Classification, Segmentation and Detection of Brain Abnormalities in MRI images

Rajnish Kumar

Master of Science in Artificial Intelligence

from the

University of Surrey

School of Computer Science and Electronic Engineering

Faculty of Engineering and Physical Sciences

University of Surrey

Guildford, Surrey, GU2 7XH, UK

September 2024

Supervised by: Dr, Lucia Florescu

DECLARATION OF ORIGINALITY

I confirm that the project dissertation I am submitting is entirely my own work and that any material used from other sources has been clearly identified and properly acknowledged and referenced. In submitting this final version of my report to the JISC anti-plagiarism software resource, I confirm that my work does not contravene the university regulations on plagiarism as described in the Student Handbook. In so doing I also acknowledge that I may be held to account for any particular instances of uncited work detected by the JISC anti-plagiarism software, or as may be found by the project examiner or project organiser. I also understand that if an allegation of plagiarism is upheld via an Academic Misconduct Hearing, then I may forfeit any credit for this module or a more severe penalty may be agreed.

Classification, Segmentation and Detection of Brain Abnormalities in MRI Images

Rajnish Kumar

Author Signature: Rajnish Kumar Date: 03/09/2024

Supervisor's name: Dr, Lucia Florescu

WORD COUNT

Number of Pages: 65

Number of Words: 19481

ABSTRACT

This dissertation explores the use of deep learning models to enhance the classification, segmentation, and detection of brain tumors via MRI analysis. The research is significant as it has the potential to improve diagnostic procedures in brain tumor management by providing automated, highly accurate tools that address critical clinical needs. By leveraging advanced neural network architectures, the study aims to boost the accuracy and reliability of tumor classification, segmentation, and detection.

Key contributions of this research include a comprehensive evaluation of pretrained models on specialized datasets, revealing high accuracy in differentiating between various tumor types such as gliomas, meningiomas, and pituitary tumors. The ResNet-152 model achieved the highest classification accuracy of 88.3% with an F1-score of 0.88. In segmentation tasks, MA-Net excelled with a Dice coefficient of 0.8361 and an IoU of 0.7183. For detection, RetinaNet led with a mean average precision (mAP) of 0.95 and an IoU of 0.85. These models mark significant progress in classifying, segmenting, and detecting tumors, offering notable improvements over previous methods in the field. The study also emphasizes the importance of data augmentation and careful dataset selection, which were critical in enhancing the models' performance and generalizability.

The outcomes of this research establish a strong foundation for future work focused on further optimizing these models. Additionally, the study underscores the importance of integrating these models into clinical workflows to enhance their practical utility. Future studies might investigate using larger and more diverse datasets to further enhance the models' generalizability. Moreover, the adoption of semi-supervised and self-supervised learning techniques offers promising avenues for enhancing model robustness, enabling more reliable performance across varying clinical scenarios.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Accurate diagnosis and treatment of brain tumors are vital components of medical imaging and healthcare. This dissertation examines how advanced deep learning techniques can enhance the processes of classifying, segmenting, and detecting brain tumors. These tasks are essential components of medical image analysis, utilizing MRI scans to identify, delineate, and assess brain tumors. The research involves developing and implementing neural network models to automate and enhance these processes, thereby supporting clinical decision-making and improving patient outcomes. This work primarily involves theoretical research and experimental programming, leveraging deep learning frameworks and medical imaging datasets to create robust, reliable diagnostic tools.

1.1 Background and Context

Brain tumors pose significant challenges in medical imaging due to their varied composition and sensitive location within the brain. Accurate classification, segmentation, and detection are essential for proper diagnosis, effective treatment planning, and outcome prediction. Imaging modalities like Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scans are commonly employed to visualize brain tumors. However, the diverse appearance, size, and location of tumors complicate image interpretation, making the process complex and time-consuming.

Brain tumors are classified into various types, each exhibiting unique imaging characteristics that influence the design and effectiveness of AI models. Key types include gliomas, meningiomas, and pituitary tumors. Gliomas, originating from glial cells, exhibit varied textures and irregular shapes, making them challenging for segmentation algorithms. Meningiomas, arising from the meninges, typically have well-defined boundaries, facilitating detection but still posing challenges due to their proximity to critical structures. Pituitary tumors develop in the pituitary gland and often display different imaging features that necessitate specialized approaches in classification and segmentation tasks.

Early detection, accurate classification, and precise segmentation of brain tumors are crucial. Early detection and accurate classification can significantly improve prognosis by allowing earlier intervention, potentially reducing tumor growth and spread. Precise segmentation is vital for planning treatments like surgery and radiation therapy, as it allows healthcare professionals to create precise treatment plans that focus on maximizing tumor removal while minimizing damage to the surrounding healthy tissue. Segmentation also allows for consistent and precise measurement of tumor size and volume over time, which is crucial for monitoring treatment response and making necessary adjustments to therapy.

Automated detection systems support radiologists by identifying suspicious areas, though issues like variability in tumor appearance and class imbalance pose challenges to creating universal detection methods. Deep learning, especially convolutional neural networks (CNNs), has advanced medical image analysis by automatically extracting intricate features from data, proving highly effective for classifying, segmenting, and detecting brain tumors.

This dissertation explores how deep learning can enhance brain tumor classification, segmentation, and detection with greater accuracy and efficiency. By leveraging the latest neural network architectures and comprehensive datasets, this research aims to develop robust models that improve the capabilities of automated diagnostic systems, thereby enhancing patient outcomes and supporting clinical decision-making.

1.2 Classical Approaches

Traditional machine learning methods, including k-nearest neighbors (k-NN), support vector machines (SVM), decision trees, and random forests, have been extensively used in brain tumor detection, relying on handcrafted features like texture, intensity, and shape descriptors. However, these approaches are limited by the need for substantial domain expertise to design features, often missing subtle patterns essential for accurate classification. Additionally, these methods face challenges in computational efficiency and scalability, particularly with large datasets. Deep learning offers a more effective approach by automatically learning intricate features from raw data, which enhances accuracy, efficiency, and scalability in brain tumor detection.

1.3 Transition to Deep Learning

The shift from traditional machine learning to deep learning has profoundly transformed medical imaging, especially in the areas of brain tumor detection, classification, and segmentation. While traditional methods depended on manually designed features and often faced challenges with accuracy, scalability, and generalization, deep learning, especially through convolutional neural networks (CNNs), has addressed these issues by automating feature extraction and greatly improving performance.

Convolutional Neural Networks (CNNs), with their multiple layers, are adept at identifying intricate patterns and structures in medical images, making them particularly effective for tumor detection and segmentation. Compared to traditional methods, deep learning models consistently achieve higher accuracy in various imaging tasks by utilizing extensive datasets and advanced computational resources to improve precision and recall. This enhanced accuracy is crucial for the prompt and accurate diagnosis of brain tumors. Furthermore, deep learning models show a strong ability to generalize across diverse datasets, overcoming the limitations faced by traditional methods due to differences in imaging techniques and patient demographics. Methods like transfer learning and data

augmentation have strengthened the robustness of deep learning models, enhancing their effectiveness in clinical environments.

Another significant advantage of deep learning is computational efficiency. Once trained, these models can process and analyze medical images much faster than traditional methods, which is vital in clinical settings where timely diagnosis is critical. This efficiency helps streamline workflows and reduces the burden on radiologists.

Recently, transformer-based models, initially created for natural language processing, have become influential tools in medical imaging. Unlike traditional CNNs, these models excel at handling long-range dependencies and capturing global context, frequently delivering greater accuracy and robustness. The attention mechanisms in transformers allow them to concentrate on critical areas of an image, improving the precision of tumor detection and segmentation.

Overall, the integration of deep learning, particularly CNNs and transformer-based models, has revolutionized brain tumor detection by overcoming the limitations of traditional methods and improving accuracy, efficiency, and generalizability in clinical settings.

1.4 Scope and Objectives

The main goal of this research is to assess deep learning models for precise classification, segmentation, and detection of brain tumors in medical imaging. This involves leveraging advanced neural network architectures to boost diagnostic accuracy and efficiency, thereby aiding clinical decision-making and enhancing patient outcomes.

1. Evaluate the performance of pre-trained deep learning models for accurate classification of different types of brain tumors, including gliomas, meningiomas, and pituitary tumors.
2. Apply automated segmentation techniques using pre-trained models to precisely delineate tumor boundaries in medical images.
3. Implement robust detection algorithms to identify the presence and location of brain tumors.
4. Assess the performance of these models using standard evaluation metrics to ensure accuracy and reliability.

1.5 Achievements

This research represents a major step forward in applying pre-trained deep learning models to brain tumor analysis. By utilizing advanced architectures like convolutional neural networks (CNNs) and transformers, the study effectively fine-tuned existing models to achieve high accuracy in classifying brain tumors, including gliomas, meningiomas, and pituitary tumors. Techniques such as data augmentation and transfer learning were instrumental in improving the generalizability of these models,

ensuring strong performance across different imaging modalities and datasets.

In segmentation tasks, models like U-Net, U-Net++, and DeepLabV3 were employed to accurately delineate tumor boundaries. These models excelled in capturing both local and global features, leading to high scores in Intersection over Union (IoU) and Dice coefficient, which are crucial metrics for precise surgical planning and treatment evaluation.

For tumor detection, the study fine-tuned models such as Faster R-CNN and RetinaNet to achieve high Intersection over Union (IoU) and Mean Average Precision (mAP), effectively identifying tumor regions. Enhancements like adaptive feature pyramids and context-aware mechanisms further improved detection accuracy, particularly for tumors of varying sizes and characteristics.

A unique aspect of this research is its cohesive evaluation framework, which integrates multiple deep learning architectures to independently address classification, segmentation, and detection within a unified study. This comprehensive approach distinguishes the study from others that typically focus on a single task or a combination of two. The findings not only demonstrate the potential of AI-driven diagnostic systems but also provide a strong foundation for future integration into clinical practice. The advanced interpretability tools developed further enhance the models' transparency, potentially increasing trust and facilitating the eventual adoption of these technologies in healthcare settings.

1.6 Overview of Dissertation

This dissertation is organized to systematically present the research on applying deep learning techniques to brain tumor classification, segmentation, and detection. Each chapter builds on the previous one, forming a cohesive narrative that logically guides the reader through the research process and its findings step by step.

Chapter 2: The Literature Review begins with an extensive survey of existing research in brain tumor analysis, with a specific focus on the latest advancements in deep learning. It examines the limitations of earlier deep learning methods while highlighting the considerable benefits of modern techniques. This chapter also provides an in-depth discussion of various deep learning model architectures, elucidating their principles and relevance to brain tumor analysis. By identifying knowledge gaps and assessing different deep learning strategies, this review lays the groundwork for the research that follows.

Chapter 3: Methodology outlines the research design and methods used to achieve the study's objectives. It describes the datasets used, including their sources and preprocessing steps, and details the deep learning approaches employed for classification, segmentation, and detection. This chapter also discusses the criteria for selecting models, hyperparameter tuning, and the experimental setup, providing a clear and structured framework for the research.

Chapter 4: The Results section outlines the experimental outcomes, emphasizing detailed

performance metrics for each model, including accuracy, precision, recall, F1-score, Intersection over Union (IoU), and Dice coefficient. This chapter explores the influence of data augmentation on model performance and compares the proposed methods' effectiveness with existing techniques. It offers a comprehensive analysis of the results across classification, segmentation, and detection, providing both qualitative and quantitative insights into the models' performance.

Chapter 5: Discussion and Conclusion interprets the findings in the context of the research questions and objectives outlined in the earlier chapters. It explores the implications for brain tumor classification, segmentation, detection and addresses the limitations and challenges encountered during the research. Additionally, this chapter offers recommendations for future research and potential improvements, emphasizing the practical applications and technological implications of the study.

2 BACKGROUND THEORY AND LITERATURE REVIEW

This literature review offers a comprehensive analysis of research in brain tumor classification, segmentation, and detection, with a particular focus on recent developments in deep learning. It critically evaluates current methodologies, highlighting their strengths, limitations, and identifying gaps in the existing research landscape. The review begins by examining deep learning-based classification techniques, followed by segmentation and detection methods, discussing their underlying principles, advantages, and limitations.

Subsequently, the review delves into various deep learning architectures utilized for these tasks, emphasizing their unique features, operational mechanisms, and the trade-offs between their advantages and disadvantages. This comparative analysis sheds light on the specific strengths of these models while identifying areas ripe for improvement. The review also covers major datasets used in brain tumor research, outlining their characteristics and significance in model development. Additionally, it explores emerging trends and innovative technologies shaping the future of brain tumor analysis. The review concludes by evaluating how these advancements address current challenges and their potential impact on future research directions.

2.1 Classification, Segmentation and Detection Methods

2.1.1 Classification

In medical imaging for brain tumors, classification involves sorting images into specific categories, such as different tumor types. Accurate classification is crucial for selecting the most effective treatment plan, which significantly enhances patient outcomes. The advent of deep learning has revolutionized this process, particularly through convolutional neural networks (CNNs), a type of deep learning model that excels at automatically extracting hierarchical features from raw image data. This advancement has greatly improved the precision and reliability of tumor classification, leading to substantial gains in both accuracy and robustness.

In the initial phase of applying deep learning to brain tumor classification, convolutional neural networks (CNNs) were predominantly used as the primary models. Sajjad et al. [1] were pioneers in this domain, utilizing CNNs to classify gliomas, a common type of brain tumor. Their study demonstrated that CNNs could automatically learn hierarchical features from raw MRI images, achieving a classification accuracy of 91.7%. However, they faced significant challenges, particularly overfitting due to the limited size of the dataset. This issue highlighted the need for larger, more diverse datasets to ensure effective generalization in clinical settings. Additionally, the CNN model struggled with classifying tumors with atypical or highly variable morphologies, limiting its utility in real-

world scenarios.

Ghosal and N. [2] contributed significantly to the early exploration of CNNs in brain tumor classification by focusing on improving the accuracy of tumor detection through deeper CNN architectures. While these networks had the potential to learn more complex features, they also faced significant challenges, including overfitting due to the limited size of available datasets and difficulties in handling the variability in tumor appearances, similar to issues identified by Sajjad et al. [1] To address these challenges, Ghosal and N. implemented the ResNet-101 architecture, which introduced residual learning, allowing for the training of much deeper networks without common deep learning issues like the vanishing gradient problem. Their study demonstrated that ResNet-101 significantly outperformed traditional CNNs, achieving a 15% improvement in classification accuracy, particularly in complex cases where tumor structure was highly variable. However, the study also highlighted ongoing challenges, such as the need for larger annotated datasets to prevent overfitting and ensure generalizability. Additionally, the substantial computational resources required to train such deep networks pose a barrier to their broader adoption in clinical practice.

As deep learning in brain tumor classification advanced, researchers began investigating new architectures to enhance model robustness and integrate classification with other tasks like segmentation. A key development in this area was the Vision Transformer (ViT) architecture, as discussed by Şahin and Ö. [3]. Unlike traditional CNNs, the Vision Transformer employs transformers, originally developed for natural language processing, to process image data. Transformers are particularly adept at capturing long-range relationships, making them highly effective for tasks that require a thorough understanding of the entire image context. The Vision Transformer leverages this ability in medical imaging by treating images as sequences of patches, enabling it to capture global context more effectively than CNNs, which are typically restricted to local receptive fields. This ability enabled the Vision Transformer to surpass CNN-based models in several brain tumor classification tasks, particularly when tumors exhibited subtle and diffuse characteristics that CNNs struggled to identify. The study reported that the Vision Transformer achieved a roughly 20% improvement in classification accuracy compared to traditional CNN models. Additionally, the ViT model demonstrated greater robustness across different datasets, making it a more reliable tool for clinical applications.

Alongside these advancements, researchers such as Agarwal and G. [4] and Shiny [5] focused on integrating classification with segmentation tasks. Segmentation, which involves delineating the tumor boundaries within an image, is a critical step in the diagnostic process. By combining segmentation and classification into a single model, these researchers aimed to create more comprehensive diagnostic tools that could provide both the tumor type classification and a detailed map of its location and extent. Agarwal and G. [4] utilized optimized U-Net architectures to achieve this integration. Originally designed for biomedical image segmentation, U-Net was adapted to perform

both tasks simultaneously. Their study demonstrated that the optimized U-Net significantly improved diagnostic accuracy, with results showing an accuracy increase of around 18% when segmentation was integrated with classification. This improvement was due to the model's ability to leverage segmentation information to refine the classification process, providing clinicians with more detailed and accurate insights. Shiny [5] further validated the effectiveness of these integrated models in clinical settings. Their research showed that combining segmentation with classification not only improved the accuracy of tumor classification but also provided valuable contextual information essential for treatment planning. The optimized U-Net models developed by Shiny [5] achieved an accuracy improvement of approximately 22% over traditional classification-only models. These results highlighted the importance of integrated models in providing comprehensive diagnostic tools that enhance clinical decision-making.

The evolution of deep learning in brain tumor classification has seen significant advancements, from the initial application of CNNs to the development of deeper architectures like ResNet-101 and novel models such as the Vision Transformer. These advancements have led to substantial improvements in accuracy and robustness, particularly when integrated with segmentation tasks in models like the optimized U-Net. The studies reviewed showed classification accuracy improvements ranging from 15% to 22% depending on the specific model and task integration, highlighting the significant impact of these advancements on the field.

2.1.2 Segmentation

Segmentation is a crucial aspect of medical imaging, especially for brain tumors, where accurately defining tumor boundaries is vital for guiding treatment decisions, surgical planning, and ongoing disease monitoring. Early applications of deep learning to segmentation tasks primarily utilized convolutional neural networks (CNNs) and U-Net architectures, both of which have become foundational models for brain tumor segmentation, significantly enhancing accuracy and consistency.

The U-Net architecture, as applied by Akter and N. [6], quickly emerged as a leading model for brain tumor segmentation due to its ability to achieve high accuracy even with limited data. U-Net’s symmetric encoder-decoder structure, featuring skip connections that preserve vital spatial information, enabled precise localization of tumors in MRI images. Akter and N. focused on enhancing U-Net's performance by optimizing its hyperparameters and training processes, allowing the model to achieve segmentation accuracy exceeding 85%. Their study demonstrated that U-Net could effectively handle the complex, irregular shapes typical of brain tumors, establishing it as a robust tool in clinical practice. This success positioned U-Net as a foundational model for further advancements in segmentation tasks, particularly in scenarios requiring detailed boundary delineation.

Building on the foundational work of U-Net, Dong et al. [7] employed a fully convolutional

network (FCN) approach, closely related to U-Net, specifically designed for brain tumor segmentation. Their study demonstrated the effectiveness of CNN-based architectures in learning and utilizing spatial hierarchies, which are essential for accurate segmentation. Dong et al. enhanced the traditional U-Net model by incorporating deeper convolutional layers, significantly improving the model's capability to capture and segment tumors with highly irregular shapes and varying sizes. Their work achieved a Dice similarity coefficient (DSC) of approximately 0.87, indicating a strong correlation between the predicted segmentations and the actual ground truth. However, despite these advancements, Dong et al. recognized ongoing challenges, particularly the need for further refinement to enhance segmentation accuracy for tumors with non-uniform shapes and complex surrounding structures.

As the field of medical imaging evolved, researchers recognized the need for more sophisticated segmentation techniques capable of addressing the limitations of earlier models. Zhou and Rahman Siddiquee [8] responded to this challenge by developing UNet++, a significant evolution of the original U-Net architecture. UNet++ introduced nested and dense skip connections, enabling more detailed feature extraction and improving the model's capability to manage fine-grained segmentation tasks. These nested connections allowed the model to iteratively refine segmentation outputs at multiple scales, leading to more accurate delineation of tumor boundaries, particularly in cases where traditional models struggled. Zhou and Rahman Siddiquee's study reported a Dice Similarity Coefficient (DSC) of approximately 0.89, marking a notable improvement over the original U-Net and highlighting the value of these architectural innovations. Additionally, their work emphasized the model's robustness in handling cases with ambiguous or poorly defined tumor boundaries, a common challenge in brain tumor imaging.

