CSC3109 Machine Learning Project Report

Woon Jun Wei (2200624)[1], Benjamin Loh Choon How (2201590)[1], Low Hong Sheng Jovian (2203654)[1], Ong Zi Xuan Max (2200717)[1], and Cleon Tay Shi Hong (2200649)[1]

[1]Team 19

June 9, 2024

Contents

1 Introduction

Machine learning (ML) has become a transformative technology, driving innovation across numerous fields such as healthcare[1], healthcare research [2] and climate science[3]. Within ML, deep learning has emerged as a particularly powerful subset, enabling machines to perform tasks that require human-like perception and decision-making. This project aims to harness deep learning techniques to tackle real-world challenges in healthcare, specifically focusing on the classification of brain MRI images. By applying deep learning methodologies to this task, the project bridges theoretical concepts with practical applications, thereby advancing our understanding and capabilities in advanced ML techniques.

Brain magnetic resonance imaging (MRI) is a non-invasive imaging technique that plays a crucial role in diagnosing and monitoring various brain conditions, including tumors. The interpretation of brain MRI images is a complex and time-consuming process, requiring expert knowledge and experience. Accurate classification of brain MRI images is essential for timely diagnosis and treatment planning, as different types of brain tumors have distinct characteristics and treatment strategies. Patients may receive imaging studies as part of their routine care or to investigate specific symptoms, such as headaches, seizures, or cognitive changes [4]. In many cases, the radiologist's report on the MRI findings is critical for guiding clinical decisions and patient management.

Accurate classification of brain MRI images is a significant challenge in the medical field. Brain MRI scans are essential for diagnosing and monitoring various neurological conditions, yet the complexity and variability of these images make classification difficult. Misclassifications can lead to delays in diagnosis and treatment, impacting patient outcomes[5]. Thus, there is a critical need for robust classification models that can reliably distinguish between different types of brain tumors.

As the availability of MRI data increases, the demand for automated tools to assist radiologists in interpreting these images grows. Current manual methods are time-consuming and prone to human error, highlighting the need for efficient, automated solutions [6]. This project seeks to address this problem by developing, evaluating, and comparing multiple deep learning models to achieve high accuracy in classifying brain MRI images. By leveraging deep learning techniques, the project aims to reduce the workload on healthcare professionals, enhance diagnostic efficiency, and improve patient care through early detection and timely medical interventions.

The overarching objective of this project is to develop deep learning models that aid radiologists and doctors in the efficient diagnosis of brain tumors, facilitating early intervention and improving patient outcomes. By applying and extending advanced deep learning techniques, this project aims to produce models capable of accurately classifying brain MRI images. These models will support healthcare professionals by reducing diagnostic time and enhancing the accuracy of diagnoses, ultimately contributing to more timely and effective medical interventions.

2 Literature Review

2.1 Existing Approaches to Brain MRI Classification Using Deep Learning

The integration of deep learning into the medical field has revolutionized diagnostic methodologies, particularly in the classification of brain MRI images. As brain tumors present diverse morphological characteristics that are detectable through MRI, the precision of image analysis can significantly influence diagnostic and treatment outcomes. This literature review examines current deep learning approaches that enhance the accuracy and efficiency of brain tumor diagnoses from MRI scans, focusing on the evolution of model architectures, data preprocessing innovations, and the strategic application of transfer learning.

2.1.1 Data preprocessing techniques

Data preprocessing is a crucial step in enhancing the accuracy of brain MRI classification models. Techniques such as data augmentation [7] and cropping along contours of brain MRIs play significant roles in improving model training and performance. Data augmentation involves creating modified versions of the original images through transformations such as rotation, scaling, and flipping. This process helps in increasing the diversity of the training dataset, which in turn aids in preventing overfitting and improves the generalization capabilities of the models [8].

Cropping along the contours of the brain in MRI images helps in focusing the model's attention on the most relevant areas, thereby reducing noise and improving classification accuracy. By eliminating irrelevant parts of the image, this technique ensures that the model learns more effectively from the essential features of the brain structure [9].

In addition, advanced preprocessing methods like Generative Adversarial Networks (GANs) for augmented data generation have shown to significantly improve the robustness and accuracy of brain MRI classification models by generating high-quality synthetic data to augment the small training set [10].

These preprocessing techniques collectively enhance the quality and quantity of the training data, leading to more accurate and reliable brain MRI classification models.

2.1.2 Convolutional Neural Networks (CNNs) and Variants

Convolutional Neural Networks (CNNs) have been extensively used for brain MRI classification. A notable approach is the use of ResNet-18 architecture for differentiating MRI sequence types, achieving an impressive accuracy of 97.9% on test sets. This model demonstrates the capability of CNNs in handling complex MRI data efficiently[11].

2.1.3 Transfer Learning in Brain MRI Classification

Transfer learning has become a prominent technique in enhancing the performance of deep learning models, especially when dealing with limited labeled data. This approach involves leveraging pre-trained models on large datasets and fine-tuning them on specific tasks such as brain MRI classification.

For example, transfer learning architectures such as InceptionV3, VGG19, DenseNet121, and MobileNet have been applied in brain MRI classification for brain tumor diagnosis, with MobileNet demonstrating the highest accuracy of 99.60% [12]. This highlights the effectiveness of using pre-trained models in achieving high accuracy with relatively small datasets.

In another study, transfer learning was utilized to address the challenges of small sample sizes and high dimensionality in whole-brain fMRI data. The approach showed improved performance compared to training models from scratch, demonstrating the benefits of transfer learning in enhancing model performance with limited data [13].

Furthermore, a study on brain tumor detection employed transfer learning by fine-tuning a deep learning model trained on a large dataset of natural images. This method resulted in comparable performance even with a smaller dataset, showcasing the practicality and effectiveness of transfer learning in medical imaging [14].

2.1.4 U-Net and Variants

U-Net and its variants have become the cornerstone in brain MRI segmentation tasks due to their proficiency in capturing spatial information effectively. The U-Net architecture has been extensively applied to

brain tumor segmentation in MRI images, achieving notable accuracy and robustness in tumor detection [15]–[17].

Various adaptations of U-Net have been proposed to enhance its performance in medical image analysis. Abd-Ellah et al. [16] introduced Two Parallel Cascaded U-Nets with an Asymmetric Residual (TPCUAR-Net) architecture as part of a two-stage detection and segmentation model. This model demonstrated high accuracy in brain tumor detection and segmentation tasks, underscoring the potential of U-Net variants.

Similarly, Imtiaz et al. [15] presented a modified U-Net architecture designed for pixel-level segmentation using a small dataset of brain tumor MRI images (322 images). Their model achieved high accuracy and robustness, further validating the efficacy of U-Net in medical image analysis.

Furthermore, Ding et al. [17] introduced the SLF-UNet architecture, which integrates U-Net with a spatial attention mechanism to enhance segmentation accuracy. The SLF-UNet model achieved significant accuracy in tumor segmentation tasks, highlighting the benefits of incorporating attention mechanisms into the U-Net framework.

Zhou et al. [18] proposed a novel U-Net architecture (U-Net++), which incorporates dense skip connections and deep supervision to improve segmentation performance. The U-Net++ model demonstrated superior performance compared to traditional U-Net architectures, showcasing the potential of advanced U-Net variants in medical image segmentation tasks.

These studies collectively demonstrate the versatility and effectiveness of U-Net and its variants in brain tumor segmentation, making them invaluable tools in medical image analysis.

