Michael Weller, Barbara Fisher, Martin JB Taphoorn, Karl Belanger, Alba A Brandes, Christine Marosi, Ulrich Bogdahn, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *New England journal of medicine*, 352(10):987–996, 2005.

[54] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017.

[55] David MJ Tax, Martijn Van Breukelen, Robert PW Duin, and Josef Kittler. Combining multiple classifiers by averaging or by multiplying? *Pattern recognition*, 33(9):1475–1485, 2000.

[56] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

[57] A.S. Vickram and C M Mathan Muthu. Segmentation of waterbodies from otsu's preprocessed satellite images using vgg-unet. In *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, pages 844–848, 2024.

[58] Bing Wan, Bingbing Hu, Ming Zhao, Kang Li, and Xu Ye. Deep learning-based magnetic resonance image segmentation technique for application to glioma. *Frontiers in Medicine*, 10:1172767, 2023.

[59] Haiying Wang and Fang Miao. Building extraction from remote sensing images using deep residual u-net. *European Journal of Remote Sensing*, 55(1):71–85, 2022.

[60] B POPE Whitney and Garth Brandal. Conventional and advanced magnetic resonance imaging in patients with high-grade glioma. *The quarterly journal of nuclear medicine and molecular imaging: official publication of the Italian Association of Nuclear Medicine (AIMN)[and] the International Association of Radiopharmacology (IAR),[and] Section of the Society of...*, 62(3):239, 2018.

[61] Qifan Yang, Huijuan Zhang, Jun Xia, and Xiaoliang Zhang. Evaluation of magnetic resonance image segmentation in brain low-grade gliomas using support vector machine and convolutional neural network. *Quantitative Imaging in Medicine and Surgery*, 11(1):300, 2021.

[62] Chunhui Zhang, Li Liu, Yawen Cui, Guanjie Huang, Weilin Lin, Yiqian Yang, and Yuehong Hu. A comprehensive survey on segment anything model for vision and beyond. *arXiv preprint arXiv:2305.08196*, 2023.

[63] Leying Zhang, Xiaokang Deng, and Yu Lu. Segment anything model (sam) for medical image segmentation: A preliminary review. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 4187–4194, 2023.

[64] Yichi Zhang, Zhenrong Shen, and Rushi Jiao. Segment anything model for medical image segmentation: Current applications and future directions. *Computers in Biology and Medicine*, page 108238, 2024.

[65] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.

[66] Équipe Blent. U-Net : le réseau de neurones populaire en Computer Vision, 2022.