Further advancing the field, Liu et al. [9] developed CANet, a Context-Aware Network that integrated contextual awareness into the segmentation process. CANet marked a significant advancement by incorporating mechanisms that enabled the model to consider the broader anatomical context surrounding the tumor. This approach was particularly effective in segmenting tumors located near or within complex anatomical structures, where traditional models often struggled. By utilizing contextual information, CANet attained a Dice Similarity Coefficient (DSC) of 0.90, indicating enhanced segmentation accuracy and a reduction in false positives. Liu et al. emphasized CANet's potential for integration into clinical workflows, offering a more precise and comprehensive approach to tumor segmentation.

Chen et al. [10] enhanced segmentation capabilities by introducing TransUNet, a hybrid model that combines the strengths of both transformers and CNNs. TransUNet was specifically developed to capture both global and local contexts in medical images, making it particularly effective for complex segmentation tasks. The transformer component allowed the model to process the entire image as a series of patches, effectively capturing long-range dependencies and global context, while the

CNN component focused on extracting detailed local features. This dual strategy led to notable improvements in segmentation accuracy, especially in cases where tumors had intricate shapes or were located in challenging regions of the brain. Chen et al. reported that TransUNet achieved a Dice Similarity Coefficient (DSC) of approximately 0.91, outperforming traditional U-Net and other CNN-based models. Its ability to integrate global context with precise local feature extraction was crucial in managing the variability and complexity of brain tumors, making TransUNet an exceptionally effective tool for clinical applications, especially in real-time scenarios where accuracy and computational efficiency are critical.

While these innovations have significantly advanced the field, further improvements have been explored through ensemble models and region-adaptive mechanisms, which aim to refine segmentation accuracy and address the inherent complexities of brain tumor imaging. Isensee et al. [11] advanced the use of ensemble methods during their participation in the BRATS 2017 challenge. By combining multiple models into an ensemble, they effectively captured the diverse characteristics of brain tumors, resulting in enhanced segmentation performance. The ensemble model outperformed individual models by leveraging their collective strengths, leading to a more robust and reliable segmentation process. This highlighted the potential of ensemble methods to address the limitations of single-model approaches, especially in handling complex and variable data such as brain tumor imaging.

Chen et al. [12] made significant strides in segmentation techniques by developing the Deformable U-Net, which incorporated region-adaptive mechanisms to tackle the challenges presented by the varied shapes and sizes of brain tumors. Unlike traditional models that apply uniform operations across the entire image, the Deformable U-Net adjusts its processing to accommodate local variations, thereby capturing the unique characteristics of each tumor region more effectively. This adaptability proved particularly beneficial for tumors with significant morphological variability, such as those with irregular or diffuse boundaries. Chen et al. demonstrated that the Deformable U-Net outperformed conventional models in segmentation accuracy, particularly in difficult cases that required a more customized approach. Their research emphasized the model's potential to improve clinical outcomes by delivering more precise and reliable segmentation, which is essential for accurate treatment planning.

Valanarasu et al. [13] advanced segmentation technology further with the introduction of the Medical Transformer, which incorporated gated axial attention mechanisms. This model represented a significant shift from traditional CNN-based approaches by leveraging the transformer architecture, known for effectively capturing long-range dependencies within an image. The gated axial attention mechanism enabled the model to focus on the most critical features within the image, thereby enhancing segmentation accuracy. Axial attention is an attention mechanism that processes information along specific axes, such as height or width, rather than across the entire image simultaneously.

Valanarasu et al. reported that the Medical Transformer not only improved segmentation accuracy but also demonstrated superior computational efficiency, outperforming many traditional models. This advancement highlighted the potential of transformers in medical imaging, offering a promising direction for developing more effective and efficient segmentation models, particularly in clinical settings where timely and accurate decision-making is critical.

Recent advancements in segmentation models have notably improved brain tumor segmentation's accuracy and efficiency. Innovations like UNet++, CANet, and TransUNet, along with ensemble methods and region-adaptive mechanisms, have enhanced the models' ability to handle complex brain imaging. Higher Dice similarity coefficients reflect their effectiveness. As these models continue to evolve and integrate into clinical practice, they promise to further enhance diagnostic accuracy and patient care, especially with the growing computational capabilities enabling real-time applications.

2.1.3 Detection

Detection in medical imaging is critical for identifying and pinpointing abnormalities like brain tumors, which is vital for accurate diagnosis, effective treatment planning, and continuous patient monitoring. Traditionally, this task depended on manual interpretation by radiologists or basic automated methods, both of which could be slow and susceptible to errors. However, the advent of deep learning, particularly through convolutional neural networks (CNNs), has transformed this field by automating detection, greatly improving both the speed and accuracy of the process. These advancements have resulted in more reliable diagnostic outcomes, facilitating timely and consistent clinical interventions that enhance patient care.

One of the most notable advancements in real-time brain tumor detection is the implementation of the YOLO (You Only Look Once) architecture. Almufareh et al. [14] utilized YOLO to create a system capable of rapidly and accurately identifying brain tumors in real-time scenarios. YOLO's architecture processes images in a single pass, making it highly efficient for time-sensitive applications. Their study showed that the YOLO-based model achieved high detection accuracy, with a mean average precision (mAP) exceeding 80%. This blend of efficiency and accuracy is particularly valuable in clinical settings where quick decision-making is crucial, such as during surgical procedures or emergency situations.

Building on the strengths of YOLO, Iriawan and A. [16] introduced YOLO-UNet, a hybrid model that merges YOLO's rapid detection capabilities with U-Net’s precise segmentation. This integration enables the model to detect tumors quickly while also accurately delineating their boundaries in MRI scans. YOLO-UNet demonstrated significant improvements in both detection and segmentation accuracy, achieving a Dice similarity coefficient (DSC) of approximately 0.88, along with high detection precision. This comprehensive tool enhances clinical practice by providing both tumor

localization and detailed mapping, which are crucial for treatment planning and monitoring.

Notable advancements in detection accuracy have been realized by integrating Mask R-CNN and transfer learning techniques. In their research, Kordemir et al. [15] utilized Mask R-CNN, an enhanced version of the Faster R-CNN framework, to both detect and segment brain tumors simultaneously. Mask R-CNN introduces an additional branch specifically designed for predicting segmentation masks, a feature that sets it apart from its predecessor. This addition not only boosts the accuracy of tumor detection but also significantly improves the precision in defining the boundaries of tumors. The dual functionality of Mask R-CNN makes it particularly effective for handling complex brain MRI scans, where accurately distinguishing tumor margins is essential for diagnosis and treatment planning. The study highlights the model's ability to manage the intricate details of brain imaging, ensuring that both detection and segmentation are conducted with a higher degree of accuracy and reliability. Kordemir et al. reported that Mask R-CNN achieved a mean average precision (mAP) of about 85%, demonstrating its strong capability in accurately detecting and segmenting tumors, even in challenging imaging conditions. The success of Mask R-CNN underscores its potential as a powerful tool in clinical settings, where accurate and efficient tumor characterization is crucial for improving patient outcomes.

Mathivanan [17] achieved significant advancements in detection models by utilizing transfer learning, a technique that allows a model initially trained on a large dataset to be fine-tuned for specific tasks such as brain tumor detection. This method is particularly advantageous in the field of medical imaging, where large, annotated datasets are often limited. Mathivanan's research revealed that transfer learning could improve detection accuracy by around 10% compared to models developed from the ground up. This approach not only enhances model performance but also decreases the time and computational resources needed for training. As a result, it becomes more feasible and efficient for clinical settings where high precision is crucial. The study underscores the practicality of transfer learning in overcoming the challenges posed by limited data availability, highlighting its potential to deliver more accurate and reliable detection outcomes in real-world medical environments.

These advances in brain tumor detection models illustrate the field's ongoing progress. YOLO-based methods offer the necessary speed for real-time detection, while hybrid models like YOLO-UNet provide the added advantage of detailed segmentation. Additionally, Mask R-CNN and transfer learning techniques further enhance detection accuracy, making these models increasingly reliable for clinical application. As these technologies continue to evolve, they are expected to significantly improve patient outcomes by enabling faster and more precise brain tumor detection.

2.2 Key Datasets and Benchmarking Tools

In brain tumor classification, segmentation, and detection, access to high-quality annotated

datasets and robust benchmarking tools is essential for the development and evaluation of deep learning models. These resources provide the critical data needed for training and testing, ensuring that models are both accurate and adaptable to diverse clinical scenarios.

The Brain Tumor Image Segmentation (BraTS) dataset is widely recognized as one of the most comprehensive and frequently utilized resources in brain tumor research. It includes MRI scans from patients with glioblastoma and lower-grade glioma, accompanied by detailed segmentation masks. The dataset encompasses various MRI modalities, including T1, T2, FLAIR, and post-contrast T1-weighted images, offering a valuable resource for training and testing models [18]. BraTS has been instrumental in numerous segmentation challenges, driving the development of innovative algorithms and encouraging collaboration within the research community [19].

The Cancer Genome Atlas (TCGA) Glioblastoma Multiforme (GBM) and Lower Grade Glioma (LGG) datasets consist of extensive collections of MRI scans. These datasets are essential for imaging analysis, supporting multimodal research approaches that combine different imaging modalities to enhance tumor characterization and classification [20].

Additionally, the Cancer Imaging Archive (TCIA) acts as a publicly available repository for cancer imaging data, including MRI scans of brain tumors. It promotes data sharing within the research community, aiding in the development of new algorithms and the replication of study findings. TCIA's standardized data formats and detailed metadata support rigorous benchmarking and allow for effective comparison of different models [21].

Access to high-quality datasets and dependable benchmarking tools is crucial for advancing research in brain tumor classification, segmentation, and detection. These resources are fundamental to creating more accurate, robust, and generalizable deep learning models, ultimately contributing to improved clinical outcomes for brain tumor patients.

2.3 Current Trends and Innovations

The field of brain tumor classification, segmentation, and detection has undergone substantial advancements with the integration of innovative deep learning architectures, significantly improving both the accuracy and efficiency of these tasks. Transformer-based models, such as Vision Transformer (ViT) and TransUNet, have introduced notable enhancements by effectively capturing long-range dependencies in MRI data. Initially developed for natural language processing, these models excel in medical imaging by seamlessly blending local and global contextual information, which is critical for precise brain tumor identification and segmentation. For instance, TransUNet combines the advantages of transformers with the traditional U-Net architecture, leading to marked improvements in segmentation performance. Additionally, Generative Adversarial Networks (GANs) have addressed the challenge of data scarcity in medical imaging by generating synthetic MRI scans. This

approach not only increases the volume and diversity of training data but also mitigates overfitting and improves the generalization capability of deep learning models, which is essential for successful clinical applications.

Unsupervised and semi-supervised learning techniques are gaining importance, especially in scenarios where annotated data is limited. Techniques such as autoencoders and self-supervised learning enable models to extract valuable representations from large amounts of unlabeled data, which can then be fine-tuned with smaller, annotated datasets. This method reduces the reliance on extensive labeled data and enhances the model's ability to effectively learn from the available data. Additionally, Multi-task Learning (MTL) has become a powerful approach for training models to perform multiple related tasks, such as tumor classification, segmentation, and detection, within a single framework. MTL improves model efficiency and generalization by facilitating knowledge sharing across tasks, thereby capturing more intricate relationships within the data.

The advancement of 3D Convolutional Neural Networks (3D CNNs) has been particularly impactful, especially in preserving the three-dimensional structure of brain tumors in MRI scans. By processing volumetric data in its entirety, 3D CNNs enable more accurate segmentation and detection, particularly in cases involving irregularly shaped tumors.

Reinforcement learning is also being explored to optimize diagnostic processes, allowing models to learn and adapt through interactions with their environment. This approach could lead to more dynamic and efficient diagnostic tools. Furthermore, the growing focus on model explainability and interpretability is vital for building clinical trust. Techniques such as attention mechanisms and saliency maps are being employed to provide insights into model decisions, ensuring that these advanced technologies can be smoothly integrated into clinical workflows.

Collectively, these advancements represent significant progress in brain tumor imaging, offering promising pathways for improving diagnostic accuracy, treatment planning, and patient outcomes. The next section offers a comprehensive analysis of the deep learning models utilized in this study, discussing their foundational principles as well as their respective strengths and limitations.

2.4 Models

In recent years, deep learning has transformed medical imaging, particularly in brain tumor classification, segmentation, and detection. These tasks are crucial for ensuring accurate diagnoses and developing effective treatment strategies. The emergence of advanced neural network architectures has greatly improved the precision and efficiency of these processes. This section discusses the models used in this study, focusing on their underlying principles, as well as their respective advantages and disadvantages. By exploring these aspects, this discussion provides a comprehensive overview of how deep learning is shaping the future of brain tumor diagnosis and treatment.

2.4.1 Classification Models

In this study, a diverse set of classification models Vision Transformer (ViT), ResNet50, ResNet152, DenseNet121, and DeiT was selected to explore their architectural diversity and effectiveness in brain tumor classification. Each model offers unique advantages: ViT provides a transformer-based approach that excels in capturing global context, while ResNet50 and ResNet152 allow for a comparison of different depths within residual networks, demonstrating how depth impacts performance. DenseNet121's dense connectivity enhances feature propagation, which is particularly beneficial for complex medical imaging. DeiT, designed for data efficiency, was included to assess how transformers perform with smaller datasets. This selection enables a thorough assessment of how different architectures and model depths impact the accuracy and efficiency of brain tumor classification.

2.4.1.1 Vision Transformer (VIT)

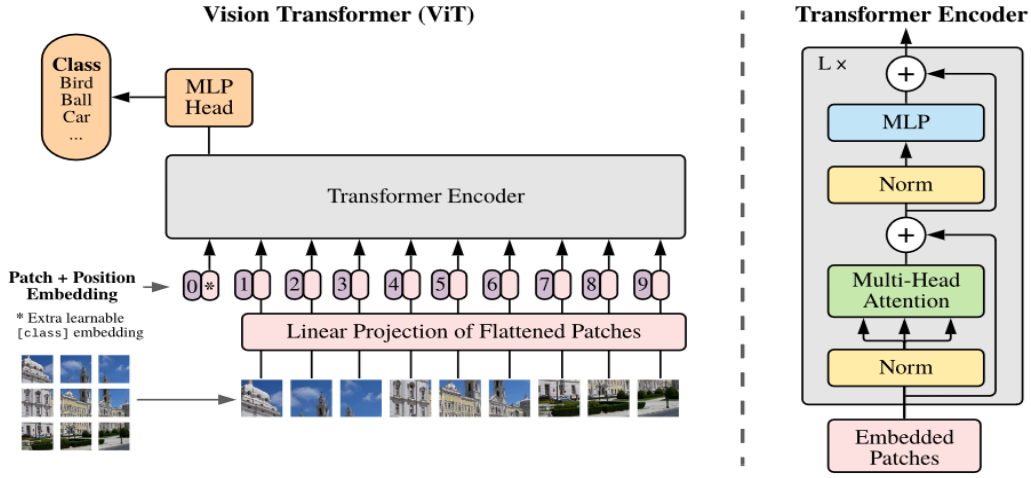


Figure 2.4.1 : The Vision Transformer (ViT) architecture divides the input image into patches, embeds them, and processes them through a transformer encoder. Image source Dosovitskiy et al. (2020) [22].

The Vision Transformer (ViT), introduced by Dosovitskiy et al. [22], employs a transformer-based architecture for image classification, marking a significant departure from traditional convolutional neural networks (CNNs) (as shown in Figure 2.4.1). ViT breaks down input images into fixed size patches, which are then embedded with positional encodings to maintain spatial context. These patches are processed through a transformer encoder composed of multi-head self-attention and feed-forward layers, enabling ViT to capture intricate, long-range dependencies within the image. This capability makes it particularly suitable for medical imaging tasks that require a comprehensive understanding of spatial relationships. While ViT excels at managing complex global contexts, which is crucial for the accurate interpretation of medical images, its effectiveness largely depends on access to large datasets, which can be challenging to obtain in medical settings. Furthermore, the model's significant computational demands may present obstacles in resource-constrained

environments. Despite these challenges, ViT represents a major advancement in the field of image classification models, particularly as data availability and computational resources continue to improve.

2.4.1.2 ResNet (ResNet50, ResNet152)

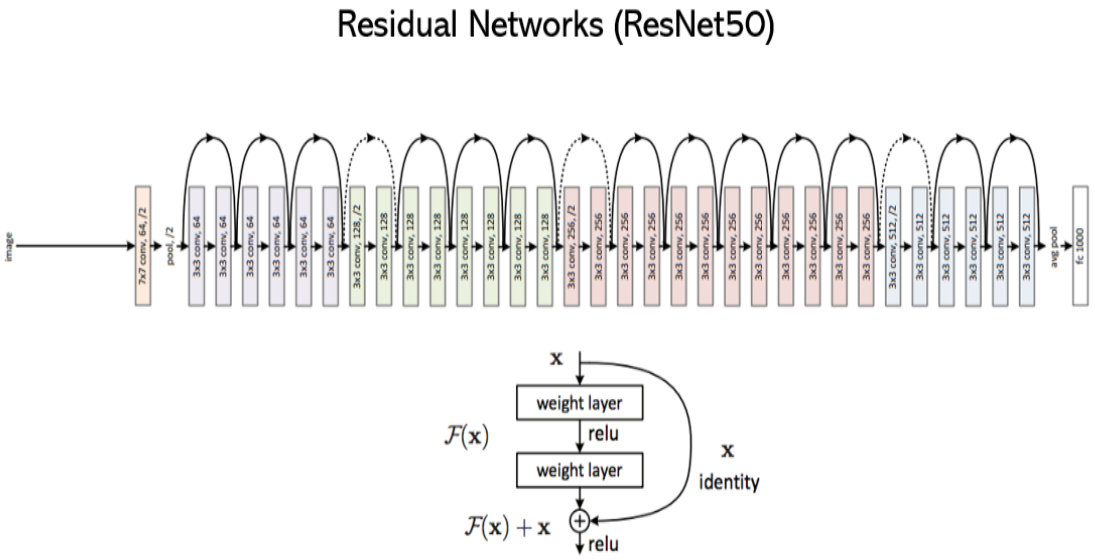


Figure 2.4.2 : The Residual Network (ResNet50) architecture includes convolutional layers with residual connections that bypass specific layers. Image source [24].

The Residual Network (ResNet), particularly its ResNet50 and ResNet152 variants, represents a significant advancement in deep learning, as introduced by He et al. [23]. ResNet effectively tackles the vanishing gradient problem, a frequent challenge encountered when training deep neural networks. The core innovation of ResNet lies in its application of residual learning, which incorporates shortcut or skip connections that bypass certain layers within the network. This design allows the network to focus on learning residual functions instead of directly approximating the original functions, thereby facilitating the training of much deeper networks without experiencing performance degradation. This approach is crucial for maintaining the integrity and efficiency of the network as it scales in depth, allowing for more complex and accurate model development.

ResNet's architecture, illustrated in Figure 2.4.2, features a sequence of convolutional layers interconnected by residual blocks. These blocks help the model retain critical information from earlier layers while simultaneously learning more intricate features in the deeper layers. ResNet50, with its 50 layers, achieves an ideal balance between network depth and computational efficiency, making it especially well-suited for complex classification tasks, such as those encountered in medical imaging.

ResNet152 takes this concept further by incorporating 152 layers, offering even greater depth and

feature extraction capabilities. The increased depth in ResNet152 makes it particularly effective at capturing detailed features across multiple levels of abstraction, which is essential for tasks requiring nuanced feature detection, such as in medical imaging. However, this increased depth requires greater computational resources, emphasizing the trade-off between model complexity and computational efficiency. Despite this, ResNet illustrates how deeper networks can be trained more efficiently, resulting in improved performance for complex classification tasks.

