Brain Tumor Segmentation and Classification

Farouk Kotob
Dept of Computer Science
Toronto Metropolitan University
Toronto, ON
Farouk.kotob@torontomu.ca

Brian Yang
Dept of Computer Science
Toronto Metropolitan University
Toronto, ON
Brian.yang@torontomu.ca

Jose John Mullor
Dept of Computer Science
Toronto Metropolitan University
Toronto, ON
Jose.johnmulloor@torontomu.ca

Mohamed Jama
Dept of Computer Science
Toronto Metropolitan University
Toronto, ON
m5jama@torontomu.ca

Abstract—This project presents a deep learning based approach for tumor detection and classification. It uses segmentation and classification tasks to improve and enhance the process of brain tumor analysis. A U-Net model segments tumor regions from MRI scans where it produces binary masks that highlight tumor locations. For classification, a fine-tuned ResNet-50 model is used to classify these tumors into three types: meningioma, glioma, and pituitary. To train and validate this model, a public dataset containing T1-weighted MRI images labelled for segmentation and classification was used in this project. The model consists of multiple operations when it comes to the data preprocessing which is a critical phase of the model, such as: normalization, and resizing. The U-Net model achieved high Dice scores for segmentation accuracy, while the ResNet 50 model attained strong classification metrics, including precision, recall, and F1-score. This hybrid approach shows a potential for deep learning to enhance brain tumor analysis moving forward.

I. INTRODUCTION

Brain tumors pose significant challenges in medical diagnosis and treatment these days due to their complex structures and different types. Early and accurate detection of tumors from medical images is critical for effective treatment planning [1]. Manual analysis of MRI images is time-consuming and prone to errors among radiologists, highlighting the need for automated solutions.

Today deep learning has emerged as a transformative approach in medical imaging, offering unparalleled accuracy and efficiency in analyzing complex datasets [2]. This project focuses on developing a hybrid deep learning approach for brain tumor analysis. It consists of segmentation and classification tasks to identify and categorize the brain tumor from T1-weighted MRI images. The segmentation component uses a U-Net model to precisely delineate tumor regions, while the classification component employs a fine-tuned ResNet 50 model to classify the detected tumors into categories such as meningioma, glioma, and pituitary.

In this project we used a segmented and a classified public dataset in order to implement pre-processing, including normalization, resizing and splitting into training, validation, and test sets. The proposed approach not only improves the accuracy of brain tumor segmentation and classification but also demonstrates the potential of deep learning models to enhance both the accuracy and efficiency of brain tumor analysis.

II. Materials and Methods

A. Dataset

The dataset utilized in this project is a publicly available collection of MRI scans, including a total of 3,064 samples. Each sample includes an MRI image, a segmentation mask and a tumor label. The dataset was sourced from a publicly available repository and is described in detail on the Papers with Code platform [3]. The dataset was split into training, validation, and testing sets by patients ids provided within the dataset. When using the datasets we made sure that no overlap of patients across the splits.

TABLE 1. Distribution of Tumor Types in Dataset

Tumor Type	*Num of Samples*	*Percentage*
Meningioma	*708*	*23.1%*
Glioma	*1,426*	*46.5%*
Pituitary	*930*	*30.4%*

This table 1 summarizes the distribution of tumors among the dataset, where it clearly shows a diverse representation of tumor types, which is beneficial when it comes to segmentation and classifications tasks.

B. Data Pre-Processing

Data preprocessing is a crucial step to ensure that the data is compatible with the models being used. The pre-processing part involved the following :

1) Data Loading: Loading the data where the files were parsed using the h5ypy library [4] to extract the MRI images, segmentations masks, and tumor labels. To store this information we used a dictionary-based structure for each sample.

2) Normalization: The pixel intensity values of the MRI images were normalized to a range of (0,1), to ensure the consistency in the input data and make sure that it helps with the model convergence during the training phase.

