CS675: Final Project Report

Sky Cen, Hanyu Zhu, Yuhan Chen

December 19, 2024

1 Abstract

Our project focuses on brain tumor MRI analysis, aiming to improve early detection and accurate diagnosis through automated image processing, particularly in clinical scenarios such as pre-surgical planning, monitoring tumor progression, and supporting radiologists in time-critical emergency cases. Brain tumors are life-threatening and require precise medical evaluation for effective treatment. Manual MRI analysis is time-consuming and prone to human error due to the variability of the tumors, as supported by studies indicating diagnostic discrepancies of up to 20% between radiologists interpreting the same scans [1]. This highlights the necessity of automated solutions to enhance accuracy and consistency in brain tumor detection.

2 Literature Review

Research in brain tumor MRI analysis underscores the need for automated diagnostic tools due to persistent challenges in manual evaluation. Studies such as Bauer et al. [1] and Menze et al. [2] highlight high inter-observer variability in manual MRI analysis, with diagnostic errors reported in up to 20% of cases. Lundervold & Lundervold [3] emphasize that deep learning models, particularly CNNs, can mitigate this issue by leveraging hierarchical feature extraction capabilities for accurate tumor segmentation. However, as Taha & Hanbury [4] note, data scarcity and imbalance remain major challenges due to limited annotated medical datasets. Advances like the inclusion of attention mechanisms [5] and transfer learning have improved model sensitivity, particularly for rare or small tumors. Furthermore, Perez et al. [6] discuss the impact of variations in imaging protocols across institutions, complicating generalization.

3 Dataset

The dataset used in this project, sourced from Kaggle, which is organized into two sections:

1) Classification Dataset: This section contains MRI images labeled in four categories—glioma, meningioma, pituitary tumor, and no tumor. The dataset is organized into training and test sets, with 826, 822, 827, and 395 images per category in the training set and 100, 115, 74, and 105 images in the test set.

2) Segmentation Dataset: Focused on tumor localization, this section includes images labeled as either tumor or no tumor. For tumor cases, spatial coordinates are provided for precise localization. The dataset is split into training (102 images and coordinate entries), validation (30 images and coordinate entries), and test (15 images and coordinate entries) sets.

The link of the dataset: https://www.kaggle.com/datasets/bilalakgz/brain-tumor-mri-dataset

4 Methodology

We employed a deep learning-based approach using convolutional neural networks (CNNs) for brain tumor classification and segmentation, leveraging their proven effectiveness in extracting spatial hierarchies from medical images through layered feature representations. The process is divided into classification and segmentation to accurately identify and locate tumors. The implementation was conducted using the PyTorch framework, leveraging the segmentation-models-pytorch library. Training and evaluation were executed in a GPU-enabled environment to expedite computations and facilitate large-scale model experimentation.

1) Classification:

- Data Preprocessing: The dataset used for brain tumor classification was subjected to a comprehensive preprocessing pipeline designed to enhance model performance and reduce overfitting. This process involved three critical stages: normalization, resizing, and data augmentation. First, all images were resized to a fixed resolution of 244×244 pixels. This step maintained uniform input size, simplifying batch processing

and model compatibility, then a variety of augmentation techniques were applied to increase dataset diversity and improve the model's ability to generalize across unseen samples. These augmentations included random rotation, horizontal and vertical flipping, random brightness and contrast adjustments, Gaussian noise addition, and elastic transformations. The use of data augmentations reduced model overfitting by exposing the neural network to more varied training samples. Lastly, all pixel values were scaled to the [0, 1] range to standardize the intensity distribution, ensuring consistent input representation for the neural network.

- Model Definition and Training: We explored two deep learning architectures, ResNet and EfficientNet, due to their proven capabilities in image classification tasks. The ResNet architecture allows for deeper networks by leveraging the optimization stability provided by residual blocks. The EfficientNet architecture allows models to be scaled efficiently, allowing for larger models and faster training of such models without performance sacrifice. The larger models were trained using transfer learning, utilizing pretrained ImageNet weights. Transfer learning strategies included freezing early convolutional layers to retain general feature extraction capabilities while fine-tuning deeper layers for task-specific learning. Additionally, learning rate scheduling and early stopping were implemented to optimize training efficiency and prevent overfitting.

- Model Evaluation: The training process employed categorical cross-entropy as the loss function, reflecting the multi-class classification nature of the task. We utilized the AdamW optimizer with a learning rate of 1e-4, coupled with weight decay regularization to combat overfitting. A cosine annealing learning rate scheduler with warm restarts dynamically adjusted the learning rate, enhancing convergence stability. Evaluation metrics included accuracy, and a confusion matrix.