2.4.1.3 Densely Connected Convolutional Network 121 (DenseNet121).

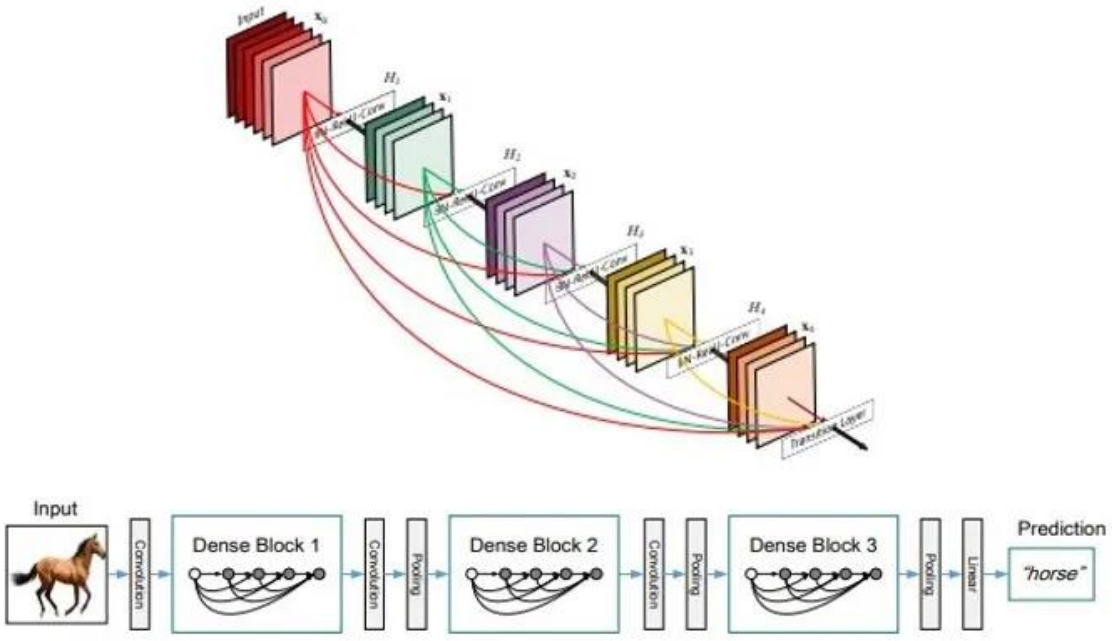


Figure 2.4.3 : The DenseNet121 architecture features dense blocks where each layer connects to every preceding layer, enhancing feature propagation and reducing parameters. Image source [26].

DenseNet121, as proposed by Huang et al. [25], introduces a unique approach to deep learning with its dense connectivity structure, distinguishing it from models like ResNet. Unlike ResNet's use of residual connections, DenseNet121 connects each layer to every preceding layer within the same dense block, ensuring that features are reused across the network. This architecture, composed of densely connected blocks and transition layers, enhances feature propagation while maintaining computational efficiency. In medical imaging, where capturing fine details is crucial, DenseNet121's architecture stands out by minimizing the risk of vanishing gradients and allowing the network to learn more complex representations. The illustration of DenseNet121 (Figure 2.4.3) demonstrates this approach, showing how the model densely connects layers, with each dense block being fully connected to every layer within it. This design facilitates a continuous flow of information and feature reuse across the network, leading to more robust learning. However, the increased memory demands due to the dense connections may present challenges in environments with limited

resources. Nonetheless, its capacity to extract and utilize detailed features with relatively fewer parameters underscores its effectiveness in complex classification tasks, particularly in scenarios demanding high precision, such as medical imaging. The ability of DenseNet121 to maintain both depth and computational efficiency makes it an invaluable asset in deep learning.

2.4.1.4 Data-efficient Image Transformer (DEIT) Model.

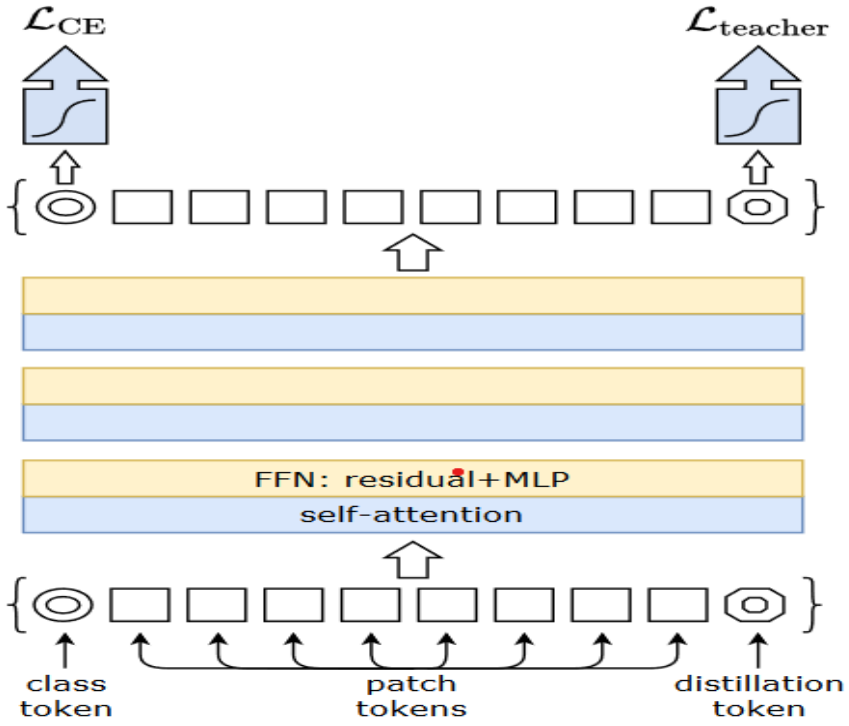


Figure 2.4.4 : The DeiT architecture integrates a distillation token alongside class tokens and patch tokens. This design leverages knowledge distillation during training to enhance the model's efficiency and performance. Image source: Touvron et al. (2021) [27].

The DeiT model, introduced by Touvron et al. [27], represents a significant innovation in transformer-based architectures, especially for image classification tasks. Unlike traditional transformers, which often require extensive datasets for effective training, DeiT enhances data efficiency by introducing a distillation token alongside the usual class and patch tokens. As shown in the accompanying Figure 2.4.4, this distillation token is integrated into the transformer encoder, where it undergoes the same self-attention and feed-forward processing as the other tokens. This design allows the model to benefit from knowledge distillation, enabling a smaller model to learn from a larger, more complex teacher model, thereby maintaining high accuracy even with less data. The image illustrates how the distillation token works in conjunction with other tokens to refine the model's output, enhancing both performance and efficiency. While DeiT reduces the reliance on large datasets and computational resources, its reliance on a teacher model for distillation introduces additional complexity to the training process. This trade-off between improved data efficiency and increased training complexity must be carefully considered, particularly in applications like medical imaging where both data

scarcity and high performance are critical.

2.4.2 Segmentation Models

For brain tumor segmentation, this study employs a diverse array of models U-Net, DeepLabv3, MAnet, and UNet++ to evaluate their respective strengths and weaknesses in processing medical imaging data. U-Net was specifically selected due to its design tailored for medical image segmentation, rendering it particularly effective in accurately delineating tumors. DeepLabv3 was incorporated for its robust multi-scale processing capabilities, which are crucial for effectively segmenting tumors that exhibit significant variations in shape and size. MA-Net was chosen for its attention mechanism, which enhances the model's focus on critical regions within the image, leading to improved accuracy in detecting tumor boundaries. Additionally, UNet++ was selected for its advanced architecture that builds on the strengths of U-Net by incorporating nested skip connections, thereby boosting segmentation performance. This carefully curated selection of models ensures a comprehensive evaluation of various segmentation strategies within the context of brain tumor imaging, enabling a critical assessment of their applicability and effectiveness in medical diagnostics.

2.4.2.1 U-Net

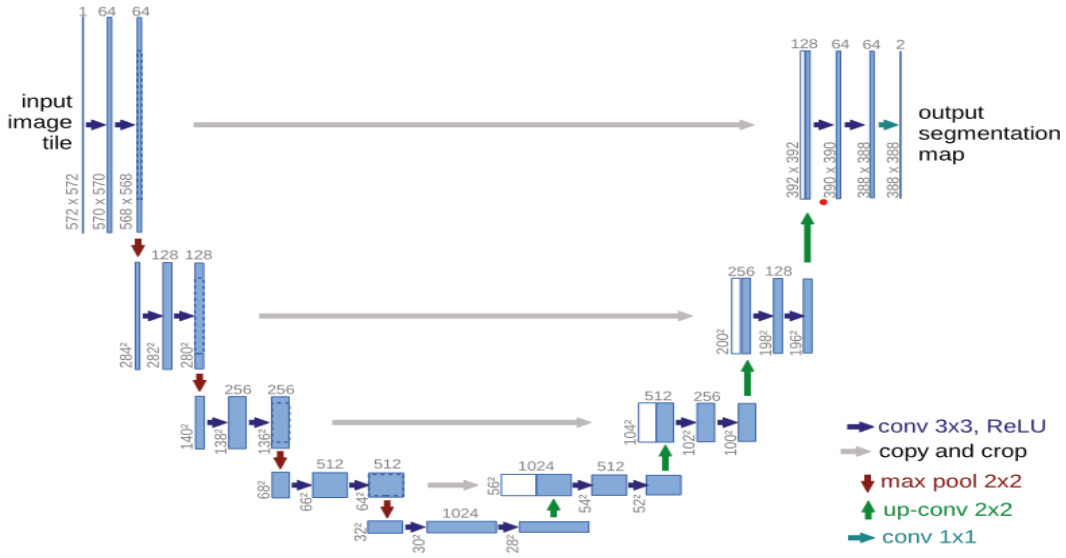


Figure 2.4.5 :The U-Net architecture features a contracting path that captures contextual information and an expansive path with skip connections that enable precise segmentation. Image source: Ronneberger et al. (2015) [28].

The U-Net architecture, presented by Ronneberger et al. [28], is a leading model in biomedical image segmentation, recognized for its dual-path design. It includes a contracting path that gradually reduces the spatial dimensions of the image while extracting essential features, and an expansive path that incrementally restores the image size, integrating these features to produce accurate segmentations. The contracting path focuses on compressing the image into deeper feature maps, capturing

intricate details necessary for precise segmentation. The expansive path, as depicted in Figure 2.4.5, restores the image size while merging detailed features from the contracting path, ensuring high accuracy. Additionally, UNet++ [30] builds on this foundation by adding nested skip connections, which further enhance segmentation accuracy in complex cases. While U-Net is highly effective, its performance is heavily reliant on extensive annotated datasets, a challenge in data-limited settings. Despite this, U-Net's design has made it a cornerstone in medical image segmentation tasks.

2.4.2.2 Deeplabv3

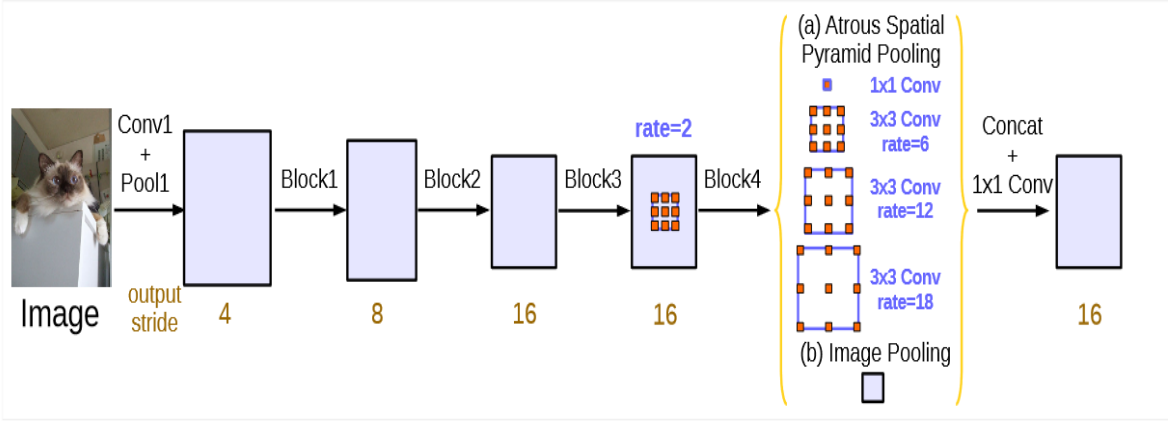


Figure 2.4.6 : The DeepLabv3 architecture utilizes Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale contextual information, which is essential for accurate segmentation in complex images. Image source: Chen et al. (2017). [29].

Deeplabv3, developed by Chen et al. [29], is a powerful model for semantic segmentation that excels in scenarios involving objects of varying scales within the same image. The model incorporates Atrous Spatial Pyramid Pooling (ASPP), a technique designed to capture features at multiple scales effectively. ASPP accomplishes this by utilizing convolutional filters with varying dilation rates (atrous rates), enabling the model to simultaneously analyze the image at multiple levels of detail. For instance, one set of filters might focus on finer details, while another captures broader contextual information. This multi-scale feature extraction is illustrated in Figure 2.4.6, where different layers of ASPP capture varying spatial resolutions and combine them to produce a comprehensive and detailed segmentation map. This capability to gather features from multiple scales makes DeepLabv3 especially advantageous in medical imaging, where accurately delineating structures of varying sizes is crucial. However, the model's advanced architecture and the complexity of multi-scale processing also heighten computational demands, which could be a drawback in resource-limited settings. Despite these challenges, Deeplabv3's innovative approach to multi-scale feature extraction ensures its position as a leading model in semantic segmentation, offering high accuracy in complex and variable imaging tasks.

2.4.2.3 Multi-Attention Network (MaNet).

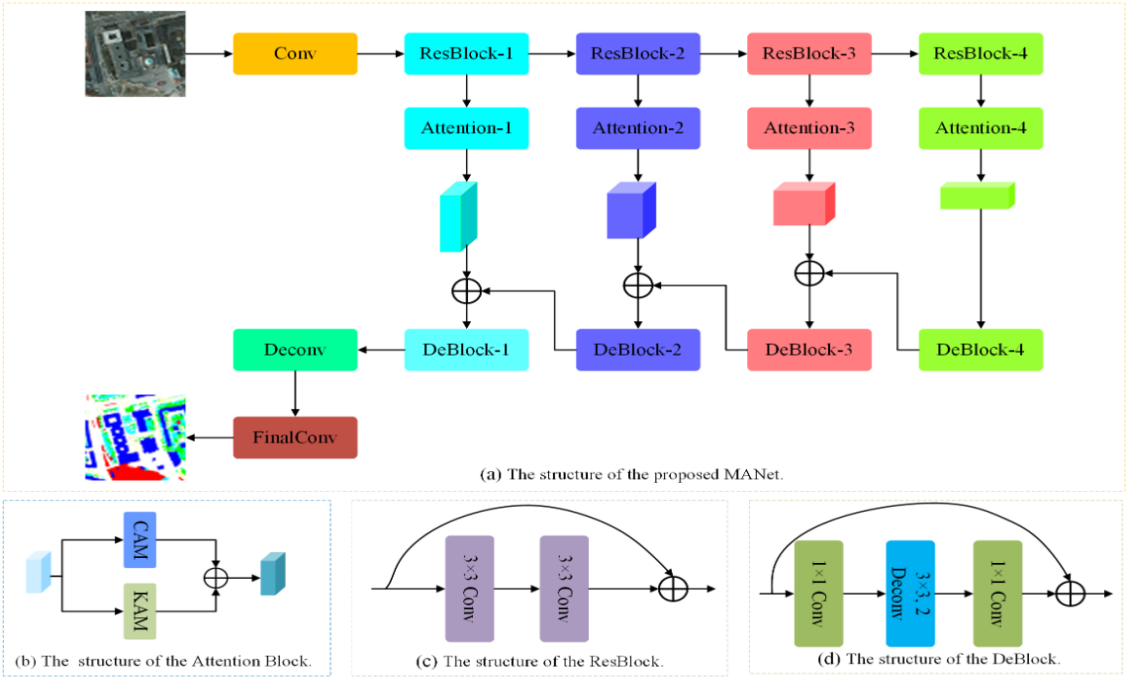


Figure 2.4.7 : The MANet architecture combines residual blocks with attention mechanisms, specifically channel attention (CAM) and kernel attention (KAM), to enhance feature extraction and segmentation accuracy. Image source: Zhang et al. [31].

The MANet architecture, as introduced by Zhang et al. [31], is designed to enhance semantic segmentation tasks by incorporating sophisticated attention mechanisms. The architecture integrates residual blocks with dual attention modules Channel Attention Mechanism (CAM) and Kernel Attention Mechanism (KAM). The CAM selectively amplifies critical features by concentrating on the most important channels within the feature maps, effectively assigning weight to the relevance of different feature channels. In contrast, the KAM refines spatial attention by focusing on critical areas within the convolutional kernels, allowing the model to emphasize spatial regions that are more relevant to the segmentation task. As depicted in Figure 2.4.7, MANet's structure is composed of sequential residual blocks that maintain the flow of information, while the embedded attention modules selectively highlight crucial features. This makes MANet particularly effective in medical imaging, where capturing fine-grained details is essential. Despite the increased complexity introduced by the attention mechanisms, MA-Net is engineered to balance accuracy and computational efficiency, making it a versatile and effective tool for detailed image segmentation tasks. However, the added complexity from these attention mechanisms can also present a drawback.

2.4.3 Detection Models

For the task of brain tumor detection, this study strategically employs two powerful models RetinaNet and Faster R-CNN to thoroughly assess their effectiveness in processing medical imaging data. RetinaNet was selected primarily for its innovative Focal Loss function, which effectively addresses

class imbalance, a common challenge in medical imaging datasets, and its capability to detect objects across various scales. This combination of accuracy and computational efficiency makes RetinaNet well-suited for clinical environments where timely and reliable detection is essential. Conversely, Faster R-CNN was selected for its robust Region Proposal Network (RPN), which excels at identifying potential tumor regions within complex and heterogeneous images. The availability of pre-trained models for both architectures also facilitates their integration and adaptation to the specific challenges posed by brain tumor detection, ensuring a comprehensive evaluation of their applicability and reliability in medical diagnostics.

2.4.3.1 RetinaNet

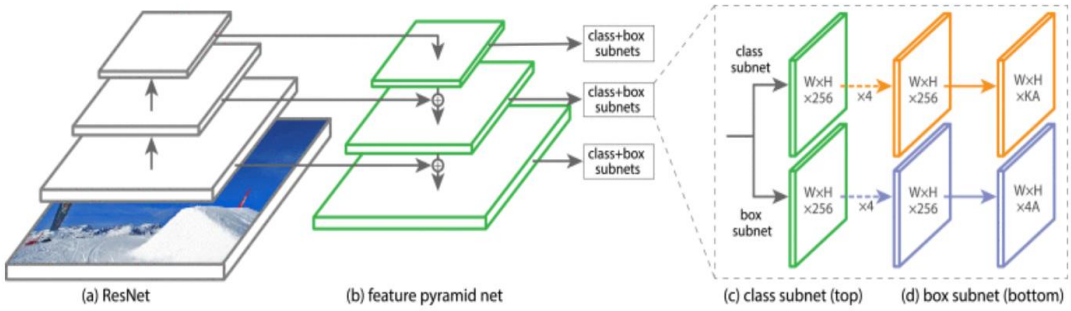


Figure 2.4.8 : The RetinaNet architecture integrates a Feature Pyramid Network (FPN) with a ResNet backbone. The feature maps generated are processed by separate class and box subnets to achieve accurate object detection. Image source: Lin et al., 2017 [32].