3) Resizing: All images and masks were resized to 224 x 224 pixels in both the UNet segmentation model and the ResNet50 classification model.

4) Data augmentation: Augmentation techniques were applied during training to prevent overfitting and improve the overall robustness of the model [5] . Some of the techniques used included rotations, flipping and cropping which created additional variations of the training data being used. Training images have a 50% chance of augmentation being applied.

C. Model Architectures

The model architecture consists of two parts one for segmentation and one for classification as described below :

1) U-Net For Tumor Segmentation: The U-Net architecture was employed for the segmentation part, where it is specifically designed for biomedical image segmentation and consists of two main components [6] .

a) Encoder: ResNet50 backbone modified for single channel (gray-scale images). This component extracts hierarchical features using convolutional layer and pooling operations.

b) Decoder: The decoder takes the encoded features and gradually increases their resolution back to the original image size. At each step, it combines upsampled features with corresponding encoder features through attention, helping highlight important areas for accurate tumor segmentation.

The final layer of the U-Net outputs a binary mask including the tumor region within the MRI image. The model we used was trained using binary cross-entropy loss and Dice Loss and optimized along with the AdamW optimizer. For evaluating the models performance we used the dice score to correctly evaluate and assess the model.

2) ResNet-50 For Tumor Classification: For the classification task, we used a pre-trained ResNet-50 model which we fine tuned on the dataset.. This model is called residual connections, where it addresses the vanishing gradient problems and allows taringfs of deep networks [5]. The ResNet-50 architecture was modified as follows:

a) 1st modification: The fully conceded (dense) layer was replaced with a custom layer to output probabilities for three classes: meningioma, glioma, and pituitary.

*b) 2nd modificatio*n: All layers except the final block, were frozen during initial training to retain the knowledge learned from imageNet [7].

c) 3rd modification: Changing the first convolutional layer from 3 channels (RGB) to 1 channel (gray-scale) and maintaining pre-trained weights by averaging out across channels.

Regarding the objective function we used the cross-entropy, with the AdamW optimizer for parameter updates. In order to measure the classification and evaluate the model, certain metrics were used such as accuracy, recall, and F1 Score.

D. Implementation

1) Tools and Frameworks: The UNet segmentation model and ResNet50 classification model were implemented using PyTorch [9] . Supporting libraries such as NumPy [10] and Python Imaging Library [11] were utilized for data preprocessing, and matplotlib along with seaborn were employed for result visualizations. All computations were performed in a GPU-enabled environment to accelerate training and evaluation.

2) Training Setup: For both the segmentation and classification tasks loss functions, optimizers and evaluation metrics were implemented to ensure the training phase is being correctly evaluated as the training phase is a critical part in the model.

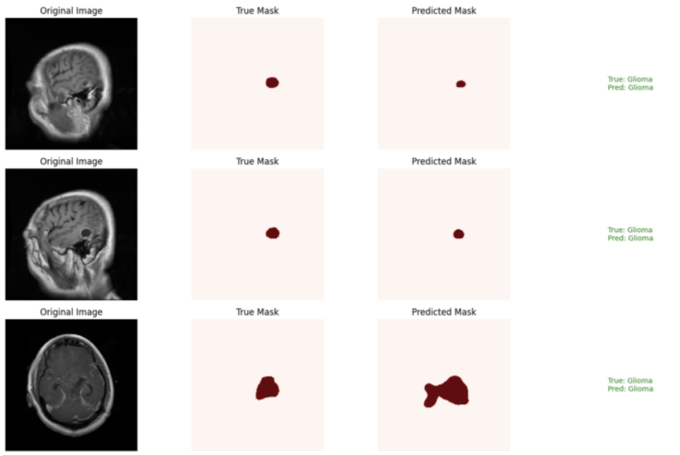
3) Training and Evaluation Workflow: The U-Net model was trained to predict binary segmentation masks from normalized MRI images resized to 224 × 224. To evaluate the quality of the segmentation Dice Scores were used [12] . The ResNet-50 model, fine-tuned with ImageNet pre-trained weights, classified MRI scans into three tumor types: meningioma, glioma, and pituitary. To assess the performance of the classification on the test set. metrics such as accuracy and F1-score were measured in this part [13]. Finally, early stopping was such an important step in the workflow as it made sure that to prevent overfitting while training.