2.1.5 Conclusion

In conclusion, deep learning has significantly advanced the field of brain MRI classification, offering robust, accurate, and efficient solutions for diagnosing and monitoring brain tumors. The continuous evolution of CNN architectures, transfer learning techniques, optimization strategies, and preprocessing methods promises further enhancements in the accuracy and reliability of these models, ultimately contributing to improved patient outcomes and reduced diagnostic times. These advancements not only highlight the transformative potential of deep learning in medical imaging but also underscore the importance of ongoing research and development. By continuing to refine and innovate upon these techniques, the medical community can look forward to even more effective and efficient diagnostic tools, paving the way for better patient care and outcomes.

3 Dataset Exploration

The dataset, referred to as `dataset_19`, comprises a collection of MRI images organized into four subfolders, each representing a distinct class of brain tumors. Each subfolder contains 120 images, resulting in a well-balanced dataset with a total of 480 images. This balance is essential for training deep learning models, as it mitigates the risk of bias towards any particular class, ensuring that the model learns to distinguish among all categories effectively.

The dataset includes images representing four types of brain tumors: glioma, meningioma, pituitary tumors, and cases with no tumors. Specifically, it consists of 120 images of glioma tumors, 120 images of meningioma tumors, 120 images of pituitary tumors, and 120 images with no tumors.

Each image in the dataset captures the brain from different perspectives, including axial, coronal, and sagittal views. These diverse views are crucial as they provide comprehensive information about the tumor's location, size, and relation to surrounding structures. Axial views offer a horizontal slice of the brain, coronal views provide a frontal slice, and sagittal views give a side slice. Including these various

perspectives ensures that the model can learn robust features and improve its ability to generalize across different cases.

The distribution of images across these classes is depicted in Figure 1. This visual representation underscores the dataset's balance and uniformity, further highlighting its suitability for developing a deep learning model for brain tumor classification.

Figure 1: Distribution of images in `dataset_19`.

Overall, `dataset_19` provides a diverse and balanced set of images, essential for training an accurate and reliable deep learning model. The inclusion of different tumor types and multiple views per case enriches the dataset, making it a valuable resource for the task of brain tumor classification.

3.1 Tumor Types

The dataset encompasses four distinct classes of brain tumors, each representing a unique type of pathology. A thorough understanding of these tumor types is vital for developing a model capable of accurately classifying them.
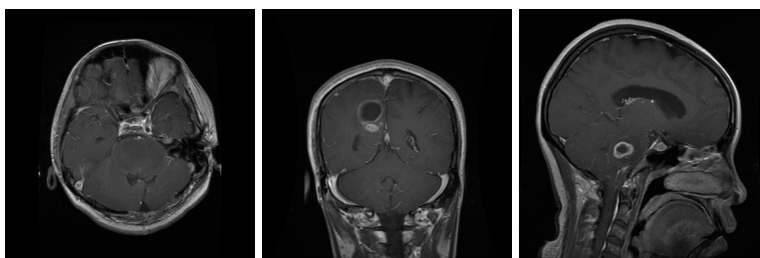

Figure 2: Glioma Tumor: Axial, Coronal, and Sagittal Views.

Gliomas are the most prevalent primary brain tumors, originating in the glial cells of the brain. These tumors can be further classified into subtypes such as astrocytoma, oligodendroglioma, and glioblastoma. Gliomas are characterized by their infiltrative nature and high recurrence rates, often presenting

with varying shapes and irregular borders. The complexity and variability in glioma morphology make them particularly challenging to classify, underscoring the importance of incorporating detailed MRI views to capture their diverse appearances [16].
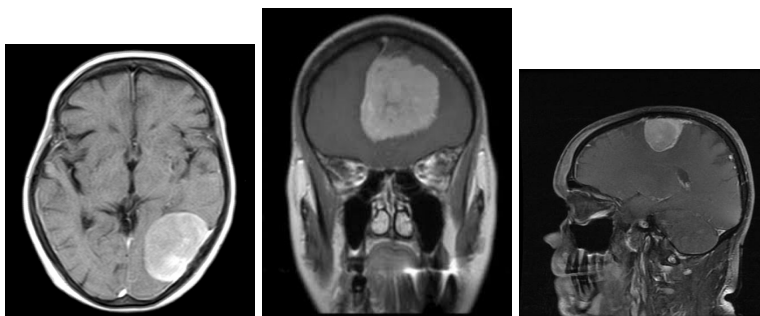

Figure 3: Malignoma Tumor: Axial, Coronal, and Sagittal Views.

Meningiomas, on the other hand, are typically benign tumors arising from the meninges, the protective layers surrounding the brain and spinal cord. These tumors are usually slow-growing and well-defined, which often makes them easier to surgically remove compared to other types. The well-defined nature of meningiomas benefits from clear imaging perspectives, which assist in their precise identification and classification.
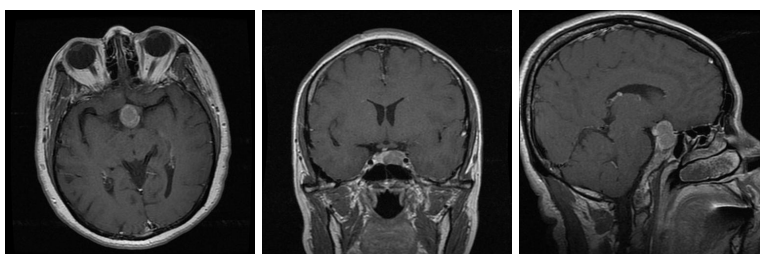

Figure 4: Pituitary Tumor: Axial, Coronal, and Sagittal Views.

Pituitary tumors develop in the pituitary gland, a small but crucial organ located at the base of the brain. These tumors can be either benign or malignant. Given the pituitary gland's critical role in hormone regulation, accurate identification of pituitary tumors is essential. Detailed MRI views help in assessing the tumor's impact on the surrounding brain structures and the gland itself.
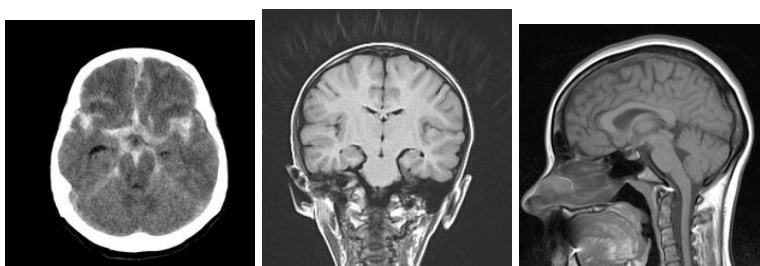

Figure 5: No Tumor: Axial, Coronal, and Sagittal Views.

Lastly, the class labeled 'No Tumor' includes images of the brain without any detectable tumors. This class is crucial for training the model to accurately distinguish between pathological and non-pathological images, thereby reducing false positives in diagnoses.

In summary, each tumor type in the dataset presents unique challenges and characteristics that must be considered in model development. The inclusion of diverse MRI views, such as axial, coronal, and sagittal perspectives, provides a comprehensive understanding of the tumor morphology and spatial relationships, thereby enhancing the model's ability to accurately classify and differentiate between various brain tumor types.

3.2 Image Sizes

The overall average size of the images is approximately $[453.36, 450.83]$ pixels. The smallest image size is $(198, 150)$ pixels, and the largest image size is $(1075, 890)$ pixels. The mean, median, and mode sizes across the dataset are consistent at $[512, 512]$ pixels, suggesting a central tendency around these dimensions.