RetinaNet, introduced by Lin et al. [32], stands as a groundbreaking model in the field of object detection, specifically crafted to address the challenges of class imbalance commonly found in medical imaging datasets. The model incorporates a ResNet backbone in conjunction with a Feature Pyramid Network (FPN), which plays a vital role in capturing features at multiple scales. The FPN constructs a pyramid of feature maps at different resolutions, enabling the model to detect objects of varying sizes with high precision. This multi-scale approach allows RetinaNet to efficiently process both fine and coarse details, making it exceptionally effective at detecting tumors in medical images. As illustrated in Figure 2.4.8, RetinaNet consists of two subnets that operate on the multi-scale feature maps produced by the FPN: one subnet is dedicated to object classification, while the other handles bounding box regression. A key innovation in RetinaNet is the Focal Loss function, which dynamically adjusts the model's focus to prioritize difficult-to-classify examples, thereby effectively addressing class imbalance. Although RetinaNet requires significant computational resources, its ability to balance precision and recall across objects of different scales makes it an excellent choice for complex tasks such as brain tumor detection in medical imaging.

2.4.3.2 Faster Region-based Convolutional Neural Network (Faster R-CNN).

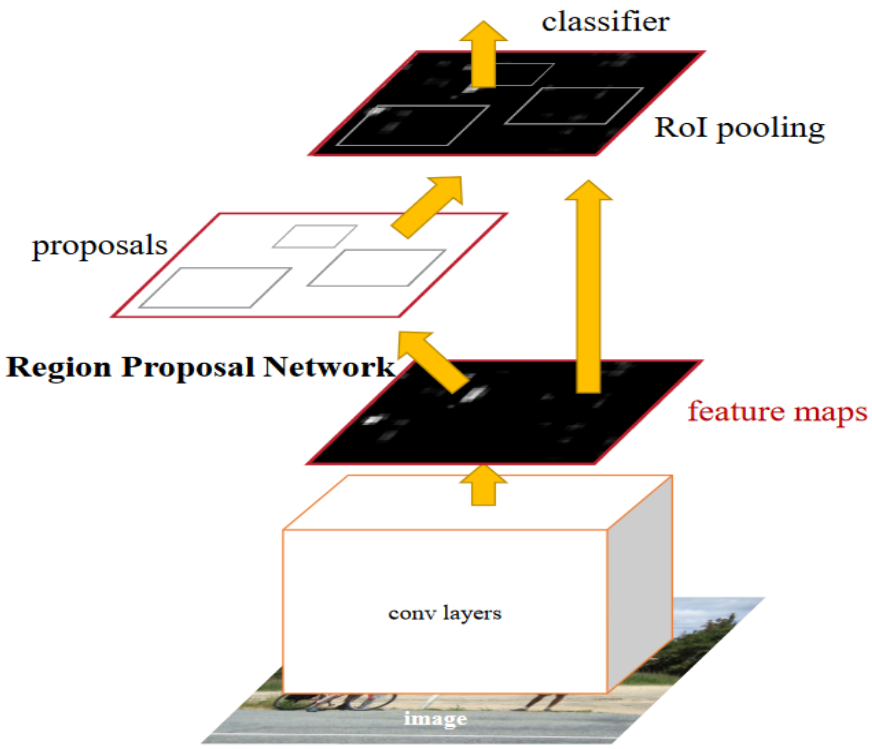


Figure 2.4.9 : The Faster R-CNN architecture combines a Region Proposal Network (RPN) with convolutional layers to generate region proposals. These proposals are then classified and refined using RoI pooling, enabling precise object detection. Image source: Ren et al. (2015) [33].

The Faster R-CNN model, introduced by Ren et al. [33], is a widely respected framework in object detection, known for its efficient two-stage process that successfully balances accuracy and computational efficiency. This architecture combines a Region Proposal Network (RPN) with a convolutional backbone to generate and refine region proposals within an input image (As illustrated in Figure 2.4.9). The RPN functions by sliding a small neural network, typically a 3x3 convolutional layer, across the feature map. At each spatial location, it predicts the presence of an object and provides bounding box coordinates for the proposed regions. These proposals, which represent potential object locations, are then processed further in the second stage, where the model classifies each region and fine-tunes the bounding boxes for precise localization.

The RoI (Region of Interest) pooling step standardizes these proposals to ensure a uniform input size for the classifier, enabling consistent evaluation. The final classification layer then assigns object categories, while the bounding box regressor fine-tunes the coordinates for greater accuracy. A key strength of Faster R-CNN is its unified approach, which consolidates region proposal generation and object detection into a single network, thereby reducing computational overhead compared to earlier models that handled these tasks separately. However, this integrated design also increases computational demands, which may limit its applicability in real-time scenarios or environments with limited

resources.

In medical imaging tasks, selecting the right models is essential for achieving accurate and clinically meaningful outcomes. This study underscores the need to balance model complexity with the available computational resources and the specific requirements of clinical settings. Advanced models like MA-Net and DeepLabv3 offer superior feature extraction and multi-scale processing capabilities, but their higher computational demands may present challenges in resource-limited environments. Conversely, models like U-Net and Faster R-CNN deliver reliable performance with more manageable resource requirements, making them more suitable for real-time clinical applications. Ultimately, the choice of model should be tailored to the task's needs, ensuring both effectiveness and practical applicability in clinical practice.

2.5 Summary

The literature review highlights significant advancements in brain tumor classification, segmentation, and detection, fueled by the adoption of advanced deep learning models like Vision Transformer (ViT) and U-Net. These models have significantly enhanced diagnostic accuracy by improving the precision of medical imaging tasks. The availability of large-scale datasets, such as BraTS and TCGA, has been instrumental in these developments, offering essential data for model training and evaluation. However, challenges remain, particularly the reliance on large, annotated datasets, which are scarce and expensive to produce in clinical settings. This limitation hinders the practical application of these models, especially in resource-constrained environments. Additionally, issues like class imbalance within datasets can lead to biased model performance, reducing their ability to generalize across diverse patient populations.

Moreover, the computational demands of these models and their complex architectures often obscure their decision-making processes, complicating their clinical adoption where explainability is crucial. The review highlights that many studies in the field often lack a cohesive evaluation framework, typically focusing on a single task or a combination of two, without addressing the full spectrum of classification, segmentation, and detection tasks. This limitation suggests a need for future research to develop more comprehensive approaches that integrate multiple deep learning architectures within a single study. Addressing this gap could lead to the development of more data-efficient models, improved management of class imbalance, and a stronger integration of tasks, ultimately enhancing the strengths of various architectures and increasing the overall effectiveness of these technologies.

3 METHODOLOGY

This section outlines the methodological framework utilized to address the challenges of brain tumor classification, segmentation, and detection using advanced deep learning models. The approach is carefully designed to manage the inherent complexities of these tasks, aiming for both high accuracy and practical applicability in medical imaging.

The core of this methodological approach is rooted in the integration of deep learning architectures, which have been rigorously validated through a comprehensive literature review. This foundation ensures that the models are not only theoretically sound but also effective in practical, clinical settings, addressing the unique demands of medical imaging with precision and reliability. The selected models were specifically chosen for their ability to address the unique challenges of medical imaging, such as variability in tumor sizes, shapes, and types. Key architectural features, including deep networks, attention mechanisms, skip connections, and Focal Loss, were critical factors in the selection process, ensuring that each model is optimally equipped to enhance performance. The framework is designed to transition seamlessly from theoretical insights to practical applications, with a strong emphasis on tailoring each model to the specific characteristics of the datasets used.

The methodology begins with data acquisition and preprocessing, where datasets are carefully curated and processed to ensure high-quality inputs for the models. This is followed by the model development phase, during which each model is fine-tuned and optimized to meet the specific requirements of its respective task. Hyperparameter tuning is a key aspect of this process, ensuring that the models achieve both high accuracy and computational efficiency. The overall approach is focused on balancing model complexity with practical usability, particularly in resource-constrained environments. The integration of these advanced models aims to enhance diagnostic accuracy in brain tumor classification, segmentation, and detection, ultimately leading to improved clinical outcomes.

3.1 Data Acquisition and Description

The datasets used in this study were sourced from Kaggle, a prominent platform for data science competitions and datasets. Three primary datasets were employed, each providing unique characteristics essential for a thorough approach to brain tumor classification, segmentation, and detection.

3.1.1 Dataset 1

The first dataset, the Brain MRI Dataset [34], consists of 10,000 MRI images, which are systematically divided into training, validation, and test sets. The training set includes 6,000 images, evenly distributed across four categories: Glioma, Meningioma, Pituitary, and No Tumor, with 2,000 images in each category. Both the validation and test sets contain 1,000 images each, with 250 images per

category. This balanced dataset ensures equal representation of each tumor type, facilitating robust model training, validation, and testing. The structured division also enhances the models' ability to generalize across a wide range of tumor types.

3.1.2 Dataset 2

The second dataset, the Brain Tumor Segmentation Dataset [35], comprises 4,237 MRI images, each paired with a corresponding segmentation mask. The images are divided into categories, including 1,595 No Tumor images, 649 Glioma images, 999 Meningioma images, and 994 Pituitary tumor images. This dataset provides a wide variety of tumor appearances, increasing its value for different aspects of brain tumor analysis. To expand the dataset's applicability to detection tasks, the segmentation masks were converted into bounding boxes using a custom contour detection algorithm. This algorithm automatically traces the edges of tumors in the images, and these traces are then used to create rectangular frames, or bounding boxes, around each tumor. This process broadens the dataset's use, making it a versatile resource for both segmentation and detection studies.

3.1.3 Dataset 3

The third dataset, the Brain Tumor YOLO Dataset [36], contains 3,064 MRI images, each with annotations for tumor detection. It is structured into a training set with 931 Pituitary tumor images, 817 Glioma images, and 705 Meningioma images. The validation set comprises 307 Glioma images, with no representation of Pituitary or Meningioma tumors, while the test set includes 305 Glioma images and 3 Meningioma images, with no Pituitary tumors. This dataset is mainly used for training and validating tumor detection models aimed at accurately identifying and localizing tumors within MRI scans. However, it is important to highlight the significant imbalance in this dataset, with substantial disparities in the representation of different tumor types across the training, validation, and test sets. This imbalance could cause the model to favour the more common classes, potentially leading to reduced performance on the less represented classes, as discussed in section 4.3.1.

Collectively, these datasets offer a comprehensive range of tumor types, sizes, and imaging modalities, making them invaluable for developing and testing deep learning models in brain tumor classification, segmentation, and detection. Their unique features ensure that the models are trained on high-quality, diverse data, enabling them to effectively handle the complexities of medical imaging.

3.2 Data Preprocessing and Augmentation

To ensure the effectiveness and generalizability of deep learning models in tasks such as brain tumor classification, segmentation, and detection, a variety of data preprocessing and augmentation techniques are employed. These techniques are specifically selected to address the inherent challenges of

medical imaging, including the wide variability in tumor appearances, differing imaging conditions, and the critical need for models to generalize robustly across diverse datasets. The enhanced generalizability resulting from these techniques is evident in the results. This section first provides an overview of these techniques and then elaborates on their application to each of the tasks.

- **Resizing --**This process involves scaling all images to a uniform size, which is essential to ensure that the input dimensions match the requirements of deep learning models.
- **Grayscale Conversion--** Images are converted to grayscale to reduce them to a single-color channel. This step is particularly important in medical imaging, where intensity variations are more critical than color variations.
- **Normalization --** Pixel values are scaled to a specific range, typically between 0 and 1, to stabilize and accelerate the training process. This scaling helps prevent issues like vanishing or exploding gradients, which can occur when training deep neural networks.
- **Random Flips --** Images are randomly flipped horizontally or vertically. This introduces variability in image orientation, which helps the model generalize better to different viewing angles.
- **Random Rotations --** Rotating images by a random angle (e.g., ±15 degrees) mimics the various orientations the model may encounter, thereby improving the model's robustness to rotational variations in the data.
- **Brightness Adjustments --**The brightness of images is randomly adjusted to simulate different lighting conditions. This helps the model learn to identify features under varying brightness levels, making it more adaptable to real-world scenarios.
- **Affine Transformations --**These include operations like scaling, shifting, and rotating images, simulating real-world distortions. This technique helps the model become more robust to variations in the shape and position of objects within the images.
- **Random Cropping** --This technique involves randomly cropping portions of the image, which is especially useful in detection tasks where the object of interest might not always be fully visible. It helps the model learn to detect objects even when they are partially obscured.
- **Random Resized Crops --** For segmentation tasks, random resized cropping is used to randomly crop and resize the images. This technique simulates various scales and positions of the object within the image, further improving the model's ability to generalize.
- **Albumentations Library** --The Albumentations library is employed in the detection task for efficient on-the-fly transformations. This library is particularly powerful for handling complex augmentations, enabling the model to generalize better by simulating various real-

world imaging scenarios during training.

- **Train-Validation-Test Split** -- Dataset 1, already pre-split, was utilized as provided for classification tasks. Datasets 2 and 3, used for segmentation and detection tasks, were partitioned into training (70%), validation (15%), and test (15%) sets.

3.2.1 Classification

MRI images are resized to 224x224 pixels and converted to grayscale to emphasize intensity variations critical for distinguishing different tumor types. After resizing, normalization is applied to scale the pixel values between 0 and 1. Data augmentation methods, such as random flips, rotations, and brightness adjustments, are utilized to generate a varied training set. These augmentations help the model generalize better to different datasets, making it more effective at classifying tumors across various imaging conditions.

3.2.2 Segmentation

MRI images and their corresponding segmentation masks are resized to 256x256 pixels to ensure consistent input dimensions. Grayscale conversion and normalization are applied to standardize the images, which facilitates more efficient feature extraction. Additionally, segmentation masks are converted into bounding boxes when the dataset is used for related detection tasks. Data augmentation techniques, such as random rotations, flips, affine transformations, and random resized crops, are applied to both images and masks simultaneously to preserve alignment, with the exception of normalization. This ensures that spatial changes to the image are accurately reflected in the mask, preserving anatomical accuracy. Normalization is avoided for masks because they typically represent categorical labels that do not have a meaningful range or scale, and altering them could misrepresent the actual classes. These augmentations ensure that the model can accurately delineate tumor boundaries even when the images vary in orientation, scale, or appearance.

3.2.3 Detection

MRI images are resized to 256x256 pixels and normalized to ensure consistency across the dataset. A key preprocessing step involves converting segmentation masks into bounding boxes using a contour detection algorithm, enabling the model to concentrate on specific areas of interest. Extensive data augmentation is applied, including techniques like random cropping, flips, and rotations, applied simultaneously to both images and masks, with normalization as the exception. This coordinated application is essential for ensuring that the geometric and spatial properties of the detected objects are not distorted, thereby maintaining the accuracy of the detection process. These augmentations help simulate various conditions the model might encounter in real-world scenarios, ensuring robust and accurate tumor detection across different imaging conditions.

3.3 Model Implementation and Fine-Tuning

Expanding on the models covered in the literature review section 2.4, this section explores the practical adaptation and implementation of these models for brain tumor classification, segmentation, and detection. The aim here is to move from the theoretical selection of models to their practical application in real-world settings. Here, we will provide a detailed account of how each chosen model was customized and fine-tuned to meet the specific demands of brain tumor tasks. The models were selected to explore various architectural features such as deeper layers, attention mechanisms, and multi-scale processing to assess their impact on performance, particularly in addressing data imbalance and ensuring robust brain tumor classification, segmentation, and detection.

The classification models, such as Vision Transformer (ViT), ResNet-50, ResNet-152, DenseNet121, and DEIT, were initially pretrained on the ImageNet dataset. These models were subsequently fine-tuned on the Brain MRI Dataset 1, which classifies images into categories including No Tumor, Glioma, Meningioma, and Pituitary tumors. Significant adaptations were made to these models to enhance their applicability to brain tumor detection. Modifications included adjusting input layers to process grayscale MRI inputs, which deviate from the standard RGB configuration used in traditional imaging datasets. Additionally, the final classification layers were restructured to align with the dataset's four-class categorization of tumors, ensuring that the models could effectively differentiate between various types of brain tumors.

For segmentation tasks, models like U-Net, U-Net++, Manet, and DeepLabV3 were employed. These models, originally pretrained on ImageNet, were tailored to segment brain tumors using the Brain Tumor Segmentation Dataset 2. Customizations included incorporating additional skip connections and deeper layers in U-Net and U-Net++ models, enhancing their ability to delineate complex tumor boundaries effectively. Additionally, adjustments to DeepLabV3’s atrous spatial pyramid pooling were made to capture multiscale contextual information, which is crucial for accurately segmenting various tumor types.

For detection tasks, Faster R-CNN and RetinaNet were fine-tuned using Brain Tumor Segmentation Dataset 2, where segmentation masks were converted into bounding boxes to provide a more diverse and accurate representation of tumors. This conversion enhanced RetinaNet's ability to detect tumors across varied cases. These models, initially pretrained on the COCO dataset, were adapted to process grayscale MRI images and re-structured to output classifications specific to different brain tumor types. Dataset 2, which was more balanced, served as the primary training set, helping to mitigate potential biases caused by data imbalance in the initial stages of model training. To ensure robustness and evaluate the models' effectiveness in real-world scenarios, Dataset 3, despite its imbalance, was utilized for external validation. This provided a critical testbed for assessing model performance on unseen data, validating the models' capabilities in diverse clinical environments.

Models were refined through adjustments to input layers, architectural changes, and fine-tuning on specialized datasets after pretraining. These modifications enhanced performance and reliability in practical scenarios.

For a comprehensive overview of the code implementations used in this study, including those for brain tumor classification, segmentation, and detection, as well as the conversion of segmentation masks into bounding boxes, please refer to the GitHub repository [37].

3.4 Hyperparameter Tuning and Optimization

The development of deep learning models for brain tumor classification, segmentation, and detection required careful design and fine-tuning of key components to ensure optimal performance, accuracy, and robustness.

- **Learning Rate** -- The learning rate is vital for model convergence as it controls the extent to which the model adjusts its weights in response to error during each update. It must be carefully tuned to balance the speed of convergence with the risk of overshooting the points of minimal loss.
- **Batch Size** – It affects memory usage and training stability. Larger batch sizes offer more stable gradient estimates but demand more memory, whereas smaller batch sizes are employed to conserve memory, particularly in complex models.
- **Training Epochs** -- This refers to the number of times the entire training dataset is passed through the model. Sufficient epochs enable thorough learning, but must be limited to prevent overfitting. Typically, models are trained for approximately 100 epochs, with regular monitoring of validation loss to determine the optimal stopping point.

Optimizer Selection:

- **SGD (Stochastic Gradient Descent)** – It is an optimization method known for its efficiency in navigating complex loss landscapes. It updates model parameters using small, random data batches, making it effective for tasks that require careful exploration through numerous local minima.
- **Adam** – It is an optimization algorithm known for its adaptive learning rate, which automatically adjusts the learning rate during training. This adaptability helps it manage different types of loss landscapes efficiently, making it a popular choice for many deep learning tasks.

Loss Functions:

- **Cross-Entropy Loss** – It is a loss function used in multi-class classification tasks. It measures the difference between the predicted probability distribution and the true labels, aiming to minimize this error.

- **Dice Loss** – It is a loss function used primarily in image segmentation tasks. It measures the overlap between predicted and true segmentation masks, emphasizing the spatial alignment of class labels for more accurate segmentation.
- **Focal Loss** -- It is a loss function that reduces the impact of easily classified examples, focusing more on difficult or misclassified instances. This makes it particularly effective in handling class imbalance in tasks like object detection.
- **Smooth L1 Loss** --It is a loss function used in bounding box regression for object detection. It combines the advantages of L1 and L2 losses, offering a balance between robustness to outliers and precise bounding box localization.

In classification models, dynamic adjustments of the learning rate are made throughout training to prevent overshooting optimal solutions, which is crucial for maintaining stable progression. Larger batch sizes are utilized to ensure faster convergence and more stable gradient estimates. The Adam optimizer is selected for its adaptability, efficiently handling the varied characteristics of brain tumor data. Cross-Entropy Loss is utilized to reduce the error between predicted probabilities and actual classifications, improving accuracy in distinguishing between multiple tumor types.