4) Computational Resources: All models were trained using Google Colab's GPU-enabled runtime with 40GB RAM and NVIDIA Tesla A100 GPU. Where it significantly reduced training time, and allowed us to carry out efficient experiments and evaluations.

III. Results

The performance of the developed deep learning pipeline was evaluated on a test set of brain MRI scans.

Figure 1. Test Results Visualizations



A. Segmentation

The U-Net model achieved a final test Dice score of 0.7495, showing a moderate accuracy in delineating tumor regions from MRI images. These results demonstrate that the model can identify and localize tumors in most cases but may have faced some difficulties when it comes to complex or ambiguous tumor boundaries.

Figure 1 showcases examples of original MRI scans, their true masks, and the predicted masks generated by the U-Net model. These examples highlight the ability of the model to approximate the tumor shape and size, though certain cases show under- or over-segmentation.

b. Classification

The ResNet-50 model attained a classification accuracy of 96% on the test set.

Figure 2. Confusion Matrix

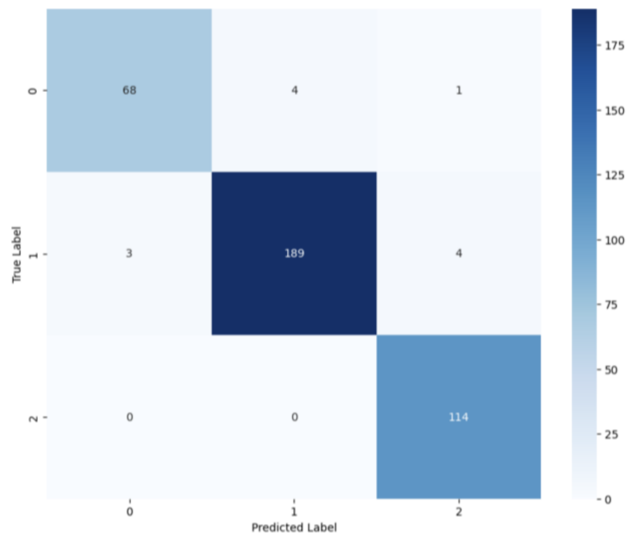
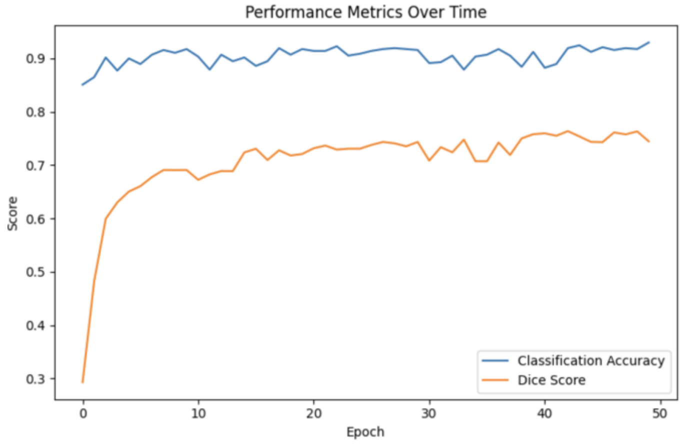


Figure 2 illustrates the classification performance, showing the number of correct and incorrect predictions for each tumor class. The model performed well for gliomas and pituitary tumors but struggled with meningioma cases, as evident from the higher number of misclassifications in the confusion matrix.