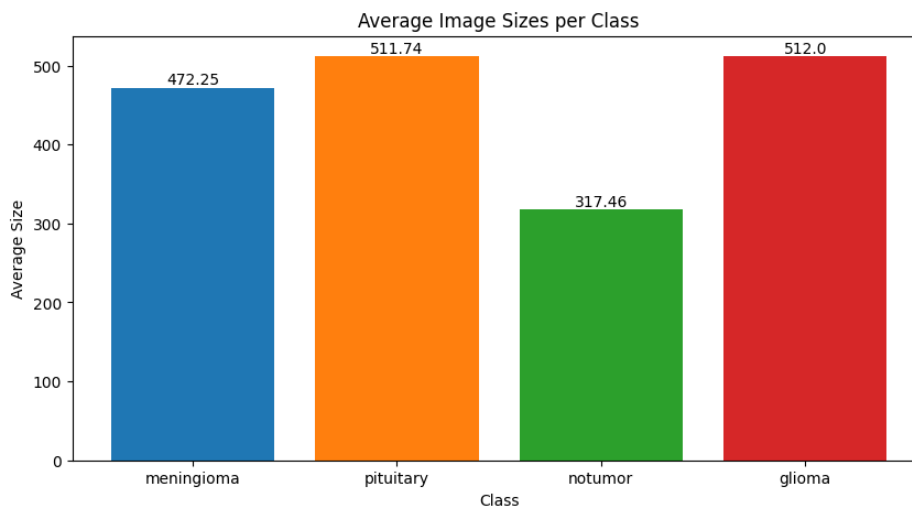
Figure 6: Distribution of brain image sizes in dataset_19.

Table 1: Overall Statistics of Brain Image Sizes

Statistic	X	Y
Mean Size	453.3625	450.83125
Median Size	512	512
Mode Size	512	512
Standard Deviation of Size	124.1067	131.5980
Variance of Size	15402.4644	17318.0236
Smallest Size	198	150
Largest Size	1075	890

Table 2: Class-Specific Statistics of Brain Image Sizes

Statistic	Meningioma	Pituitary	Notumor	Glioma
Mean Size (X)	472.25	511.7417	317.4583	512
Mean Size (Y)	466.4083	510.5417	314.3750	512
Median Size (X)	512	512	236	512
Median Size (Y)	512	512	236	512
Mode Size (X)	512	512	236	512
Mode Size (Y)	512	512	236	512
Standard Deviation of Size (X)	100.7078	43.5696	154.5843	0
Standard Deviation of Size (Y)	108.4743	30.7834	174.3213	0
Variance of Size (X)	10142.0708	1898.3083	23896.3149	0
Variance of Size (Y)	11766.6749	947.6149	30387.9010	0
Smallest Size (X)	223	256	198	512
Smallest Size (Y)	200	256	150	512
Largest Size (X)	650	903	1075	512
Largest Size (Y)	591	721	890	512

Class-specific statistics reveal a notable variation in image sizes. For the *meningioma* class, image sizes range from $(223, 200)$ to $(650, 591)$ pixels, with a mean size of approximately $[472.25, 466.41]$ pixels. The *pituitary* class exhibits sizes from $(256, 256)$ to $(903, 721)$ pixels, with a mean size of around $[511.74, 510.54]$ pixels. The *notumor* class has sizes ranging from $(198, 150)$ to $(1075, 890)$ pixels, and a mean size of approximately $[317.46, 314.38]$ pixels. In contrast, all images in the *glioma* class are consistently sized at $[512, 512]$ pixels.

This variability in image sizes across different classes underscores the importance of size normalization during the preprocessing phase. While the *glioma* class images are uniformly sized, other classes exhibit significant variation. If not properly addressed, this could introduce biases or inconsistencies in the training process.

Proper preprocessing, including size normalization, is vital for several reasons. Firstly, neural networks require inputs of a fixed size, and resizing is essential to facilitate batch processing of images. Secondly, ensuring all images are of the same size guarantees consistent feature representation across the dataset, aiding in better model learning. Additionally, smaller, consistent image sizes can significantly reduce computational load and memory requirements, thus making training more efficient. Lastly, properly preprocessed images can lead to improved model convergence and accuracy, as the network can learn from a standardized set of inputs.

It is important to note that the image size fed into each model will differ, as each pretrained model is trained on different image sizes. This will be further elaborated in subsequent sections. In conclusion, resizing images to a uniform dimension during the preprocessing step is a critical consideration in developing a robust and efficient deep learning model for brain tumor classification. This step not only facilitates smooth training but also contributes to the overall performance and generalization capability of the model.

4 Data Preprocessing

Data preprocessing is a vital step in the data mining process, particularly in the context of deep learning for image classification. This step involves techniques such as image augmentation, affine transforma-

tions, and resizing, which are essential for enhancing model performance. For this project, preprocessing is applied to the provided dataset (dataset_19) to ensure it is suitable for the brain tumor classification task. The specific steps involved in the preprocessing of this dataset are outlined below.

4.1 Image Cropping and Enhancement

A fundamental step in our preprocessing workflow is the cropping and enhancement of images to focus exclusively on the brain, effectively eliminating irrelevant background elements. The following sequence outlines the improved procedures:

Conversion and Blurring: The images are first converted to grayscale to diminish computational complexity and concentrate on structural attributes without the influence of color. This is immediately followed by the application of a Gaussian blur with a 3×3 kernel to soften the image noise, which facilitates more reliable edge detection in the subsequent stages.

Thresholding and Morphology: Using a threshold value of 45, adaptive thresholding converts the grayscale images into binary images to segregate the brain tissue from the background. The binary images then undergo a series of erosions and dilations—two iterations each—to eradicate small noise regions and clarify the brain's structural outline.

Contour Detection and Cropping: Contours are extracted from the thresholded image using external retrieval mode, with the largest contour presumed to represent the brain's boundary. This contour's extreme points (left, right, top, bottom) are determined, and the image is cropped to this bounding rectangle, expanded by a configurable pixel margin, to ensure the brain is isolated without clipping.

Further Processing: Each cropped image is then converted back to grayscale, enhanced using bilateral filtering to further reduce noise and sharpen edges, and mapped to a 'bone' color scheme to improve visual contrast and detail. Images are resized to 224×224 pixels for uniformity and computational efficiency.

These steps, as shown in Figure 7, not only decrease the computational load but also significantly enhance the clarity and focus on essential features for more effective classification.
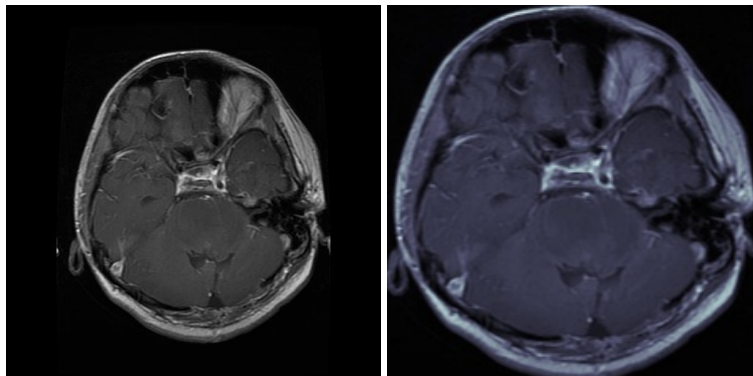

Figure 7: Cropping the MRI image along its contour.

4.2 Data Augmentation

Following the initial preprocessing steps, we apply image augmentation techniques using the `ImageDataGenerator` class from the Keras library. Each model is fine-tuned with a subset of these techniques, which include horizontal flips, random rotations, and normalization. As referenced in [19], certain affine augmentations can be advantageous for brain tumor classification tasks. However, it is crucial to note that some

augmentations, such as vertical flips, may be detrimental due to the potential misrepresentation of tumor locations.

Importantly, these augmentation techniques are applied exclusively to the training dataset, while the validation dataset remains unaltered. This strategy ensures that the model encounters a diverse array of augmented images during training, thereby enhancing its generalization capabilities without compromising the integrity of the evaluation process.

For each model evaluation, we provide a detailed commentary on the preprocessing and augmentation techniques employed, along with the rationale behind their selection. This comprehensive analysis elucidates how these techniques contribute to the overall performance and generalization capabilities of the models.