2) Segmentation:

- Data Acquisition and Preprocessing: The initial stage focused on acquiring and standardizing input data to enhance the model's learning capabilities. MRI images were resized to 256x256 pixels to optimize computational efficiency. Pixel intensity normalization was performed to scale the input values between 0 and 1. To improve model generalization and mitigate overfitting, extensive data augmentation techniques were applied, including random rotations, flipping, zooming, and elastic deformations. The dataset was subsequently divided into training (80%) and validation (20%) subsets to ensure robust performance evaluation.

- Mask Image Generation: Accurate mask images are pivotal for effective model training. Ground-truth masks delineating tumor regions were preprocessed to match the resized MRI images. Manual verification was conducted to ensure high-quality segmentation labels, allowing for corrections of any misalignments or labeling errors detected during inspection. This step was essential for establishing reliable training targets.

- Model Training and Selection:

  The models we trained include U-Net, Feature Pyramid Network (FPN), and Pyramid Scene Parsing Network (PSPNet), each known for their strengths in feature extraction, multi-scale representation, and contextual understanding. All of them used ResNet as the encoder. U-Net's architecture facilitated precise localization by combining low-level and high-level features. FPN exploited multi-scale feature maps for improved boundary detection, while PSPNet integrated contextual information through multi-level pooling.

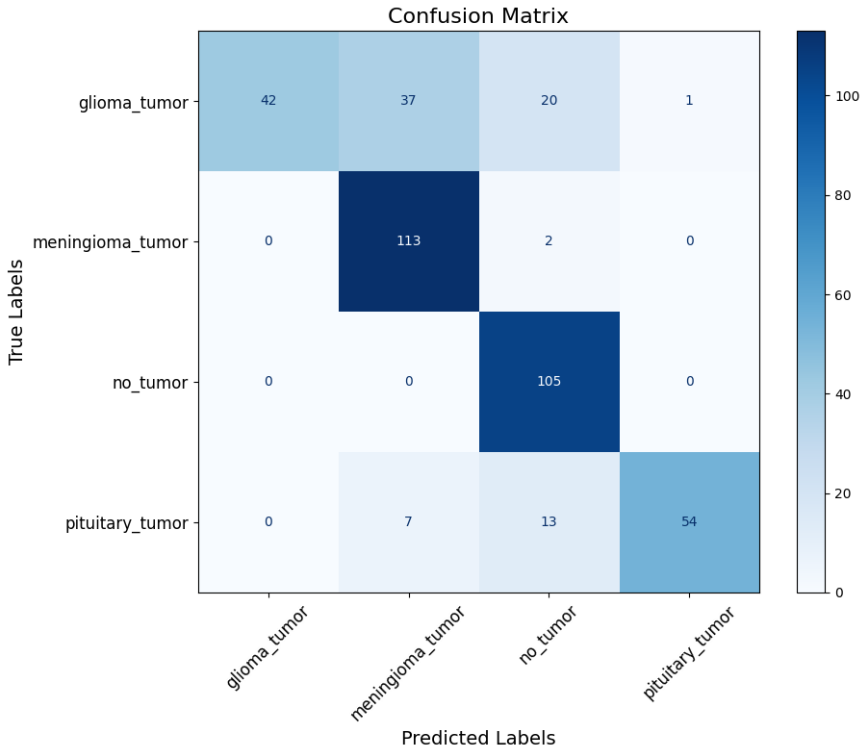
  We employed the Cross Entropy as loss function, which suitable for addressing class imbalances inherent in tumor segmentation tasks. The AdamW optimizer, was initialized with a learning rate of 0.001 to ensure stable convergence. Evaluation metrics included the Dice coefficient, providing a comprehensive assessment of segmentation performance.

- Model Evaluation: The final stage involved a rigorous evaluation of segmentation performance using the defined metrics. Post-processing steps enhanced the predicted segmentation masks. Binary thresholding converted the model's probability outputs into discrete masks, applying a threshold of 0.5. Morphological operations such as opening and closing were used to smooth segmentation boundaries and reduce noise artifacts, ensuring more precise and clinically relevant segmentation outputs.

5 Results

For classification, our final EfficientNet-B7 model achieved a classification accuracy of 79.7%. This performance underscores the model's effectiveness in distinguishing between different brain tumor types, validating its

suitability for clinical diagnostic support applications. The figure below shows the confusion matrix produced by our model when evaluated on held-out test data.



The table below shows the accuracies of the best model for each architecture we were able to train.