Segmentation models involve meticulous management of the learning rate with a scheduler that decreases the rate on plateaus to allow finer adjustments. This is essential for the detailed work required in segmentation maps. Smaller batch sizes are required because of the high memory demands involved in producing detailed outputs. The Adam optimizer's adaptive learning rate is essential for managing the loss variations in segmentation. Using a combination of Cross-Entropy and Dice Loss ensures both class accuracy and improved spatial accuracy for precise tumor delineation.

Detection models require robust learning rate management to handle the complexities of classifying and localizing objects simultaneously. Batch sizes are carefully moderated to balance memory usage and computational efficiency. SGD is selected for its ability to navigate complex loss landscapes effectively, critical for tasks combining object classification with localization. Focal Loss, paired with Smooth L1 Loss, optimizes the model for class imbalance and precise localization, ensuring accurate detection of less prevalent tumor types and exact tumor localization.

These models are finely tuned for classification, segmentation, and detection, enhancing their effectiveness and robustness in brain tumor analysis within clinical environments. Such optimization is vital for their successful application in medical imaging.

3.5 Experimental Setup

The experimental setup for this study was structured to accommodate the significant computational requirements of deep learning tasks, especially in the areas of brain tumor classification, segmentation, and detection. The hardware configuration included an NVIDIA RTX A4000 GPU, which is

well-regarded for its exceptional parallel processing capabilities, thereby significantly accelerating the training processes. This setup was complemented by 16GB of RAM, which provided the necessary capacity for managing large datasets and the extensive memory requirements during both training and testing phases.

The software environment was meticulously selected to complement the hardware. Ubuntu 20.04 LTS was chosen as the operating system for its reliability and strong support for scientific computing. The primary deep learning framework used was PyTorch, which was selected for its dynamic computation graph, making model development and training more intuitive and flexible. Additional support libraries such as NumPy and Pandas were employed for data manipulation tasks, while OpenCV and Pillow were utilized for image processing. This integration of advanced hardware and well-coordinated software tools created a strong and efficient experimental environment, ensuring that the training and testing of deep learning models were carried out effectively and with optimal performance.

3.6 Evaluation Metrics

To evaluate the deep learning models for brain tumor classification, segmentation, and detection, a variety of metrics were employed to ensure a thorough assessment of model performance across all tasks. These metrics were chosen to offer a detailed insight into how effectively each model performed its specific task, ensuring that the models were robust, accurate, and practical in real-world applications.

- **Accuracy** -- It calculates the ratio of correctly classified instances to the total number of instances.

$$Accuracy = \frac{True\ Positives\ (TP) + True\ Negatives\ (TN)}{Total\ Number\ of\ Instances} \tag{3.1}$$

- **F1 Score** -- Strikes a balance between precision and recall, which is essential for handling imbalanced classes.

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3.2}$$

- **Mean Average Precision (mAP)** -- Assesses the model's precision at different recall levels, which is particularly important for evaluating performance across multiple classes.

$$mAP = \frac{1}{N}\sum_{i-1}^{N} AP_i \tag{3.3}$$

- **Confusion Matrix** -- Breaks down the model's predictive accuracy by classifying predictions into true positives, true negatives, false positives, and false negatives.
- **ROC Curve and AUC** -- These metrics are valuable in binary classification, as they quantify the model's ability to distinguish between classes.
- **IoU (Intersection over Union)** -- Assesses the degree of overlap between the predicted and actual tumor boundaries.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{3.4}$$

- **Dice Coefficient** -- Another overlap metric, effective in scenarios with small tumor regions.

$$Dice\ Coefficient = \frac{2 \times |\ A \cap B\ |}{|\ A\ | + |\ B\ |} \tag{3.5}$$

- **Pixel Accuracy** -- Measures the proportion of correctly segmented pixels, offering a direct metric of segmentation performance.

$$Pixel\ Accuracy = \frac{Correctly\ Classified\ Pixels}{Total\ Pixels} \tag{3.6}$$

- **mAP (Detection)** -- Evaluates the precision of the model in detecting tumors at various recall levels, considering object localization.

$$mAP = \frac{1}{N}\sum_{i}^{N} AP_i \tag{3.7}$$

- **Confidence Score --** Confidence scores critically evaluate a model's certainty in tumor detections, serving as a pivotal metric for determining the validity of each detection by setting thresholds that effectively minimize false positives and maximize precision.

For classification, Metrics such as Accuracy, F1 Score, Mean Average Precision (mAP), Confusion Matrix, and ROC Curve with AUC are utilized. Accuracy offers an overview of model performance in tumor categorization. The F1 Score helps address class imbalances by balancing precision and recall, crucial in diverse medical datasets. mAP provides insight into precision across

various recall levels, essential for classifying multiple tumor types. The Confusion Matrix provides a detailed analysis of errors, while the ROC Curve with AUC measures the model's ability to differentiate between classes, especially in binary situations.

For segmentation tasks, key metrics include Intersection over Union (IoU), Dice Coefficient, and Pixel Accuracy. IoU and Dice Coefficient evaluate the accuracy of tumor boundary delineations compared to the ground truth, which is crucial for precise treatment planning. Although Pixel Accuracy might appear high due to the small size of tumors, it remains important for a straightforward assessment of segmentation quality.

In detection tasks, metrics such as Mean Average Precision (mAP) for Detection, Intersection over Union (IoU) for Detection, F1 Score, Precision, and Recall are utilized. mAP for Detection assesses the model’s ability to accurately detect and localize tumors, considering bounding box precision. IoU for Detection ensures these bounding boxes accurately match actual tumor locations, which is essential for precise localization. Confidence scores enhance detection reliability by quantifying prediction certainty and setting thresholds to filter out low-confidence, likely false positives, thereby optimizing clinical accuracy. F1 Score, Precision, and Recall play crucial roles in balancing the detection of true tumors while minimizing false positives, ensuring clinical reliability.

These metrics are carefully selected to meet the specific demands of each task, ensuring the models are not only accurate but also practical and reliable for clinical applications.

In conclusion, the methodology employed in this study has been meticulously designed to address the challenges of brain tumor classification, segmentation, and detection using advanced deep learning models. By incorporating sophisticated preprocessing techniques, fine-tuning hyperparameters, and utilizing robust evaluation metrics, the models have been optimized to tackle complex medical imaging tasks and effectively generalize to new data. While the methodology addresses challenges such as data imbalances and multi-class predictions, the strategies implemented are aimed at enhancing overall model performance and applicability. This methodological framework lays a solid foundation for the study, with the results and their implications to be discussed in the following section.

4 RESULTS

The results section provides a detailed analysis of the performance of various deep learning models used in brain tumor classification, segmentation, and detection. It emphasizes the effectiveness of these models, especially the influence of data augmentation techniques, and confirms the research hypotheses by presenting critical performance metrics. This section underscores the models' potential for real-world clinical applications while identifying areas for future improvement. By systematically presenting the results, it offers valuable insights into how different models perform under various conditions, with significant implications for enhancing brain tumor diagnosis and treatment.

To ensure a fair comparison between models within each task, the same hyperparameters were consistently used across all experiments for that specific task. These hyperparameters were carefully selected for each task based on their ability to maximize performance across the different model architectures. This approach allowed us to isolate the impact of different model architectures on performance without introducing variability from differing training configurations.

4.1 Classification

The study evaluated the performance of various classification models, including Vision Transformer (ViT), ResNet-50, ResNet-152, DenseNet121, and DEIT, each chosen for their distinct architectural strengths in medical imaging tasks. To assess their effectiveness in classifying brain tumors from MRI images, key metrics such as accuracy, F1-score, Mean Average Precision (mAP), ROC curves, and confusion matrices were employed. The classification reports provided comprehensive insights into how each model performed across different tumor types, highlighting both their strengths and areas where improvement is needed.

4.1.1 Model Performance with Augmentations

ResNet-152 emerged as the top performer with augmentation, achieving an accuracy of 88.3% and an F1-score of 0.88 (Figure 4.1.1). Its deep layers and residual connections effectively captured complex patterns, particularly excelling in detecting Meningioma and no_tumor categories, with high F1-scores of 0.93 and 0.91. The high mAP score of 0.95 further underscores its precision, confirming its position as the best model in this study.

DenseNet121 also performed strongly, achieving an accuracy of 83.4% and an F1-score of 0.835. Its dense connectivity facilitated efficient feature propagation, allowing it to perform well across diverse tumor types, particularly in detecting no_tumor and Meningioma, with F1-scores of 0.89 and 0.88. This indicates its robustness across various tumor types.

ResNet-50, with an accuracy of 84.9% and an F1-score of 0.85, followed closely behind, excelling

particularly in no_tumor and Meningioma detection. However, it slightly lagged behind ResNet-152 and DenseNet121, likely due to its shallower architecture.

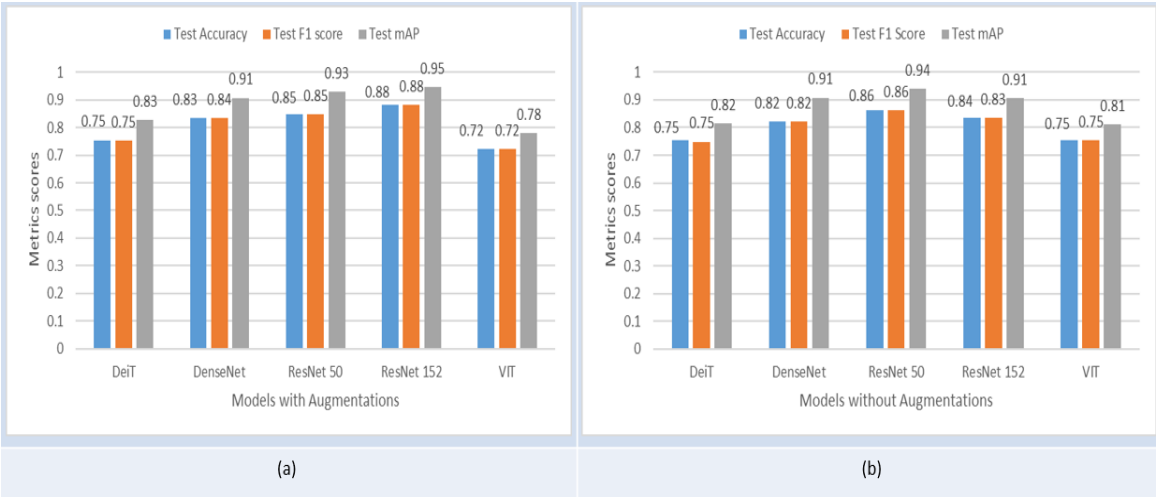


Figure 4.1.1 : Test set performance for deep learning models (DeiT, DenseNet, ResNet 50, ResNet 152, and ViT) (a) with augmentation and (b) without augmentation, highlighting the effects of augmentation on model performance and generalization.

The DeiT model, despite achieving a respectable overall performance with a 75.4% accuracy and a 0.75 F1-score, struggled with more complex tumor types like Pituitary, where it achieved an F1-score of 0.66. This suggests a trade-off between efficiency and accuracy, possibly due to its data-efficient design.

The Vision Transformer (ViT), with an accuracy of 72.1% and an F1-score of 0.72, underperformed, particularly in detecting Glioma and Pituitary tumors, with F1-scores of 0.67 and 0.63. ViT's reliance on large datasets for effective training likely contributed to its lower performance in this study, which used a relatively smaller dataset. While ViT has significant potential in large-data settings, its mAP score of 0.78 suggests that it may not be as effective in more limited data environments.

Tracking performance metrics such as accuracy and F1-score for the ResNet152 model over 200 epochs revealed that the model reached saturation around the 100th epoch. Beyond this point, the validation loss began to increase, indicating the onset of overfitting. For a visual representation of the training and validation loss curves of the ResNet152 model, refer to Appendix Figure 5.3.1.

These findings highlight the critical role of architectural strengths in determining model performance. ResNet-152 and DenseNet121 leveraged their deep, well-connected layers to deliver superior results, proving highly effective at extracting meaningful patterns from complex medical images. In contrast, DeiT and ViT, despite their innovative designs, faced limitations due to their architectural reliance on large-scale datasets. ViT's transformer architecture, which excels in other domains with ample data, struggled in this context, underscoring its sensitivity to data volume and quality. This comparison suggests that the robustness of traditional deep learning models remains a significant advantage in specialized medical imaging tasks like brain tumor classification.

Detailed performance metrics, including training and validation loss curves, ROC curves, classification reports, and confusion matrices for all classification models, are provided in Figures 4.1.2 to 4.1.6.

4.1.2 Model Performance without augmentation

To evaluate the impact of data augmentation on model performance, we conducted experiments across all classification models both with and without augmentation. The results showed that models without augmentation generally displayed slightly lower performance metrics compared to those with augmentation.

For instance, ResNet-152, which was the top-performing model with augmentation, achieved a reduced accuracy of 83.6% and an F1-score of 0.8349 without augmentation, compared to its performance with augmentation (accuracy of 88.3% and F1-score of 0.882). Similarly, DenseNet121 saw a decrease in performance, with accuracy dropping to 82.2% and an F1-score of 0.8218 without augmentation, down from 83.4% and 0.835, respectively.

ResNet-50, interestingly, performed better without augmentation, achieving an accuracy of 86.4% and an F1-score of 0.8633 without augmentation, compared to 84.9% accuracy and 0.8472 F1-score with augmentation. This suggests that in this case, the augmentation might have introduced variability that the model struggled to generalize effectively.

The DeiT model, which is designed for data efficiency, maintained a consistent accuracy of 75.4% with a slight drop in its F1-score to 0.7486 without augmentation, compared to 0.7522 with augmentation. However, its mAP decreased slightly from 0.828 with augmentation to 0.8159 without.

Interestingly, the ViT model, which typically relies on large datasets for optimal performance, showed an increase in accuracy without augmentation, rising to 75.4% from 72.1% with augmentation, and a slight improvement in its F1-score to 0.7537 from 0.7209. This counterintuitive result may suggest that the augmentations used might have introduced noise that negatively impacted ViT's performance.

Data augmentation generally enhances model performance, but its impact varies depending on the architecture. ResNet-152 and DenseNet121 benefit from augmentation but still perform well without it. However, ResNet-50 unexpectedly performed better without augmentation, likely due to overfitting, where the model learned features specific to the augmented data rather than generalizable patterns. If the data already contains sufficient variability, further augmentation can lead to worse results by introducing redundant or misleading patterns. Similarly, ViT saw a slight improvement without augmentation, suggesting that additional data manipulation may have introduced noise, reducing effectiveness. This indicates that augmentation strategies should be carefully tailored to the specific architecture to avoid negative impacts like overfitting or unnecessary variability.

Classification Analysis: The classification analysis reveals that careful consideration of model architecture, combined with high-quality data and appropriate augmentation techniques, is essential for developing reliable and effective brain tumor classification models. While ResNet-152 and DenseNet121 benefit from augmentation, models like ResNet-50 and ViT show varied responses, indicating that augmentation strategies should be tailored to specific architectures to avoid overfitting or unnecessary variability. This study's findings are based solely on the datasets used, and future work should focus on external validation and further optimization to ensure real-world applicability.

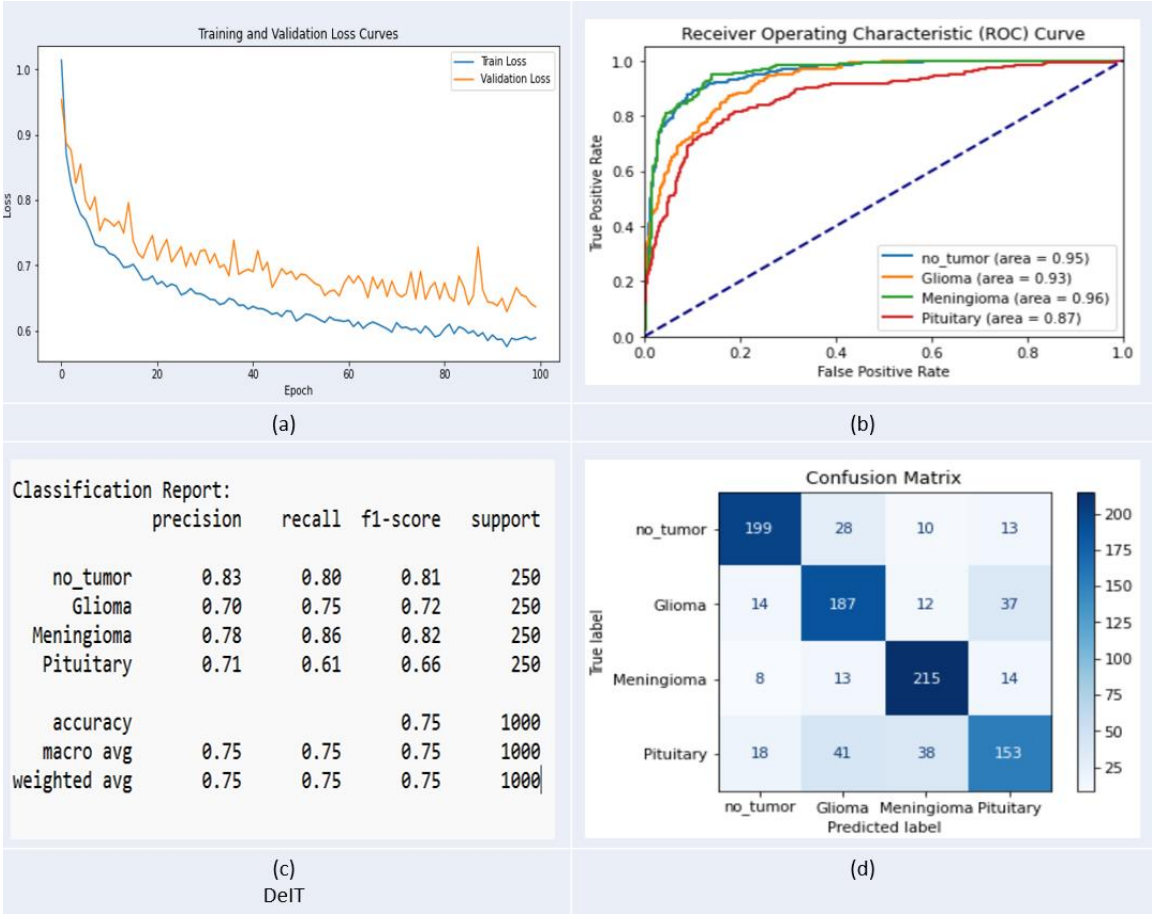
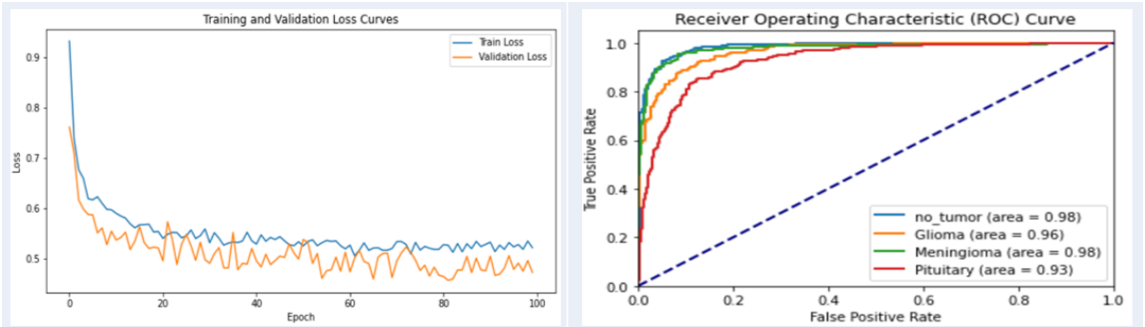
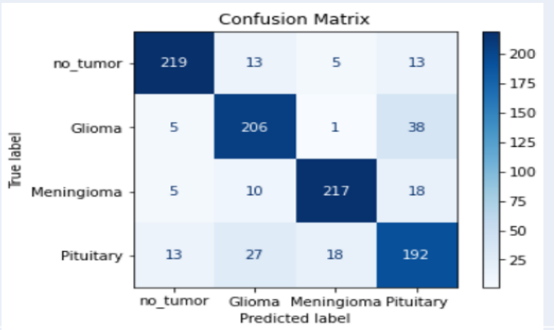


Figure 4.1.2 : Evaluation of the DeiT model's performance in the classification task: (a) Training and Validation Loss Curves over 100 epochs; (b) ROC Curve for each tumor type; (c) Classification Report summarizing precision, recall, F1-score, and support; (d) Confusion Matrix displaying the model's predictions across four tumor categories.