4.3 Data Splitting

After preprocessing and augmentation, the dataset is divided into training and validation sets, with the training set comprising 80% of the data and the validation set the remaining 20%. This distribution allows the model to be trained extensively while also being evaluated on a separate set of data to assess its performance, as depicted in Figure 8.
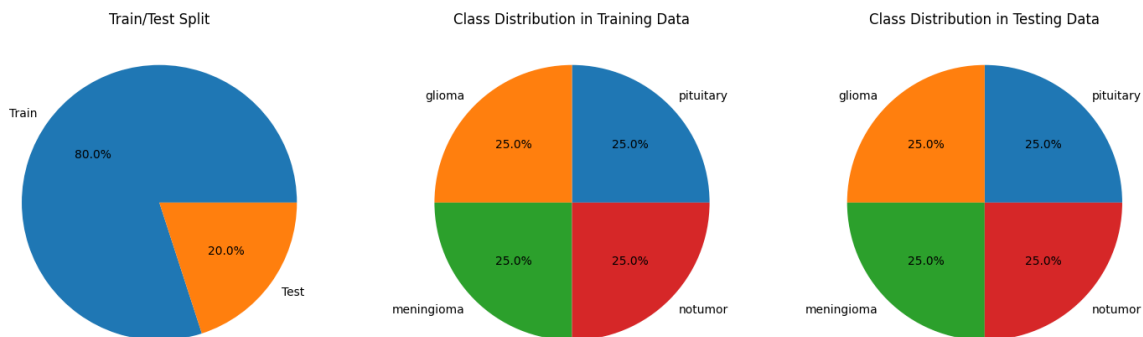

Figure 8: Splitting the dataset into training and validation sets.

4.4 Summary and Justifications

Given the relatively small size of dataset_19 (480 images in total), data augmentation is crucial for enhancing the dataset's variability and improving the model's generalization capabilities. Flipping images horizontally is justified based on the anatomical symmetry of the brain's hemispheres, allowing for effective augmentation without misrepresenting tumor locations [19]. This technique increases the diversity of the training data, which is particularly important given the limited size of the dataset compared to larger datasets, such as those used in the BraTS competition.

Similarly, random rotations are applied since brain images can be rotated in various directions, further increasing the dataset's variability and aiding in model training. The augmentation strategies, including horizontal flipping and random rotations, are essential for compensating for the smaller dataset size by exposing the model to a wider variety of image orientations and perspectives. This approach helps create a more robust model capable of accurately classifying brain tumors despite the limited amount of original training data.

5 Data Mining and Evaluation

In this section, various data mining techniques are employed to extract meaningful insights from the preprocessed data.

5.1 Overview

All models were trained on Google's Colab plaform, utilizing the GPU runtime for faster training. In the following subsections, we will discuss the data mining techniques used to model the brain tumor classification task. The techniques include data augmentation, data splitting, and model selection.

5.2 Model Selection

Model selection is a critical step in developing an effective machine learning model, especially for complex tasks like brain tumor classification. The choice of model impacts the accuracy, efficiency, and overall performance of the solution. This section discusses the use of transfer learning with pretrained models such as VGG16 and ResNet50, as well as deep learning approaches like U-Net, which have demonstrated high performance in related tasks such as the BraTS Competition.

The BraTS (Brain Tumor Segmentation) Competition has established itself as a benchmark for brain tumor segmentation tasks over several years. The competition has highlighted various models and techniques that consistently achieve superior results. Among the most commonly used models by top teams in the competition are U-Net, VGG16, and EfficientNet. These models have been chosen due to their proven ability to effectively handle medical imaging tasks, making them ideal candidates for our project.

Transfer Learning with Pretrained Models: Transfer learning involves leveraging pretrained models, which have already been trained on large datasets, and fine-tuning them for specific tasks. This approach is particularly useful when dealing with small datasets, as it allows the model to benefit from the knowledge gained during pretraining.

VGG16: VGG16 is a convolutional neural network (CNN) known for its simplicity and effectiveness. It has been pretrained on the ImageNet dataset, which contains millions of images across thousands of categories. For our brain tumor classification task, VGG16 can be fine-tuned to learn the specific features of MRI images, leveraging its deep architecture to capture intricate details.

ResNet50: ResNet50 is another pretrained model that has gained popularity for its performance and efficiency. It introduces a residual learning framework to ease the training of deep networks. ResNet50's architecture allows it to maintain performance while being computationally efficient, making it a suitable choice for our dataset.

Deep Learning with U-Net: U-Net is a deep learning model specifically designed for biomedical image segmentation. Its architecture consists of a contracting path to capture context and a symmetric expanding path for precise localization. U-Net has been widely used in medical imaging due to its ability to segment images accurately. In the context of brain tumor classification, U-Net can be adapted to highlight and classify different regions of MRI scans, making it a powerful tool for our project.

Model Selection Process: The model selection process involves evaluating the suitability of each model based on the dataset characteristics and the specific requirements of the classification task. Given the relatively small size of our dataset (dataset_19), models that can generalize well from limited data are preferred.

By leveraging these models, we aim to achieve high accuracy in classifying brain MRI images, thereby supporting healthcare professionals in the efficient diagnosis of brain tumors and facilitating early intervention. The combination of transfer learning and deep learning techniques ensures that our approach is both robust and effective, despite the limited size of the dataset.

5.3 Metrics

After training the models, their performance is evaluated using a variety of metrics, providing a comprehensive understanding of their effectiveness in classification tasks. Below are the commonly used metrics, along with their mathematical formulations based on True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values.

The confusion matrix offers a detailed breakdown of the model's predictions, illustrating the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This matrix is crucial for identifying the model's strengths and weaknesses in classifying different classes, allowing for a granular analysis of performance.

Precision is a metric that measures the proportion of true positive predictions among all positive predictions made by the model. It is calculated using the formula Precision $= \frac{TP}{TP+FP}$. A higher precision value is preferred as it indicates the model's ability to avoid false positives, making it a critical measure in scenarios where the cost of false positives is high.

Recall, also known as sensitivity, measures the proportion of true positive predictions among all actual positive instances in the dataset. It is computed using the formula Recall $= \frac{TP}{TP+FN}$. A higher recall value is desirable as it indicates the model's ability to capture all positive instances, which is particularly important in medical diagnoses where missing a positive case can have severe consequences.

The F1 score, which is the harmonic mean of precision and recall, provides a balanced measure of the model's performance by considering both false positives and false negatives. The formula for the F1 score is F1 Score $= 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$. A higher F1 score signifies better performance, especially in cases where a balance between precision and recall is essential.

Support refers to the number of instances in each class. It helps identify class imbalances and assess the model's performance across different classes, providing context to other metrics.

The Dice Similarity Coefficient (DSC) is a metric commonly used in medical image segmentation tasks. It measures the overlap between the predicted and ground truth segmentation masks, with a higher DSC indicating better segmentation accuracy. The formula for DSC is DSC $= \frac{2\times TP}{2\times TP+FP+FN}$.

Specificity measures the proportion of true negative predictions among all actual negative instances in the dataset. It is calculated as Specificity $= \frac{TN}{TN+FP}$. A higher specificity value is preferred as it indicates the model's ability to avoid false positives, which is crucial in ensuring that negative cases are accurately identified.

Accuracy measures the proportion of correct predictions made by the model across all classes, providing an overall assessment of the model's performance. It is computed using the formula Accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$. A higher accuracy value signifies a better-performing model.

Finally, the Receiver Operating Characteristic (ROC) curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The Area Under the ROC Curve (AUC) measures the entire two-dimensional area underneath the ROC curve, providing an aggregate measure of performance across all classification thresholds. A higher AUC value is preferred as it indicates better model performance.