Architecture	Accuracy
ResNet34	67.01%
ResNet50	70.05%
EfficientNetB0	76.90%
EfficientNetB1	75.13%
EfficientNetB2	74.62%
EfficientNetB4	73.86%
Pretrained EfficientNetB7	79.70%

For segmentation task, Unet achieved the best Dice coefficient of 0.82, indicating robust tumor segmentation performance. These results suggest that the model generalizes well across diverse tumor morphologies and could be a valuable tool in clinical applications.

The table blow shows the dice scores of the best decoder for each architecture with Resnet as the encoder on segmentation task.

Architecture	Accuracy
Unet+ResNet	82.18%
FNP+ResNet	72.99%
PSPNet+ResNet	79.69%

6 Significance

The abnormal growth of brain cells that characterizes brain tumors poses a significant health risk due to potential organ dysfunction and even death. Brain tumors exhibit a wide range of sizes, textures, and locations, complicating detection and diagnosis. Magnetic resonance imaging (MRI) serves as a critical diagnostic tool for identifying these tumors, enabling timely and accurate clinical evaluation. Automated MRI analysis can support radiologists by reducing diagnostic errors and accelerating the evaluation process, ultimately enhancing healthcare accessibility and improving patient outcomes.

7 Difficulties and Challenges

We face several challenges when conducting MRI analysis using CNN models, including limited dataset and time-consuming model training.

1) Data Imbalance: The no-tumor dataset was smaller than other categories, potentially leading to biased model predictions and reduced generalization capabilities.

2) Hyperparameter Tuning: Finding optimal model configurations was time-consuming and required extensive experimentation. In classification, we haven't been able to find out the model with accuracy higher than 80%, although outperforming the baseline, it is still not satisfactory and needs further improvements. For both tasks, due to high computational requirements and time constraints, we were unable to experiment in depth with hyperparameters, which may result in the model becoming less optimal.

3) Dataset Size: For classification, the dataset size limited our abilities to effectively leverage the benefits of deeper models. This was evident during experimentation: as the EfficientNet size increased, the test accuracy decreased. As a result, pretrained weights were used for larger models. With a larger dataset, we would be able to train deeper models from scratch, without using pretrained weights. For the segmentation task, there are only 102 images for training. Since the size of training data is strictly limited, we also used pretrained weights to make model perform better.

8 Further Improvements

1) Classification: In training, implementing dynamic learning rate adjustments using schedulers combined with advanced optimizers can improve convergence. Adding an early stopping mechanism can prevent overfitting during prolonged training. For evaluation, introducing metrics like F1-Score, Precision, Recall, and AUC-ROC ensures a more comprehensive assessment, while using a normalized confusion matrix helps interpret model performance across all classes.

2) Segmentation: In model training, the learning rate is fixed, and the loss function only uses CrossEntropyLoss, limiting optimization potential. Implementing learning rate schedulers such as ReduceLROnPlateau or CosineAnnealingLR, along with combining loss functions like Dice Loss and IoU Loss, could boost model performance. Evaluation metrics are limited to the Dice Coefficient, and visualization outputs are minimal.

9 Implications and Extentions

Our work highlights the potential integration of automated brain tumor detection into clinical workflows. Automatic classification and segmentation supports radiologists by reducing diagnostic delays, improving early detection, and enabling more personalized treatment planning. Reliable automated systems can lower the diagnostic burden in resource-constrained settings and support telemedicine-based consultations, especially in rural or underserved areas.

Future research could expand the model to handle additional MRI sequences, incorporate multimodal fusion, and explore novel architectures such as transformers for improved contextual understanding. Developing explainable machine learning methods could enhance clinical interpretability, fostering trust among healthcare professionals. Deploying the model through a cloud-based clinical decision support system could enable real-time diagnostics and remote healthcare services.

References

[1] Bauer, S., et al., "A Survey of MRI-based Medical Image Analysis for Brain Tumor Studies," Journal of Medical Imaging, 2013.

[2] Menze, B. H., et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Transactions on Medical Imaging, 2015.

[3] Lundervold, A. S., & Lundervold, A., "An Overview of Deep Learning in Medical Imaging with Focus on MRI," Frontiers in Neuroscience, 2019.

[4] Taha, A. A., & Hanbury, A., "Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool," BMC Medical Imaging, 2015.

[5] Isensee, F., et al., "nnU-Net: A Self-configuring Method for Deep Learning-based Biomedical Image Segmentation," Nature Methods, 2021.

[6] Perez, F., et al., "Generalization Challenges in Medical Imaging Datasets," IEEE Medical Imaging Conference, 2021.