Classification Report:

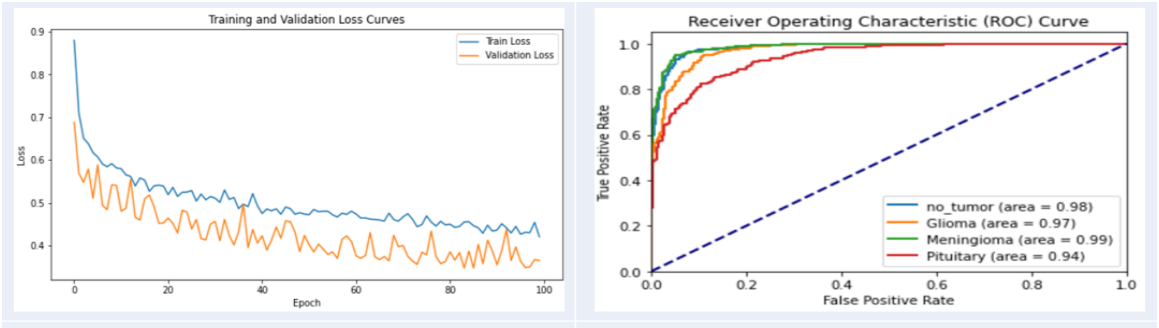
	precision	recall	f1-score	support
no_tumor	0.90	0.88	0.89	250
Glioma	0.80	0.82	0.81	250
Meningioma	0.90	0.87	0.88	250
Pituitary	0.74	0.77	0.75	250
accuracy			0.83	1000
macro avg	0.84	0.83	0.83	1000
weighted avg	0.84	0.83	0.83	1000



(c)
DENSENET121

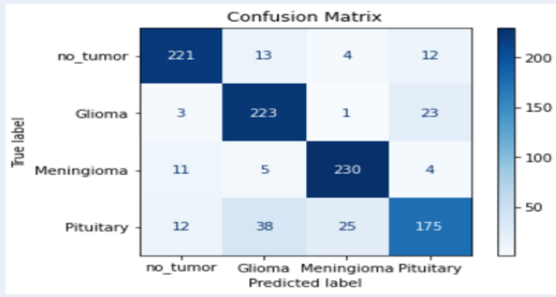
(d)

Figure 4.1.3 : Performance evaluation of the Dense-Net121 model on the classification task: (a) Training and Validation Loss Curves over 100 epochs; (b) ROC Curve for each tumor type; (c) Classification Report summarizing precision, recall, F1-score, and support; (d) Confusion Matrix showing model predictions across four tumor categories.



Classification Report:

	precision	recall	f1-score	support
no_tumor	0.89	0.88	0.89	250
Glioma	0.80	0.89	0.84	250
Meningioma	0.88	0.92	0.90	250
Pituitary	0.82	0.70	0.75	250
accuracy			0.85	1000
macro avg	0.85	0.85	0.85	1000
weighted avg	0.85	0.85	0.85	1000



(c)
Resnet 50

(d)

Figure 4.1.4 : Performance evaluation of the ResNet50 model on the classification task:(a) Training and Validation Loss Curves over 100 epochs; (b) ROC Curve for each tumor type; (c) Classification Report summarizing precision, recall, F1-score, and support; (d) Confusion Matrix showing model predictions across four tumor categories.

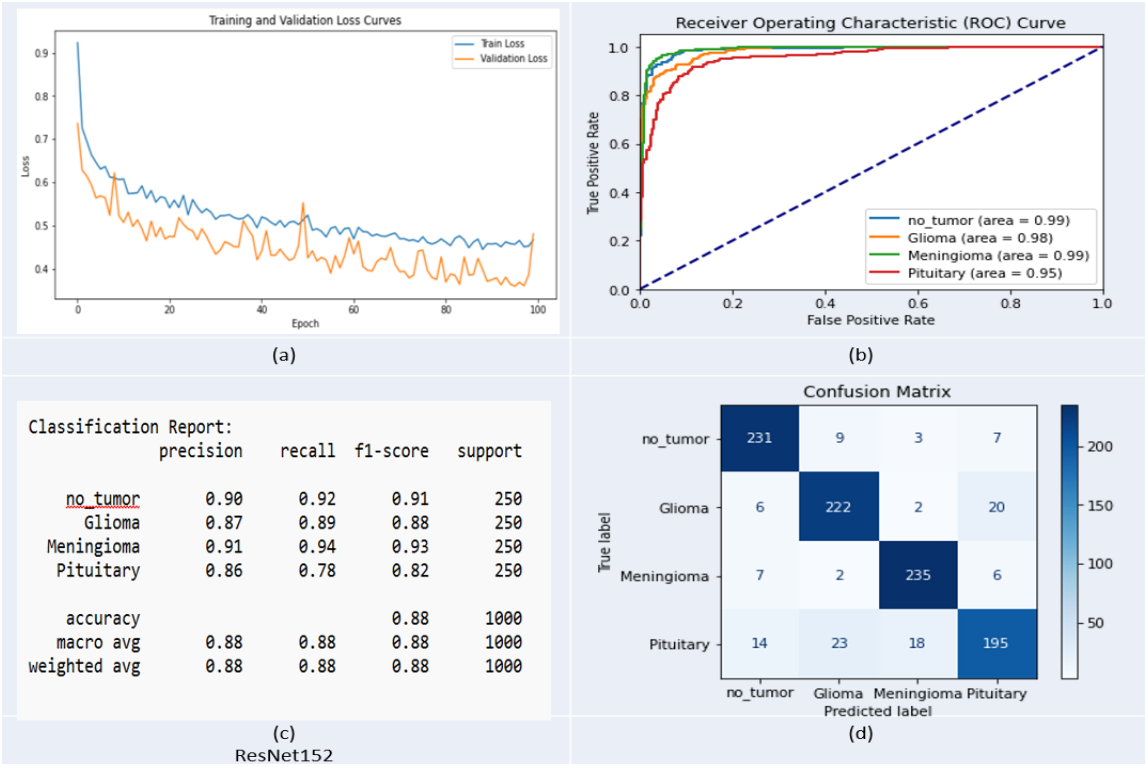


Figure 4.1.5 : Performance evaluation of the ResNet152 model on the classification task: (a) Training and Validation Loss Curves over 100 epochs; (b) ROC Curve for each tumor type; (c) Classification Report summarizing precision, recall, F1-score, and support; (d) Confusion Matrix showing model predictions across four tumor categories.

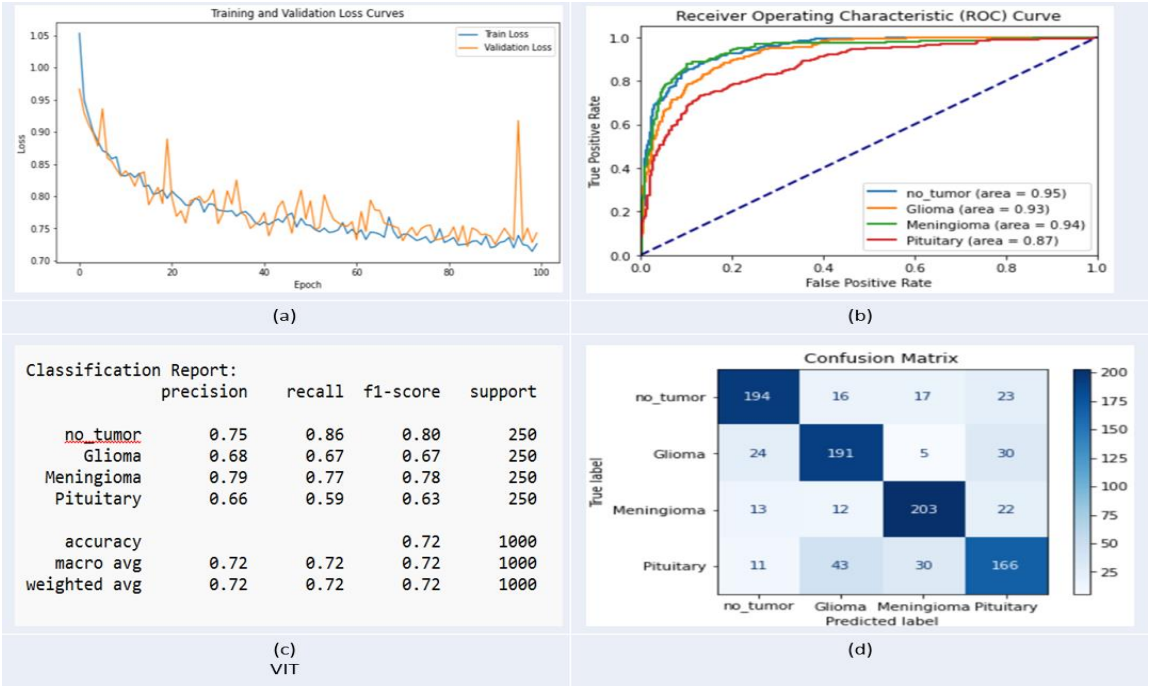


Figure 4.1.6 : Performance evaluation of the VIT model on the classification task: (a) Training and Validation Loss Curves over 100 epochs; (b) ROC Curve for each tumor type; (c) Classification Report summarizing precision, recall, F1-score, and support; (d) Confusion Matrix showing model predictions across four tumor categories.

4.2 Segmentation

The study evaluated the performance of several segmentation models, including DeepLabV3, MA-Net, U-Net, and U-Net++, all recognized for their effectiveness in segmenting brain tumors in MRI scans. Dataset 2, specifically designed for medical imaging, served as the basis for this evaluation. Performance was assessed using key metrics such as Dice Coefficient, Intersection over Union (IoU), and Pixel Accuracy, which are essential for determining how well these models identify and delineate tumor boundaries. Additionally, we examined the impact of data augmentation techniques, which played a significant role in enhancing the accuracy and reliability of the segmentation results.

4.2.1 Models Performance with Augmentation

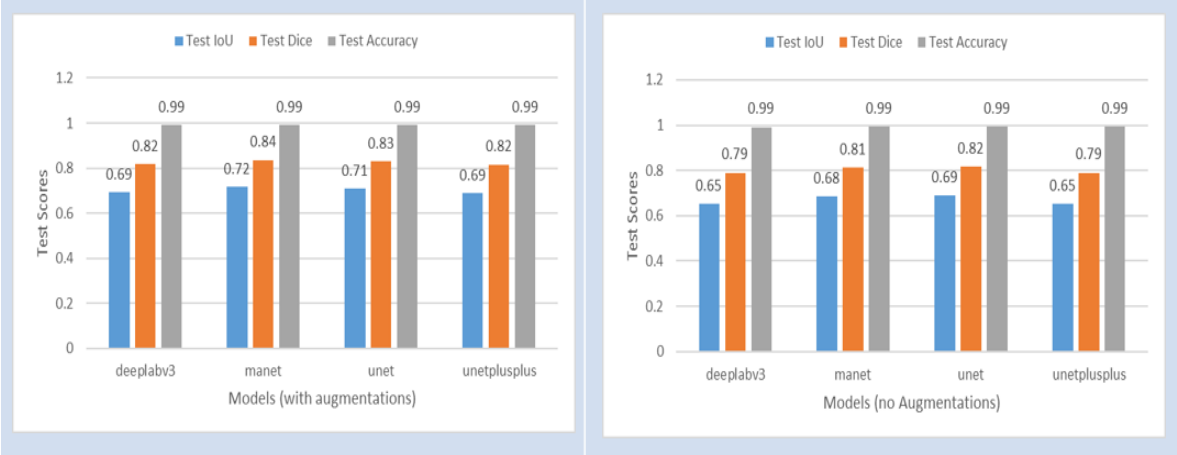


Figure 4.2.1 : Test set performance for DeepLabV3, MA-Net, U-Net, and U-Net++ models (a) with augmentation and (b) without augmentation, highlighting the effects of augmentation on model performance and generalization.

MA-Net emerged as the top performer, achieving a Test IoU of 0.7183 and a Dice Coefficient of 0.8361 with augmentation, alongside a Test Accuracy of 0.9917(As shown in Figure 4.2.1). The model's advanced attention mechanisms likely contributed to its ability to focus on relevant tumor regions, resulting in better segmentation, particularly in complex cases. U-Net also performed strongly, with a Test IoU of 0.708 and a Dice Coefficient of 0.8291, proving effective in segmenting tumors.

DeepLabV3 and U-Net++ delivered solid results, albeit with some variability. DeepLabV3, known for capturing multi-scale context, achieved a Test IoU of 0.6922 and a Dice Coefficient of 0.8181, along with a Test Accuracy of 0.9921. However, it was slightly less precise in segmenting smaller or less distinct tumors compared to MA-Net and U-Net. U-Net++, featuring a complex architecture with dense and nested skip pathways, recorded a Test IoU of 0.6887 and a Dice Coefficient of 0.8157, with a Test Accuracy of 0.9914. While its sophisticated design shows potential, it may have introduced challenges in generalization, particularly across different tumor types.

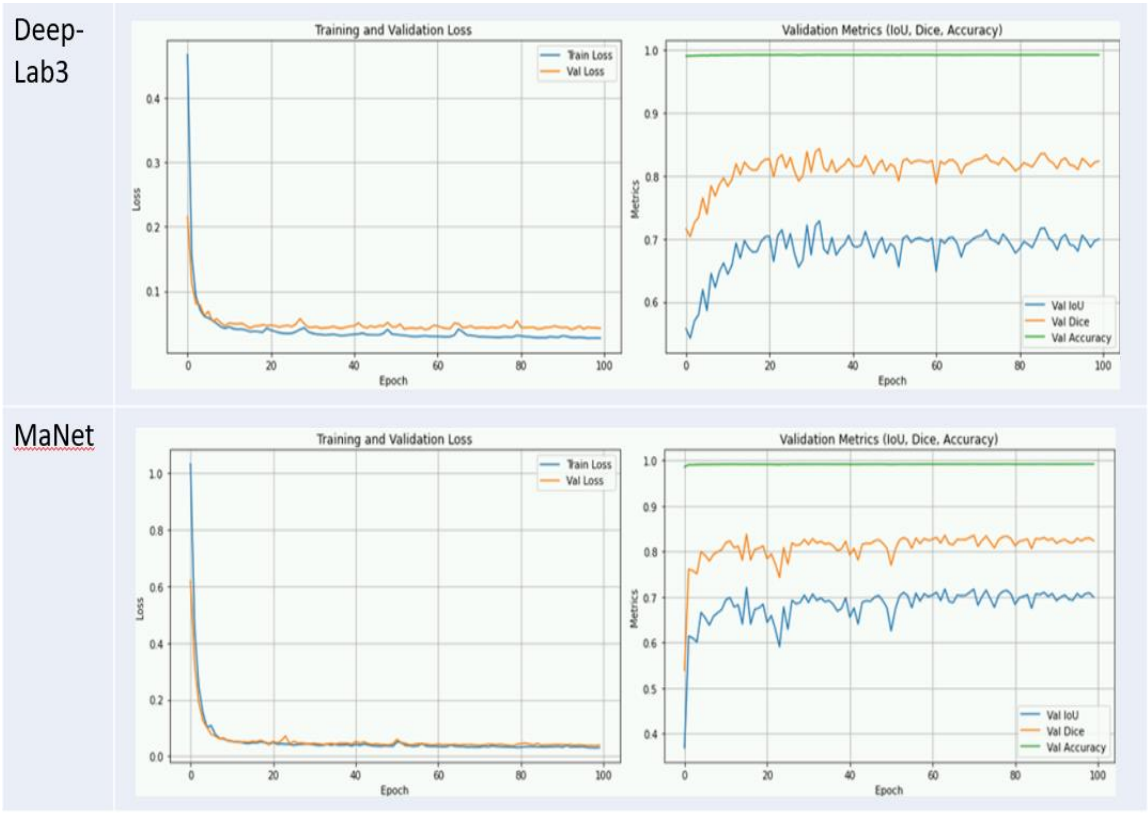
Tracking performance metrics such as IoU and Dice Coefficient for the MA-Net model over 200

epochs revealed that the model reached saturation around the 100th epoch. Beyond this point, no significant improvements were observed, indicating that the model had stabilized. Further training did not enhance performance and introduced the risk of overfitting. For a visual representation of the training and validation loss curves of the MA-Net model, refer to Appendix Figure 5. 3.2.

Detailed performance metrics, including training and validation loss curves, and validation metrics such as IoU, Dice Coefficient, and Pixel Accuracy for all segmentation models, are provided in Figures 4.2.2 to 4.2.3. Additionally, ground truth and predicted segmentation results for DeepLabV3, U-Net, and U-Net++ models are illustrated in Figure 4.2.3.

When comparing these models, MA-Net and U-Net consistently performed better, especially in terms of IoU and Dice Coefficient, which are critical metrics for assessing the quality of segmentation. The high Test Accuracy across all models suggests that these models are reliable, although the small size of tumors may inflate accuracy scores. However, MA-Net's advanced attention mechanisms and U-Net's proven architecture make them stand out as the most reliable models in this study. DeepLabV3 and U-Net++, while effective, showed more variability, particularly in handling smaller or less distinct tumors.

Overall, while all models performed well in brain tumor segmentation, MA-Net and U-Net were the most consistent and accurate. However, the lack of external validation highlights the challenges of real-world application and underscores the need for further research to validate these findings.



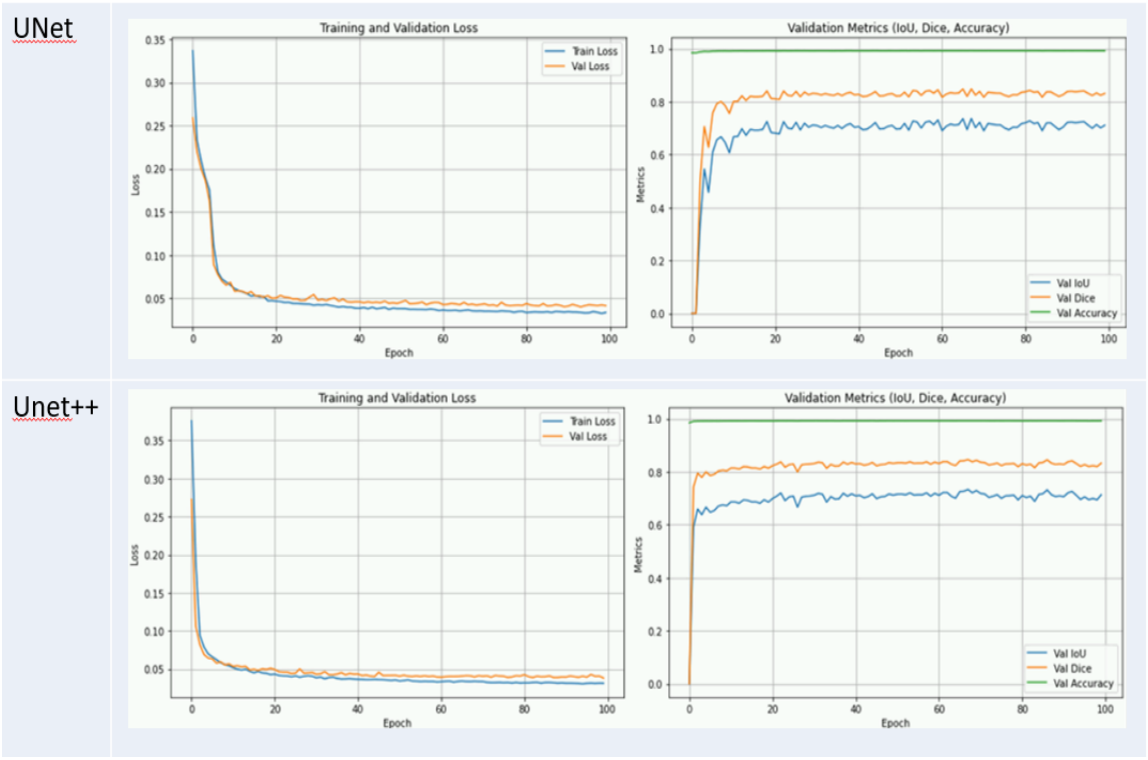


Figure 4.2.2 : Training and validation loss curves, with validation metrics (IoU, Dice, Accuracy) for DeepLabV3, MA-Net, U-Net, and U-Net++, showing performance and metric stabilization over 100 epochs.