These metrics, when analyzed together, offer a comprehensive evaluation framework. Precision, recall, and F1 score primarily assess the model's performance on individual classes, while the confusion matrix and support provide detailed insights into prediction distribution. DSC is particularly crucial for

segmentation tasks, ensuring accurate overlap measurement between predicted and actual segmentation masks. Specificity and accuracy offer additional layers of performance evaluation, ensuring robust and reliable model assessment.

5.4 U-Net

Woon Jun Wei (2200624)

The U-Net architecture, introduced by Ronneberger et al. in 2015, is a pioneering model for biomedical image segmentation, widely adopted for various image segmentation tasks [20]. U-Net features a symmetric architecture with a contracting path to capture context and an expansive path for precise localization. The architecture includes skip connections that concatenate feature maps from the contracting path to the expanding path, enhancing the segmentation performance (see Figure 11).
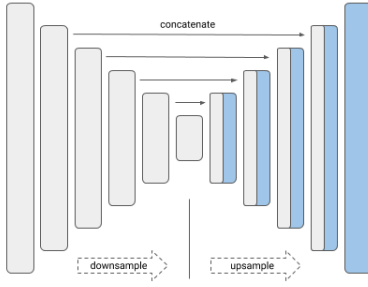

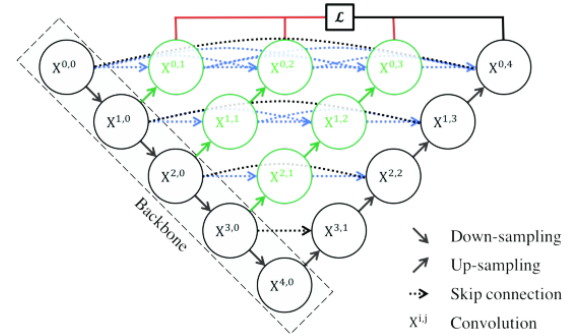Figure 9: (a) Traditional U-Net architecture [21]

Figure 10: (b) U-Net++ architecture [22]

Figure 11: Comparison of Simple U-Net and U-Net++ architectures

5.4.1 Implementation

U-Net++, an extension introduced by Zhou et al. in 2018, improves the original architecture by incorporating dense skip connections. These connections concatenate feature maps from all previous layers in the contracting path to the expanding path, enhancing feature reuse and segmentation accuracy [18], [22]–[24] (See Figure 11). The U-Net++ architecture offers improved performance and efficiency compared to the original U-Net model, making it well-suited for medical image segmentation tasks.

Due to the complexity and time constraints associated with training, evaluation, and tuning, a traditional U-Net model was attempted to be architected from scratch, but was found to be infeasible within the scope of this project. Instead, this study employed the U-Net++ architecture using the `segmentation_models` library [21]. This library offers implementations of various deep learning models for image segmentation, supporting pre-trained models such as VGG, ResNet, and EfficientNet as downsampling backbones.
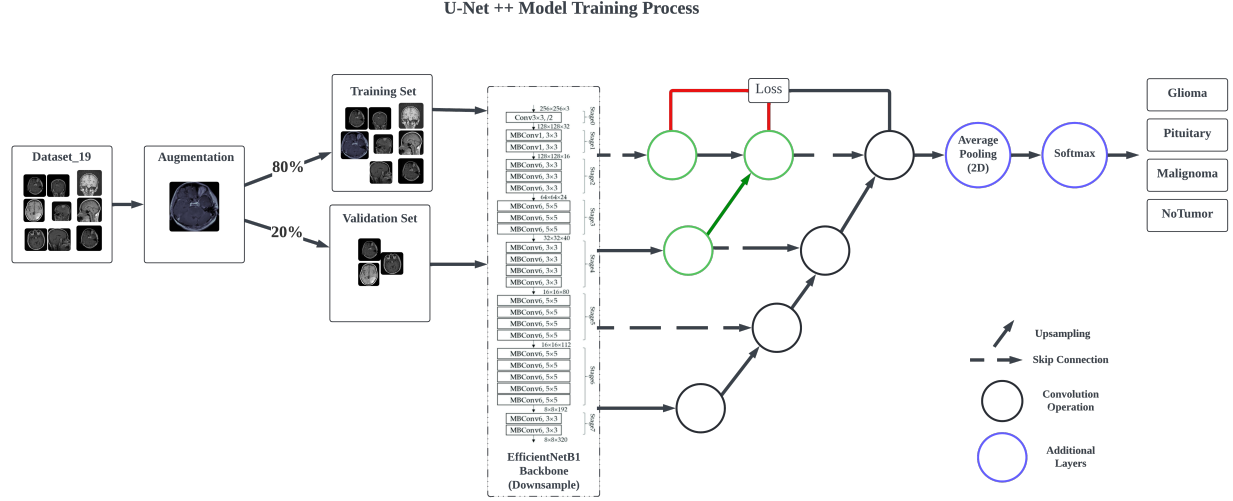
Figure 12: U-Net++ Training Process

Input images were preprocessed by cropping, enhancing, and resizing to $224 \times 224 \times 3$, as detailed in Section 4.1. Further preprocessing included normalization, horizontal flipping, and random rotation by 20 degrees, managed by the `ImageDataGenerator` class, which also handled the 80:20 training-validation split.

The implemented model architecture was U-Net++ with an EfficientNetB1 backbone, initialized with ImageNet weights. To obtain classification results for the full image instead of pixel-wise classification, a Global Average Pooling layer followed by a Dense layer with four units and a softmax activation function was added. The model was compiled using the categorical cross-entropy loss function and accuracy metric.

EfficientNetB1 was chosen as the backbone due to its balance between model size and performance, offering a good trade-off for the brain tumor segmentation task. EfficientNet models are known for their efficiency in terms of accuracy and computational resources, making them suitable for medical image segmentation tasks [25]. The EfficientNetB1 backbone provides a strong feature extractor while maintaining a relatively compact size, enabling efficient training and inference.

The U-Net++ Model's training process is illustrated in Figure 12, the U-Net++ Model is referenced from [22].

Training was conducted using the Adam optimizer with a learning rate of 0.0001 and a batch size of 10 for 100 epochs. The best model was saved based on validation loss, incorporating Checkpointing, Early Stopping, and Reduce Learning Rate on Plateau. Training halted at 33 epochs due to early stopping, achieving a validation loss of 0.1862 and a validation accuracy of 0.9444.

5.4.2 Fine-Tuning

To further optimize the performance of the U-Net++ model for brain tumor segmentation, fine-tuning was conducted using Optuna, a hyperparameter optimization framework. This process involved testing various backbone architectures to identify the most effective model configuration. The backbones evaluated included VGG16, MobileNetV2 and DenseNet121.

The primary objective of this fine-tuning effort was to minimize the validation loss, thus enhancing the model's accuracy and generalizability. The model configuration for each trial consisted of the U-

Net++ architecture with a specified backbone, followed by a Global Average Pooling layer to reduce the spatial dimensions and a Dense layer with a softmax activation function for the final classification into four classes. The models were compiled using the Adam optimizer with an initial learning rate of 1×10^{-4}, employing categorical cross-entropy as the loss function and accuracy as the performance metric.

Table 3 summarizes the results of the fine-tuning trials, listing the validation loss for each tested backbone.

Table 3: Fine-Tuning Results for Different Backbones

Backbone	Validation Loss
MobileNetV2 (Trial 1)	0.4158
DenseNet121 (Trial 2)	0.3684
MobileNetV2 (Trial 3)	0.7039
VGG16 (Trial 4)	0.7004

The Optuna optimization process involved creating a model with a backbone suggested by Optuna for each trial. Each model was then trained for 35 epochs using the Adam optimizer, with callbacks for learning rate reduction, early stopping, and model checkpointing based on validation loss. The best validation loss achieved during training was recorded as the objective metric for each trial.

While DenseNet121 achieved the lowest validation loss of 0.3684 during the fine-tuning trials, it was observed that the EfficientNetB1 backbone, used during the initial training phase, demonstrated superior performance with a lower validation loss and higher accuracy. Specifically, the EfficientNetB1 model achieved a validation loss of 0.1862 and a validation accuracy of 0.9444, outperforming the DenseNet121 in terms of both loss and accuracy metrics. Consequently, despite the promising results from DenseNet121, EfficientNetB1 was retained as the preferred backbone for the final U-Net++ model due to its demonstrated efficacy and robust performance during the training phase.

The fine-tuning process underscored the significance of backbone selection in enhancing the model's performance. Although DenseNet121 showed potential with the lowest validation loss among the tested backbones, EfficientNetB1 was ultimately chosen for its superior overall performance, contributing to the model's effectiveness in accurately segmenting brain tumors. The results are summarized in Table 3.
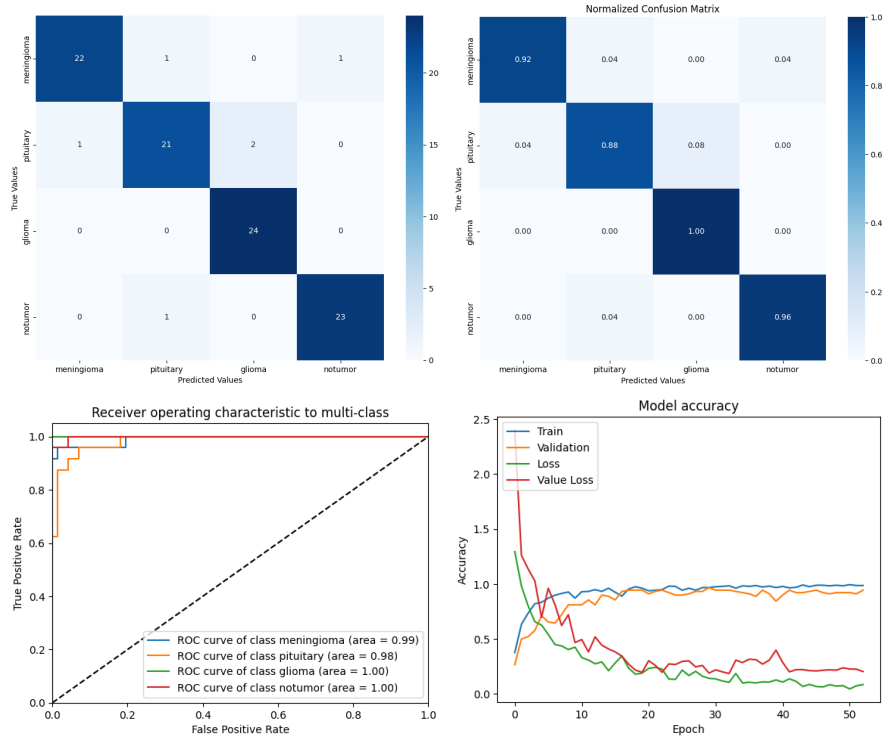
5.4.3   Results and Evaluation

Figure 13: Confusion Matrix, ROC Curve, and Learning Curve for Brain Tumor Segmentation

Table 4: Classification Report for Brain Tumor Segmentation

Class	Precision	Recall	F1-Score	Support
meningioma	0.96	0.92	0.94	24
pituitary	0.91	0.88	0.89	24
glioma	0.92	1.00	0.96	24
notumor	0.96	0.96	0.96	24
micro avg	0.94	0.94	0.94	96
macro avg	0.94	0.94	0.94	96
weighted avg	0.94	0.94	0.94	96
samples avg	0.94	0.94	0.94	96

Table 5: Additional Metrics for Brain Tumor Segmentation

DSC	Sensitivity	Specificity	Accuracy
0.9370	0.9375	0.9792	0.9375

The performance metrics presented in Tables 4 and 5 demonstrate the robust capability of the U-Net model in brain tumor segmentation. The classification report (Table 4) highlights the model's high

precision, recall, and F1-scores across all classes, with an overall accuracy of 0.94. Specifically, the model achieved precision scores of 0.96, 0.91, 0.92, and 0.96 for the meningioma, pituitary, glioma, and notumor classes, respectively. Corresponding recall values were 0.92, 0.88, 1.00, and 0.96, indicating the model's effectiveness in identifying true positives for each class.

The confusion matrix in Figure 13 (top-left) provides a visual representation of the model's classification accuracy, revealing a minimal number of misclassifications. The ROC Curve (bottom left) further substantiates the model's excellent performance, demonstrating a strong ability to distinguish between different classes.

Additional evaluation metrics in Table 5 confirm the model's robustness. The Dice Similarity Coefficient (DSC) of 0.9370 indicates a high overlap between the predicted and actual tumor regions, underscoring the model's precise segmentation capabilities. Sensitivity and specificity values of 0.9375 and 0.9792, respectively, further highlight the model's accuracy in detecting tumor presence while minimizing false positives.

Mentioned in [16], the varying nature and size of glioma tumors pose significant challenges for accurate segmentation. Despite these challenges, the U-Net model achieved an impressive F1-score of 0.96 for glioma, indicating a low incidence of false positives and false negatives. This performance suggests that the model is highly effective in segmenting glioma tumors, although there remains potential for further improvement. Fine-tuning the model or employing additional data augmentation techniques could enhance the segmentation accuracy for glioma tumors, thereby improving overall model performance.
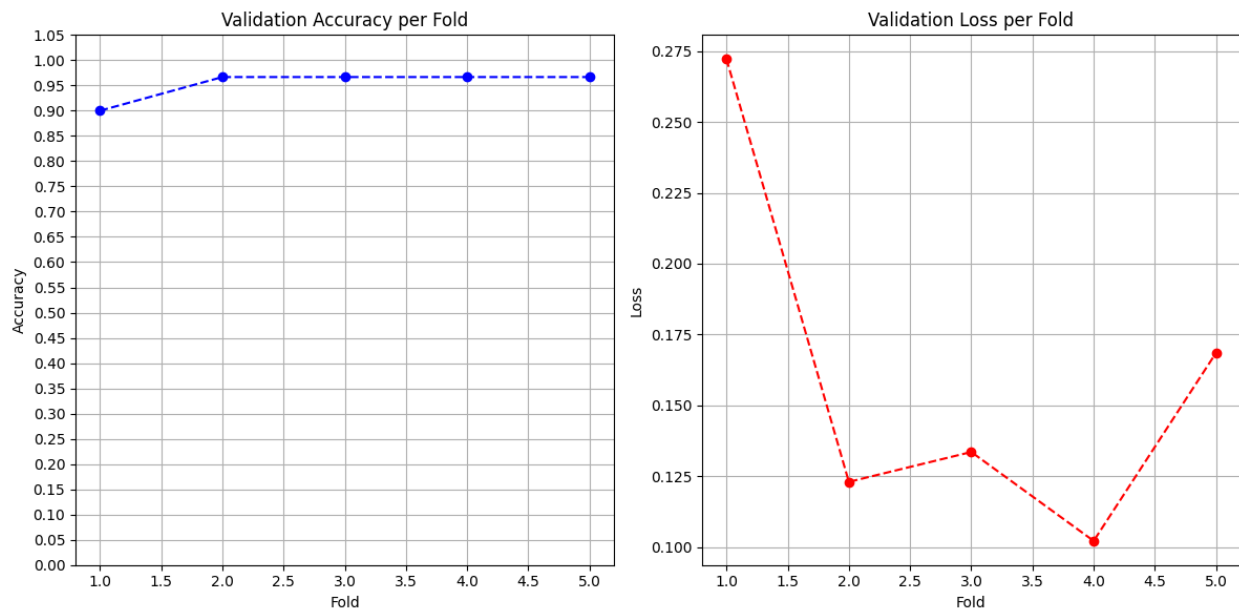
5.4.4 K-Folds Cross-Validation

Figure 14: K-Folds Cross-Validation for Brain Tumor Segmentation

K-Folds cross-validation was conducted to rigorously assess the model's generalization performance on the brain tumor segmentation task. The dataset was partitioned into five folds, ensuring each fold served as the validation set exactly once while the remaining four folds were utilized for training. This procedure was repeated five times, allowing each fold to be used for validation, thus providing a robust

estimate of the model's performance. The model was trained for 35 epochs for each fold, with the initial training stopping at 33 epochs due to early stopping. Model checkpoints were removed for efficiency during this process.