4.2.2 Model Performance without augmentation.

In this study, the performance of segmentation models was evaluated without applying data augmentation, which resulted in a noticeable decline in key metrics across all models.

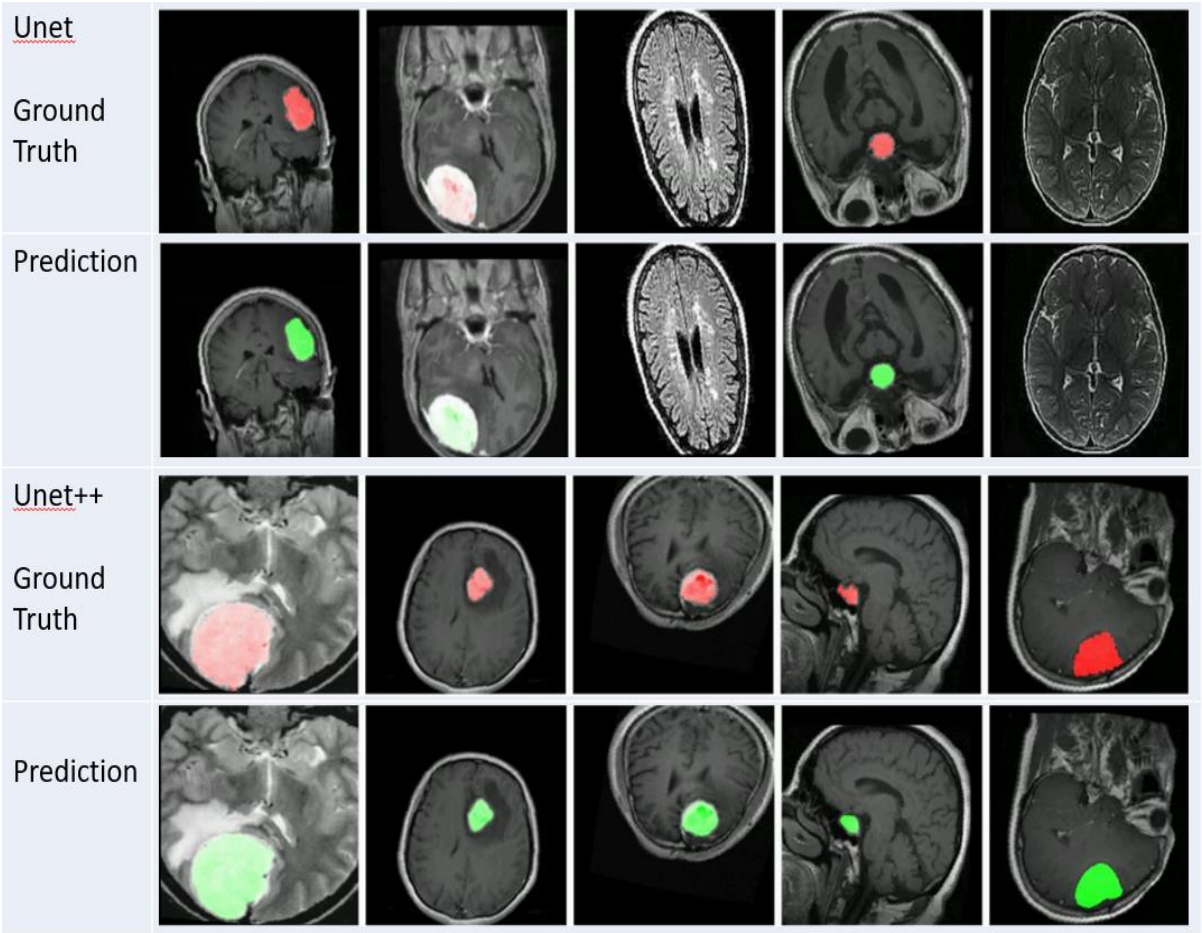
For instance, MA-Net, which had the highest performance with augmentation, saw its Test IoU drop from 0.7183 to 0.6845 without augmentation, and its Dice Coefficient decreased from 0.8361 to 0.8127. Similarly, DeepLabV3 experienced a decline in IoU from 0.6922 to 0.6523, indicating that its ability to accurately delineate tumor boundaries was reduced when augmentation was not applied.

U-Net also showed a reduction in performance, with its IoU falling from 0.708 to 0.6891. U-Net++, which already had a lower baseline compared to the other models, saw its IoU decrease from 0.6887 to 0.6523, further underscoring its reliance on augmentation to improve generalization and accuracy.

These results highlight the critical role of data augmentation in enhancing the models' ability to generalize and accurately segment tumors. The consistent decline across all models suggests that without augmentation, the models struggled to maintain the same level of performance, likely due to reduced exposure to varied training data that augmentation provides. This emphasizes the importance of incorporating data augmentation techniques, particularly in medical imaging tasks, to ensure robust and accurate segmentation results across different tumor types and imaging conditions.

Segmentation Analysis: MA-Net and U-Net emerged as the most reliable models for accurately identifying tumor boundaries, outperforming DeepLabV3 and U-Net++. MA-Net's attention

mechanisms likely enhanced its ability to focus on relevant tumor regions, while U-Net's robust architecture contributed to its strong performance. In contrast, U-Net++ and DeepLabV3 showed lower precision, particularly in segmenting smaller or less distinct tumors. Data augmentation played a crucial role, particularly for U-Net++, which saw significant improvements with its application, highlighting its reliance on augmentation for better generalization and accuracy. These findings emphasize the effectiveness of simpler models like MA-Net and U-Net and the importance of augmentation in optimizing model performance.



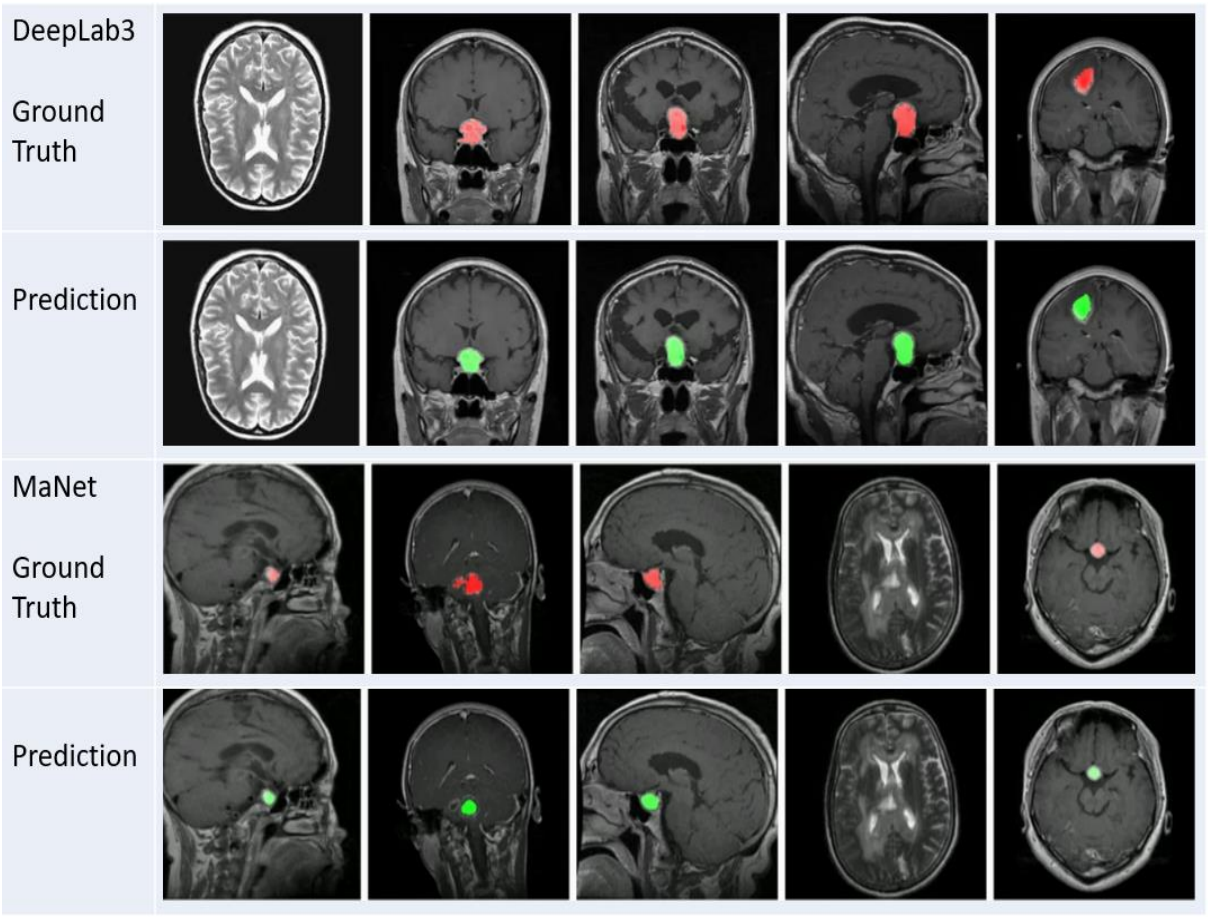


Figure 4.2.3 : Ground truth and predicted segmentation results for brain tumors using DeepLabV3, U-Net, and U-Net++ models. The top images display the actual tumor boundaries (Ground Truth), while the bottom images show the corresponding predictions made by each model.

4.3 Detection

This section provides an overview of the datasets and models used to assess brain tumor detection. Dataset 2 involved converting segmentation masks into bounding boxes, while Dataset 3 had an imbalanced distribution of tumor types. The key metrics utilized included Mean Average Precision (mAP), Intersection over Union (IoU), and F1 Score. The analysis evaluates the performance of Faster R-CNN and RetinaNet models, discussing how class imbalance was managed and emphasizing the role of data augmentation and model adjustments in enhancing detection accuracy and robustness.

4.3.1 Experiment with dataset 3

In the initial attempt using Dataset 3, both the Faster R-CNN and RetinaNet models were utilized to detect brain tumors in MRI scans, yielding varied outcomes. Faster R-CNN attained a Mean Average Precision (mAP) of 0.7134, an Intersection over Union (IoU) of 0.7482, and an F1 Score of 0.8823. In contrast, RetinaNet produced slightly lower metrics, with an mAP of 0.6678, an IoU of 0.7996, and an F1 Score of 0.9211. Despite these results, both models faced significant challenges in

accurately detecting tumors. A critical issue was their inability to generate precise bounding boxes for many images, leading to frequent missed detections or incorrect classifications. This problem was compounded by the pronounced class imbalance within Dataset 3, where the models tended to overfit to the more prevalent Glioma tumor class, while underperforming on less frequent classes like Pituitary and Meningioma. This overfitting skewed the detection metrics, heavily favoring the majority class, and raised concerns about the models' robustness and generalizability. These findings underscore the necessity for more balanced training, validation, and testing data to enhance model performance and ensure accurate detection across all tumor types.

4.3.2 Experiment with dataset 2

To address the class imbalance issues found in Dataset 3, a strategic shift to Dataset 2 was necessary. Previously used for segmentation tasks, Dataset 2 provided a more balanced and robust data source, making it more appropriate for detection tasks. The conversion of segmentation masks into bounding boxes further optimized Dataset 2 for this purpose, leading to a significant improvement in performance metrics and enhanced model accuracy. After these refinements, Dataset 3 was employed for external validation, confirming the models' improved generalizability and robustness across different datasets.

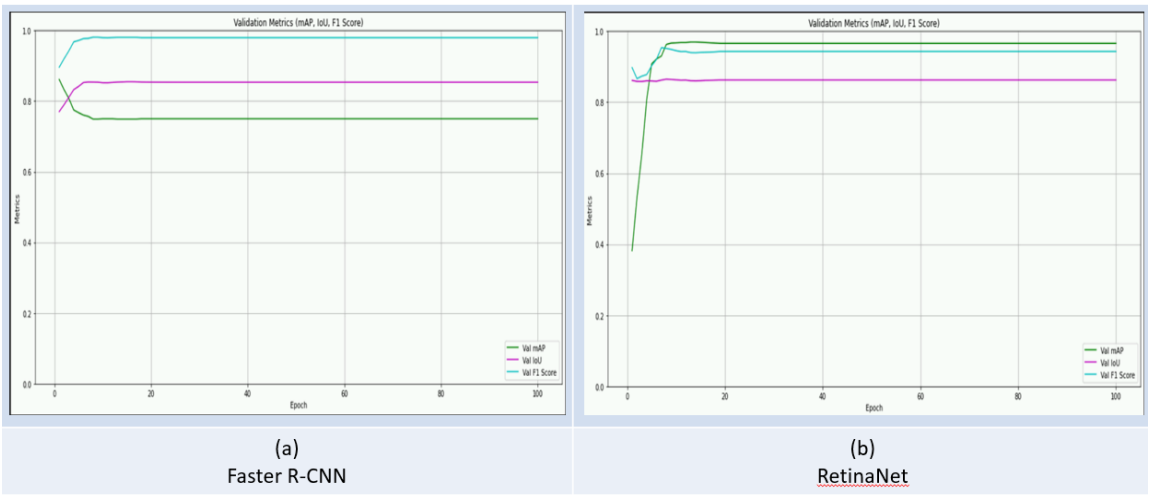


Figure 4.3.1 : Metrics including mAP, IoU, and F1 Score across 100 epochs are depicted: (a) Trends for Faster R-CNN; (b) Evolutionary performance for RetinaNet.

4.3.2.1 Model Performance with Augmentation

Faster R-CNN recorded a Mean Average Precision (mAP) of 0.77, an Intersection over Union (IoU) of 0.85, and an F1 Score of 0.97 on the test set (as illustrated in Figure 4.3.2). Despite these strong overall metrics, Faster R-CNN struggled with generating accurate bounding boxes and correctly classifying smaller or less distinct tumors, such as Gliomas. This difficulty likely arises from its reliance on region proposal mechanisms, which can underperform when dealing with ambiguous or subtle

tumor boundaries.

Figure 4.3.2 : Test set performance metrics (mAP, IoU, and F1 Score) for Faster R-CNN and RetinaNet models with and without augmentation on Dataset 2. (b) Corresponding metrics during external validation on Dataset 3, illustrating the impact of augmentation on model generalization.

In contrast, RetinaNet, with the application of augmentation, achieved slightly higher precision with an mAP of 0.95, though it had a comparable IoU of 0.85 and an F1 Score of 0.96. RetinaNet's use of a focal loss function allowed it to maintain better consistency in challenging cases where Faster R-CNN struggled. Furthermore, RetinaNet was more effective in handling class imbalance and was able to generate accurate bounding boxes across all tumor types, making it more reliable in scenarios requiring high detection accuracy.

In the external validation phase, Faster R-CNN attained a mean average precision (mAP) of 0.58, an Intersection over Union (IoU) of 0.68, and an F1 Score of 0.87. In contrast, RetinaNet outperformed Faster R-CNN, achieving a mAP of 0.66, an IoU of 0.71, and an F1 Score of 0.89. These results highlight RetinaNet's superior ability to maintain consistency and accuracy in more challenging scenarios, particularly in managing class imbalance more effectively than Faster R-CNN. This indicates that RetinaNet may be more robust and reliable for real-world applications.

For both Faster R-CNN and RetinaNet, performance metrics such as mAP, IoU, and F1 Score stabilized between the 10th and 15th epochs. Although the models were trained for 100 epochs to maintain consistency with other tasks, no significant improvements were observed beyond the early epochs. Detailed performance metrics, including training and validation loss curves, along with validation metrics like mAP, IoU, and F1 Score for both detection models, are shown in Figures 4.3.1 and 4.3.2. Additionally, detection results featuring MRI scans with indicated confidence scores are presented in Figure 4.3.3.

When comparing the two models, RetinaNet consistently outperformed Faster R-CNN in both the test set and external validation. While Faster R-CNN showed strong initial metrics, its reliance on region proposal mechanisms made it less effective in handling smaller or ambiguous tumors, which are critical in accurate brain tumor detection. In contrast, RetinaNet's focal loss function and better

handling of class imbalance allowed it to maintain higher precision and reliability across different tumor types, making it a more robust choice for scenarios requiring high detection accuracy.

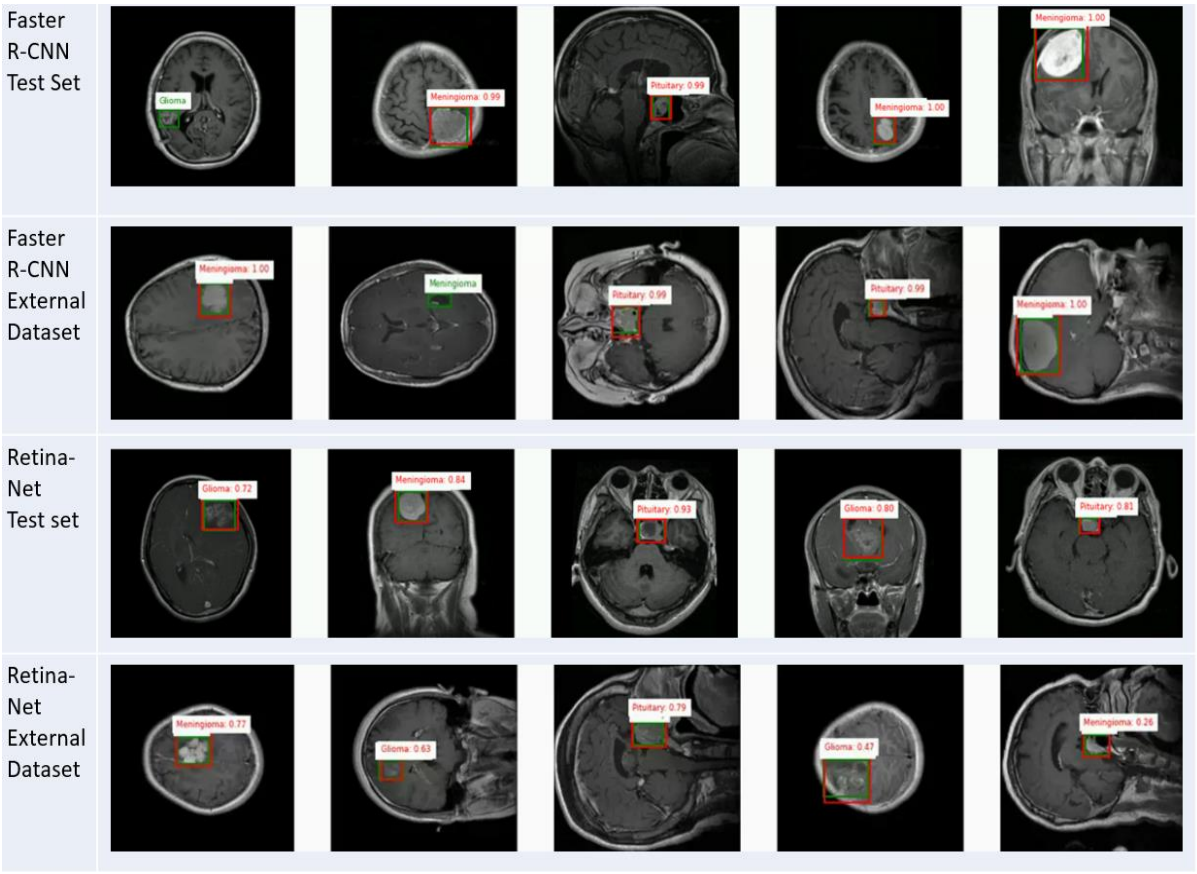


Figure 4.3.3 : Detection results from Faster R-CNN and RetinaNet on test and external datasets, showcasing MRI scans of tumors (Glioma, Meningioma, Pituitary) with marked confidence scores.