The left plot in Figure 14 illustrates the validation accuracy for each fold. The accuracy values per fold indicate high performance across all splits, with a slight variation. Specifically, the accuracy ranged from approximately 0.90 to 1.00, demonstrating the model's strong capability in accurately segmenting brain tumors. This high level of accuracy across different folds suggests that the model generalizes well to unseen data, maintaining its performance across various subsets of the dataset.

The right plot in Figure 14 presents the validation loss for each fold. The loss values show greater variability compared to the accuracy, with the highest loss observed in fold 1 and the lowest in fold 4. The variability in the loss values suggests that while the model generally performs well, there are certain instances where the model's performance can be further optimized. This indicates potential areas for improvement, particularly in achieving more consistent performance across different validation sets.

It is noteworthy that during the K-Folds validation, there were instances where the validation loss was smaller than the training's validation loss, and the accuracy during K-Folds was also higher. This phenomenon underscores the model's robustness, indicating that it is capable of performing well on unseen data. However, this occurrence may also be attributed to the differences in data distribution between the training and validation sets in each fold. This disparity can sometimes lead to lower validation loss and higher accuracy, suggesting that while the model is strong, the evaluation metrics might slightly overestimate the model's performance. Therefore, it is important to consider these factors when interpreting the results.

Overall, the K-Folds cross-validation results indicate that the model achieves a high average accuracy of 0.9533 with a standard deviation of 0.0267, reflecting robust performance with minimal variability. The average validation loss was 0.1599 with a standard deviation of 0.0601, suggesting that while the model's predictions are generally reliable, there is room for enhancement in terms of reducing the prediction error. These metrics collectively demonstrate that the model possesses strong generalization capabilities, although further refinement could help in achieving even greater consistency and reliability in its performance.

5.4.5 Conclusion

In this study, we implemented and fine-tuned the U-Net++ model for the challenging task of brain tumor segmentation, utilizing various backbone architectures. Our comprehensive approach involved initial training with the EfficientNetB1 backbone, followed by fine-tuning with alternative backbones such as VGG16, MobileNetV2, and DenseNet121. The EfficientNetB1 backbone demonstrated superior performance, achieving a validation loss of 0.1862 and a validation accuracy of 0.9444. This backbone was retained for the final model due to its robust performance metrics and efficacy in accurately segmenting brain tumors.

The results of our fine-tuning process, supported by Optuna's hyperparameter optimization framework, underscored the importance of backbone selection in enhancing model performance. Despite DenseNet121 achieving the lowest validation loss during fine-tuning trials, the overall performance of EfficientNetB1 in terms of validation loss and accuracy solidified its selection as the preferred backbone.

Our evaluation included rigorous testing through K-Folds cross-validation, which demonstrated the model's strong generalization capabilities with high accuracy and relatively low variability across different data folds. The average validation accuracy of 0.9533 and validation loss of 0.1599 (from K-Folds Evaluation) reaffirm the model's robustness and reliability in segmenting brain tumors.

The study highlighted the U-Net++ model's ability to handle the complexity of medical image

segmentation tasks effectively. The high precision, recall, and F1-scores achieved across various tumor classes indicate the model's capability to provide accurate and reliable segmentation results. The findings suggest that with further optimization and potential enhancements in data augmentation techniques, the model's performance can be further improved, making it a valuable tool in clinical applications for brain tumor diagnosis and treatment planning.

Overall, this work demonstrates the effectiveness of leveraging advanced deep learning architectures and optimization techniques to address critical challenges in medical image segmentation, paving the way for more accurate and efficient diagnostic tools in healthcare.

5.5 VGG16

This is a simple implementation of the VGG16 model

5.6 ResNet50 Implementation

Benjamin Loh Choon How (2201590)

ResNet50, a revolutionary architecture in the field of deep learning, employs residual connections which are its defining feature. These connections are crucial for training very deep networks by enabling the direct propagation of gradients from deeper to shallower layers in the network. This architectural choice reduces the problem of vanishing gradients, thus allowing the model to learn effectively even when many layers deep.

For our brain tumor classification task, we utilize the ResNet50 model pre-trained on the ImageNet dataset. This pre-training provides a robust feature extraction base, which we then tailor to our specific task through further training and adaptation.

Model Architecture: Our adaptation of ResNet50 begins with the standard base architecture loaded with ImageNet weights, but excludes the top layer to accommodate our specialized output needs. This modification allows the network to maintain the intricate feature extraction capabilities developed through ImageNet training while adapting to the nuances of brain tumor imagery.

> Imagine ResNet50 like a skilled artisan skilled in restoration. This artisan (the model) is brought into a renovation project (our classification task) where the fundamental structure is sound but needs adaptation and specific enhancements. Just as an artisan would preserve valuable original elements while updating others, our model retains valuable pre-learned features while adapting to new data.

Following the base layers, a Global Average Pooling 2D layer is introduced. This layer reduces the spatial dimensions of the feature maps to a single vector per map, thus summarizing the most critical features while significantly reducing the number of parameters and the computational burden. This pooling step is crucial for maintaining efficiency and avoiding overfitting.

Next, we include a Dropout layer with a 40% drop rate, which randomly disables a fraction of the neurons during training. This randomness helps to prevent overfitting by ensuring that the model does not rely too heavily on any specific set of neurons, encouraging the network to develop redundant pathways to maintain performance even if some neurons are inactive.

The network culminates in a Dense layer with four units corresponding to the different classes of brain tumors, employing the softmax activation function. This layer outputs the probability distribution over the tumor classes, providing a clear, interpretable classification result.

Compilation and Callbacks: The model is compiled using the Adam optimizer with a starting learning rate of 0.0001 and a loss function of categorical crossentropy, which is appropriate for this multi-class classification scenario. To optimize training, we employ several callbacks:

ModelCheckpoint to save the best version of the model based on validation loss *EarlyStopping* to halt training if the validation loss does not improve for a specified number of epochs, preventing overtraining and resource wastage

6 Conclusion
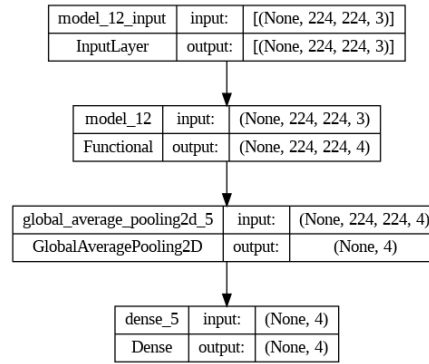
7 Future Work

8 Appendix A

Figure 15: U-Net++ model architecture

Where `model_12` is the U-Net++ model with EfficientNetB4 encoder.

References

[1] M. Nazar, M. M. Alam, E. Yafi, and M. M. Su'ud, "A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques," *IEEE Access*, vol. 9, pp. 153 316–153 348, 2021, Conference Name: IEEE Access, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3127881. [Online]. Available: https://ieeexplore.ieee.org/document/9614151 (visited on 06/08/2024).

[2] P. Doupe, J. Faghmous, and S. Basu, "Machine learning for health services researchers," *Value in Health*, vol. 22, no. 7, pp. 808–815, Jul. 1, 2019, Publisher: Elsevier, ISSN: 1098-3015, 1524-4733. DOI: 10.1016/j.jval.2019.02.012. [Online]. Available: https://www.valueinhealthjournal.com/article/S1098-3015(19)30146-9/fulltext?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1098301519301469%3Fshowall%3Dtrue (visited on 06/08/2024).

[3] D. Rolnick, P. L. Donti, L. H. Kaack, *et al.*, "Tackling climate change with machine learning," *ACM Computing Surveys*, vol. 55, no. 2, 42:1–42:96, Feb. 7, 2022, ISSN: 0360-0300. DOI: `10.1145/3485128`. [Online]. Available: `https://dl.acm.org/doi/10.1145/3485128` (visited on 06/08/2024).

[4] S. Lapointe, A. Perry, and N. A. Butowski, "Primary brain tumours in adults," *The Lancet*, vol. 392, no. 10145, pp. 432–446, Aug. 4, 2018, ISSN: 0140-6736. DOI: `10.1016/S0140-6736(18)30990-5`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0140673618309905` (visited on 06/02/2024).

[5] J. B. Iorgulescu, M. Torre, M. Harary, *et al.*, "The misclassification of diffuse gliomas: Rates and outcomes," *Clinical Cancer Research*, vol. 25, no. 8, pp. 2656–2663, Apr. 15, 2019, ISSN: 1078-0432. DOI: `10.1158/1078-0432.CCR-18-3101`. [Online]. Available: `https://doi.org/10.1158/1078-0432.CCR-18-3101` (visited on 06/08/2024).

[6] L. Lenchik, L. Heacock, A. A. Weaver, *et al.*, "Automated segmentation of tissues using CT and MRI: A systematic review," *Academic Radiology*, vol. 26, no. 12, pp. 1695–1706, Dec. 1, 2019, Publisher: Elsevier, ISSN: 1076-6332, 1878-4046. DOI: `10.1016/j.acra.2019.07.006`. [Online]. Available: `https://www.academicradiology.org/article/S1076-6332(19)30353-8/abstract` (visited on 06/08/2024).

[7] D. L. Krishna, N. V. Dedeepya Padmanabhuni, and G. JayaLakshmi, "Data augmentation based brain tumor detection using cnn and deep learning," in *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, 2023, pp. 317–321. DOI: `10.1109/CISES58720.2023.10183465`.

[8] J. S. Paul, A. Plassard, B. Landman, and D. Fabbri, "Deep learning for brain tumor classification," vol. 10137, 2017. DOI: `10.1117/12.2254195`.

[9] S. Asif, W. Yi, Q. Ain, J. Hou, T. Yi, and J. Si, "Improving effectiveness of different deep transfer learning-based models for detecting brain tumors from mr images," *IEEE Access*, vol. 10, pp. 34716–34730, 2022. DOI: `10.1109/ACCESS.2022.3153306`.

[10] F. Fahimi, S. Došen, K. Ang, N. Mrachacz-Kersting, and C. Guan, "Generative adversarial networks-based data augmentation for brain–computer interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, pp. 4039–4051, 2020. DOI: `10.1109/tnnls.2020.3016666`.

[11] M. A. Mahmutoglu, C. J. Preetha, H. Meredig, *et al.*, "Deep learning–based identification of brain mri sequences using a model trained on large multicentric study cohorts," *Radiology: Artificial Intelligence*, vol. 6, no. 1, e230095, 2024, PMID: 38166331. DOI: `10.1148/ryai.230095`. eprint: `https://doi.org/10.1148/ryai.230095`. [Online]. Available: `https://doi.org/10.1148/ryai.230095`.

[12] M. M. Islam, P. Barua, M. Rahman, T. Ahammed, L. Akter, and J. Uddin, "Transfer learning architectures with fine-tuning for brain tumor classification using magnetic resonance imaging," *Healthcare Analytics*, vol. 4, p. 100270, Dec. 2023. DOI: `10.1016/j.health.2023.100270`.

[13] A. W. Thomas, K.-R. Müller, and W. Samek, "Deep transfer learning for whole-brain fmri analyses," in *OR 2.0 Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging*, L. Zhou, D. Sarikaya, S. M. Kia, *et al.*, Eds., Cham: Springer International Publishing, 2019, pp. 59–67, ISBN: 978-3-030-32695-1.

[14] A. B. Mitta, A. H. Hegde, A. R. K. P., and G. S., "Brain tumor detection: An application based on transfer learning," in *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, 2023, pp. 1424–1430. DOI: `10.1109/ICOEI56765.2023.10125766`.

[15] R. Imtiaz, M. W. Mirza, A. Siddiq, M. Farooq-i-Azam, I. R. Khan, and S. Rahardja, "Brain tumor segmentation from MR images using customized u-net for a smaller dataset," in *2023 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, ISSN: 2766-4465, Oct. 2023, pp. 1–5. DOI: `10.1109/BioCAS58349.2023.10389092`. [Online]. Available: `https://ieeexplore.ieee.org/document/10389092` (visited on 05/28/2024).

[16] M. K. Abd-Ellah, A. I. Awad, A. A. M. Khalaf, and A. M. Ibraheem, "Automatic brain-tumor diagnosis using cascaded deep convolutional neural networks with symmetric u-net and asymmetric residual-blocks," *Scientific Reports*, vol. 14, no. 1, p. 9501, Apr. 25, 2024, ISSN: 2045-2322. DOI: `10.1038/s41598-024-59566-7`. [Online]. Available: `https://www.nature.com/articles/s41598-024-59566-7` (visited on 06/02/2024).

[17] H. Ding, J. Lu, J. Cai, Y. Zhang, and Y. Shang, "SLf-UNet: Improved UNet for brain MRI segmentation by combining spatial and low-frequency domain features," in *Advances in Computer Graphics*, B. Sheng, L. Bi, J. Kim, N. Magnenat-Thalmann, and D. Thalmann, Eds., Cham: Springer Nature Switzerland, 2024, pp. 415–426, ISBN: 978-3-031-50075-6. DOI: `10.1007/978-3-031-50075-6_32`.

[18] 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, 2018, pp. 3–11.

[19] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, "Data augmentation for brain-tumor segmentation: A review," *Frontiers in Computational Neuroscience*, vol. 13, Dec. 11, 2019, ISSN: 1662-5188. DOI: `10.3389/fncom.2019.00083`. [Online]. Available: `https://www.frontiersin.org/articles/10.3389/fncom.2019.00083` (visited on 06/02/2024).

[20] O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, May 18, 2015. DOI: `10.48550/arXiv.1505.04597`. arXiv: `1505.04597[cs]`. [Online]. Available: `http://arxiv.org/abs/1505.04597` (visited on 06/08/2024).

[21] P. Iakubovskii, *Segmentation models*, `https://github.com/qubvel/segmentation_models`, 2019.

[22] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, *UNet++: A nested u-net architecture for medical image segmentation*, Jul. 18, 2018. DOI: `10.48550/arXiv.1807.10165`. arXiv: `1807.10165[cs,eess,stat]`. [Online]. Available: `http://arxiv.org/abs/1807.10165` (visited on 06/08/2024).

[23] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, 2019.

[24] Z. Zhou, "Towards annotation-efficient deep learning for computer-aided diagnosis," Ph.D. dissertation, Arizona State University, 2021.

[25] W. Hastomo, A. Satyo, E. Sestri, *et al.*, "Classification of brain image tumor using EfficientNet b1-b2 deep learning," *Semesta Teknika*, vol. 27, pp. 46–54, May 2, 2024. DOI: `10.18196/st.v27i1.19691`.