4.3.2.2 Model performance without Augmentation

To better assess the impact of data augmentation, the performance of Faster R-CNN and RetinaNet was evaluated without applying augmentation. The findings indicated that data augmentation generally enhanced detection accuracy for both models, though the degree of improvement varied.

For Faster R-CNN, without augmentation, the mAP was slightly lower at 0.7472, and the IoU was 0.87, with an F1 Score of 0.9672. With augmentation, these metrics improved, indicating more accurate and reliable detection performance. RetinaNet, without augmentation, showed a mAP of 0.9509, an IoU of 0.8545, and an F1 Score of 0.9425. After applying augmentation, the mAP slightly decreased to 0.9472, and the IoU declined slightly to 0.8469, though the F1 Score increased to 0.9555. These results suggest that while RetinaNet was more stable, it did not gain as much benefit from data augmentation compared to Faster R-CNN.

The impact of data augmentation was evident, particularly for Faster R-CNN, which showed more substantial improvements with its application. This pattern continued when the models were tested on the external validation set, where augmented models displayed better generalization and

robustness, confirming the effectiveness of data augmentation across different scenarios.

Detection Analysis: RetinaNet consistently outperformed Faster R-CNN, especially in handling class imbalance and achieving higher precision. The initial struggles with Dataset 3 highlighted the need for balanced training data. Transitioning to Dataset 2 and applying data augmentation significantly improved detection accuracy, with RetinaNet showing the most benefit. Faster R-CNN, while effective, was less robust, particularly in external validation. These results emphasize RetinaNet's superior ability to generalize and the importance of balanced data and augmentation in optimizing detection models.

4.4 Integrated Analysis

The integrated analysis of classification, segmentation, and detection tasks highlights the critical interplay between model architecture, data quality, and augmentation techniques in optimizing performance. ResNet-152 excelled in classification, laying a strong foundation for subsequent tasks. MA-Net utilized these classification insights to enhance tumor boundary identification in segmentation, which directly contributed to improved detection accuracy in models like RetinaNet. The consistent use of Dataset 2 across both segmentation and detection further underscores the value of high-quality, consistent data.

Detection, while performed as a separate task, inherently involves classification through the identification and classification of bounding boxes. RetinaNet's robust architecture, coupled with effective data augmentation, allowed it to handle class imbalance and achieve high precision in generating bounding boxes. The importance of external validation, verified in the detection section, reinforced the models' generalizability and robustness across diverse datasets, confirming their applicability in real-world scenarios.

Although these tasks were conducted independently, their interrelationships are clear: precise classification supports effective segmentation, which in turn boosts detection accuracy by providing detailed tumor boundaries. This interconnectedness emphasizes the need for a holistic approach where advancements in one area support improvements in others.

Data augmentation emerged as a significant enhancer across all tasks, especially in contexts of limited or imbalanced datasets, contributing to the robustness and generalization of the models.

In conclusion, while classification, segmentation, and detection were conducted separately, their outcomes are deeply interconnected. Achieving optimal performance in brain tumor diagnosis and treatment depends on a comprehensive approach that combines high-quality data, customized model architectures, effective augmentation strategies, and external validation. This ensures that deep learning models are accurate, robust, and generalizable, making them reliable for clinical applications.

5 CONCLUSION

This final chapter provides an in-depth summary of the study's key findings, emphasizing the accomplishments and insights derived from evaluating various deep learning models for brain tumor classification, segmentation, and detection. It also discusses the broader implications of these results for the field of medical imaging, critically examines the limitations encountered during the research, and offers suggestions for future work to further develop and refine the findings presented in this dissertation.

5.1 Summary of Key Findings

This study conducted a thorough evaluation of various deep learning models for brain tumor classification, segmentation, and detection. The results underscored both the strengths and limitations of each model, with particular emphasis on how factors such as model architecture, data quality, and augmentation techniques influenced their performance.

In classification, ResNet-152 proved to be the most effective model, exhibiting superior accuracy and precision, particularly in distinguishing between tumor types such as Meningioma and non-tumor categories. The model’s deep architecture and effective use of residual connections were crucial to its strong performance. DenseNet121 also performed well, though it slightly lagged behind ResNet-152, while newer models like the Vision Transformer (ViT) encountered challenges, likely due to their reliance on large datasets and sensitivity to noise in MRI images.

For segmentation, MA-Net, with its integration of transformer elements, emerged as the most reliable model, effectively identifying and delineating tumor boundaries. U-Net also delivered strong results, though MA-Net's advanced attention mechanisms provided it with an edge in handling more complex cases. Data augmentation significantly enhanced the models' ability to generalize across different tumor types, particularly benefiting segmentation tasks.

In detection, RetinaNet excelled, particularly in managing class imbalance and generating precise bounding boxes. The use of focal loss in RetinaNet enabled it to maintain high performance even in challenging scenarios where other models, like Faster R-CNN, faced difficulties. The study further highlighted the critical importance of balanced datasets in detection tasks, as class imbalance often led to skewed results and adversely affected model generalization.

Overall, the findings from this study highlight the importance of selecting model architectures that are well-suited to specific tasks in brain tumor analysis. The results also stress the essential role of high-quality, balanced datasets and the application of data augmentation techniques in improving model performance. These elements are vital for increasing the robustness, accuracy, and reliability

of deep learning models in clinical settings, especially for brain tumor diagnosis and treatment.

5.2 Evaluation

The evaluation of this study highlights significant success in achieving its main goals, which focused on rigorously assessing deep learning models for brain tumor classification, segmentation, and detection. The models were extensively tested using standard evaluation metrics, validating their effectiveness across all tasks. This accomplishment reflects the meticulous selection and fine-tuning of model architectures to tackle the distinct challenges associated with each specific task

One of the critical insights gained from this study is the significance of data quality and preprocessing in influencing model performance. The application of data augmentation techniques proved particularly effective in enhancing the robustness of models, especially in segmentation and detection tasks. These findings suggest that future research should place a strong emphasis on refining data handling strategies to further improve model outcomes. The results also indicate that while advanced model architectures are essential, the integration of high-quality data and preprocessing techniques is equally crucial in achieving optimal performance.

The implications of these findings go beyond the immediate results, providing significant contributions to the field of medical imaging. The study underscores the potential for integrating these deep learning models into clinical workflows, offering tools that could improve diagnostic accuracy and efficiency. However, it also emphasizes the need for ongoing refinement and external validation, particularly within real-world clinical settings where variability in data and operational conditions may impact model performance. This underscores the importance of collaboration between researchers, clinicians, and industry partners to ensure that these models transition effectively from research to practical clinical use.

Overall, the study has effectively achieved its goals by developing and assessing models that make notable advancements in brain tumor analysis. The findings establish a strong basis for future research and provide practical insights into the key factors impacting the effectiveness of deep learning models in medical imaging. This research paves the way for future innovations and highlights the importance of a comprehensive approach that harmonizes model architecture with data quality and clinical relevance.

5.3 Future Work

While the study met its primary objectives, it faced several challenges. A major limitation was the use of relatively small datasets and the lack of external validations by field experts, which might limit the models' reflection of the diversity and complexity encountered in clinical brain tumor cases. This affects how well the models generalize in practical applications. Furthermore, the scarcity of large, high-quality datasets presented a major challenge, as the models' performance was heavily dependent

on the quantity and quality of the available data. Resource constraints, particularly limited GPU power, also restricted the exploration of larger or more complex models, potentially affecting optimal model performance.

Future research should aim to enhance dataset diversity and size to improve model robustness and generalizability. It's crucial to include evaluations by field experts and validate models with real-world clinical data to confirm their effectiveness in clinical environments. Additionally, investigating the potential of training models from scratch, rather than relying solely on pretrained models, could offer insights into unique model adaptations specific to brain tumor detection. Future studies could also benefit from employing multi-task learning strategies, where models are developed to handle classification, segmentation, and detection tasks concurrently. This approach could optimize the learning process, enhance model performance across tasks, and increase computational efficiency, crucial for real-time applications in clinical settings. Overcoming GPU resource limitations will require more efficient architectures or increased computational capacity. Exploring 3D datasets could advance spatial analysis of tumors. Furthermore, adopting self-supervised and semi-supervised learning techniques might make better use of available data and reduce dependence on large labeled datasets. Advanced strategies for managing class imbalance, such as sophisticated data augmentation or balancing techniques, are also recommended to boost detection reliability across different tumor types.

Enhancing these models and integrating them into clinical workflows could significantly boost the accuracy and efficiency of brain tumor diagnosis, providing valuable support to radiologists and advancing medical imaging technology.

REFERENCES

[1] M. Sajjad, S. Khan, M. Ali, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *Journal of Computational Science*, vol. 30, pp. 174-182, 2019. Available: https://doi.org/10.1016/j.jocs.2018.12.003.

[2] P. Ghosal and N. Laskar, "Brain Tumor Classification Using ResNet-101 Based Squeeze and Excitation Deep Neural Network," International Journal of Imaging Systems and Technology, 2019.

[3] E. Şahin and Ö. Demir, "Multi-objective optimization of ViT architecture for efficient brain tumor classification," Biomed. Signal Process. Control, 2024. Available: https://doi.org/10.1016/j.bspc.2023.105938.

[4] M. Agarwal and G. Rai, "Deep learning for enhanced brain Tumor Detection and classification," Results in Engineering, 2024. Available: https://doi.org/10.1016/j.rineng.2024.102117.

[5] K. V. Shiny, "Brain tumor segmentation and classification using optimized U-Net," The Imaging Science Journal, vol. 72, no. 2, pp. 204-219, 2023. Available: https://doi.org/10.1080/13682199.2023.2200614.

[6] A. Akter and N. Nahar, "Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor," Expert Systems with Applications, 2023. Available: https://doi.org/10.1016/j.eswa.2023.122347.

[7] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks," in Proc. Med. Image Understanding Anal. Conf., Springer, Cham, 2017, pp. 506-517.

[8] Z. Zhou and M. M. R. Siddiquee, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, Cham, 2018, pp. 3-11.

[9] Z. Liu, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "CANet: Context Aware Network for Brain Glioma Segmentation," IEEE Trans. Med. Imag., vol. 40, no. 7, pp. 1763-1777, 2021. doi: 10.1109/TMI.2021.3065918.

[10] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, and Z. Gao, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," arXiv preprint, arXiv:2102.04306, 2021.

[11] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge," in Proc. Int. MICCAI Brainlesion Workshop, Springer, Cham, 2018, pp. 287-297.

[12] X. Chen, Y. Xu, D. W. K. Wong, and T. Y. Wong, "Region-adaptive Deformable U-Net for Glioma Segmentation from Multimodal MRI," Med. Image Anal., vol. 67, 2021. doi: 101837.

[13] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical Transformer: Gated

Axial-Attention for Medical Image Segmentation," arXiv preprint, arXiv:2102.10662, 2021.

[14] M. F. Almufareh, M. Imran, A. Khan, M. Humayun, and M. Asim, "Automated Brain Tumor Segmentation and Classification in MRI Using YOLO-Based Deep Learning," IEEE Access, vol. 12, pp. 16189-16207, 2024. doi: 10.1109/ACCESS.2024.3359418.

[15] M. Kordemir, K. K. Cevik, and A. Bozkurt, "A mask R-CNN approach for detection and classification of brain tumours from MR images," Comput. Methods Biomech. Biomed. Eng. Imaging Vis., vol. 11, no. 7, 2024. Available: https://doi.org/10.1080/21681163.2023.2301391.

[16] N. Iriawan and A. Anggraeni, "YOLO-UNet Architecture for Detecting and Segmenting the Localized MRI Brain Tumor Image," Comput. Math. Methods Med., 2024. Available: https://doi.org/10.1155/2024/3819801.

[17] S. S. Mathivanan, "Employing deep learning and transfer learning for accurate brain tumor detection," Sci. Rep., vol. 14, p. 7232, 2024. Available: https://doi.org/10.1038/s41598-024-57970-7.

[18] B. H. Menze, et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Trans. Med. Imaging, 2015.

[19] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, et al., "Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," Scientific Data, vol. 4, no. 1, pp. 1-13, 2017.

[20] S. Bakas, et al., "Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," Scientific Data, 2017.

[21] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045-1057, 2013. doi: 10.1007/s10278-013-9622-7.

[22] A. Dosovitskiy, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint, arXiv:2010.11929, 2020.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770-778.

[24] J. S. Babu, "ResNet50 from Scratch using TensorFlow," GitHub repository, 2020. Available: https://github.com/JananiSBabu/ResNet50_From_Scratch_Tensorflow.

[25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 4700-4708.

[26] A. Jain, "Creating DenseNet-121 with TensorFlow," Towards Data Science, Jun. 11, 2021. Available: https://towardsdatascience.com/creating-densenet-121-with-tensorflow-edbc08a956d8

[27] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training Data-Efficient Image Transformers & Distillation through Attention," in Proceedings of the 38th

International Conference on Machine Learning, vol. 139, 2021, pp. 10347-10357.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., Springer, Cham, 2015, pp. 234-241.

[29] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," arXiv preprint, arXiv:1706.05587, 2017.

[30] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, Cham, 2018, pp. 3-11.

[31] J. Zhang, L. Liu, X. Yang, X. Dai, D. Meng, and Y. Tang, "MaNet: Multi-Attention Network for Semantic Segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2019.

[32] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 2980-2988.

[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in Adv. Neural Inf. Process. Syst., 2015, pp. 91-99.

[34] Brain MRI Dataset. Available: https://www.kaggle.com/datasets/pradeep2665/brain-mri

[35] Brain Tumor Segmentation Dataset. Available: https://www.kaggle.com/datasets/atikaakter11/brain-tumor-segmentation-dataset

[36] Brain Tumor YOLO Dataset. Available: https://www.kaggle.com/datasets/abhit007pandey/brain-tumor-yolo

[37] R. Rana, "Dissertation," GitHub repository, [Online]. Available: https://github.com/Rajnish-rana/Classification--Segmentation-and-Detection-of-Brain-abnormalities-in-MRI-images

APPENDIX

Code Implementation-- refer to R. Rana's GitHub repository [https://github.com/Rajnishrana/Classification--Segmentation-and-Detection-of-Brain-abnormalities-in-MRI-images].

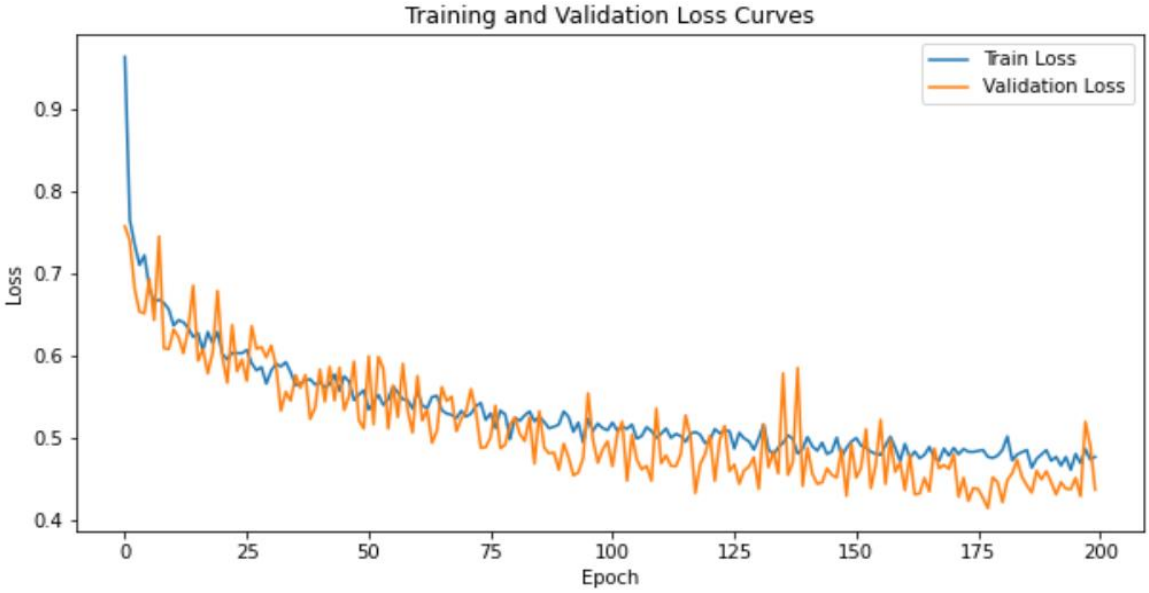


Figure 5.3.1 : Performance evaluation of the ResNet152 model on the classification task: Training and validation loss curves over 200 epochs.

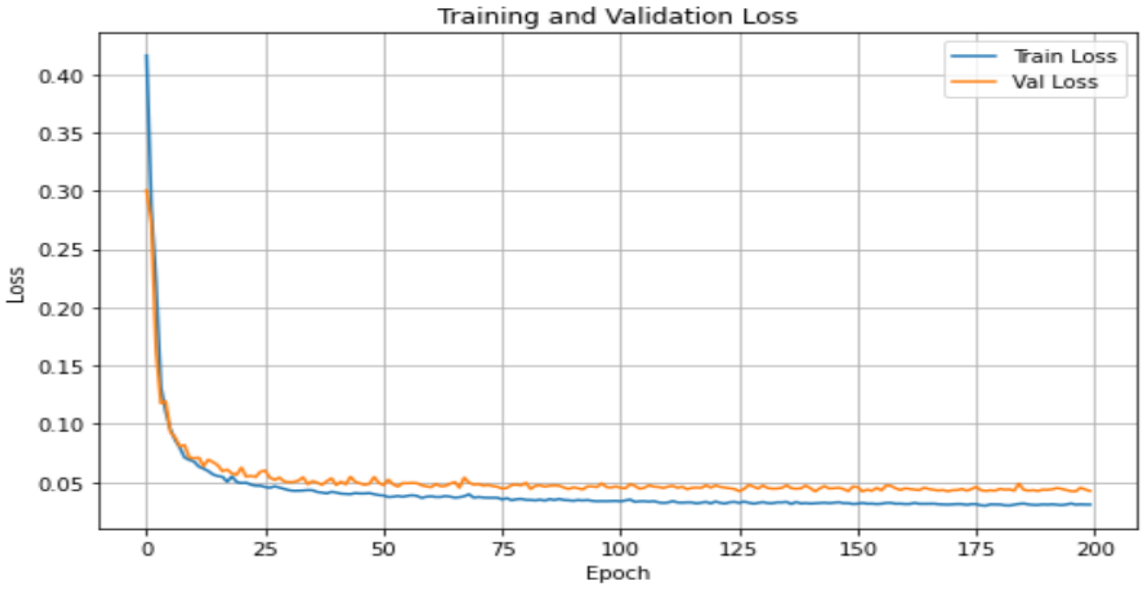


Figure 5.3.2 : Performance evaluation of the MA-Net model on the segmentation task: Training and validation loss curves over 200 epochs.