C. Training and Validation Performance

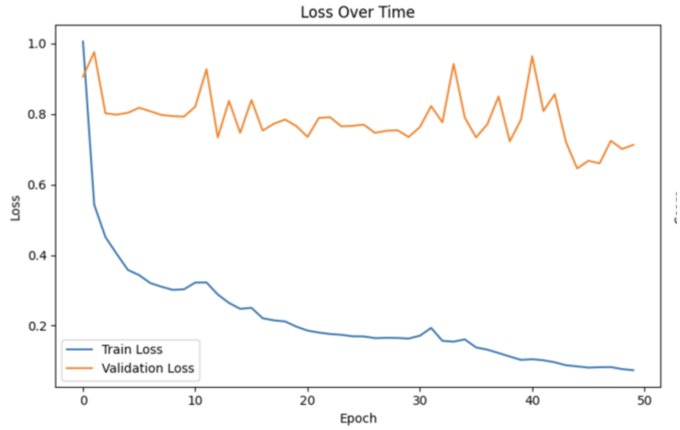
The training and validation accuracy curves (Figure 3) show rapid convergence for both U-Net and ResNet-50, with validation accuracy stabilizing around the 7th epoch.

Figure 3. Accuracy vs. Epoch



The training and validation loss curves as shown in Figure 4 demonstrate consistent learning behavior, though the gap between training and validation losses indicates mild overfitting, particularly for the classification task.

Figure 4. Loss vs. Epoch



IV. Discussion

The results show the possibility of using a combined segmentation and classification approach for brain tumor analysis. However, overall performance points to several lines of improvement in the following respects:

A. Segmentation Challenges

The proposed segmentation model, based on the U-Net architecture, achieved an average Dice score of 0.7495, reflecting a moderate success in delineating tumor regions. This would indicate that the model was able to capture the general shape and location of tumors in most cases but struggled with precision, particularly for complex or irregularly shaped tumors. Poor contrast between tumor and normal tissue, variation in tumor sizes, and several artifacts in MRI images can be some of the limiting factors for the model's performance [14]. As shown in Figure 1, the model generally behaves well for tumors whose boundary is clear but is inconsistent for cases where the edges of the tumor are ambiguous. Under-segmentation, where the region of the tumor is smaller than the ground truth, and over-segmentation, where the model predicts regions outside the tumor, were common errors. These observations suggest that enhancing the dataset with more diverse and high-quality samples could improve the model's generalization and accuracy.

B. Classification Challenges

The ResNet-50 model was able to classify images with an accuracy of 96%, showing a good discrimination capability for the three kinds of tumors: meningioma, glioma, and pituitary [15] .

The example predictions shown in Figure 3 further extend insights into the model's behavior on classification.

Although the model is confident with many correct classifications of glioma and pituitary tumor classes, it often incorrectly classified meningiomas that have very subtle visual variations. This again indicates more data augmentation and additional training on less represented or complicated samples are required.

C. General Observations

The training and validation performance curves for Figures 4 and 5 respectively seem to indicate good convergence with the U-Net and ResNet-50 converging at the

7th epoch. However, a minimal gap between training and validation accuracy and divergence in loss curves pointed out mild overfitting. This behavior has been particularly evident in the classification task of ResNet-50, where the training accuracy went up to almost 100%, whereas the validation accuracy had been plateauing at around 87%. In future work, dropout, weight decay, or early stopping may be used to address overfitting. Also, including more diverse datasets with greater variability in tumor appearance and patient demographics would likely improve the generalization capability of the models across unseen samples.

D. Strengths and Limitations

The combined segmentation and classification pipeline provides a broad framework for the analysis of brain MRI scans. In fact, integrating the two tasks together offers both localization and diagnostic capabilities, thus making this approach very applicable in clinical settings. However, a Dice score of moderate value for segmentation and 96% accuracy in classification shows that optimization is needed to meet high accuracy requirements for segmentation, which medical applications require. Despite these challenges, the results show the potential of this pipeline to support radiologists in the diagnosis of brain tumors. This can enhance diagnostic efficiency and consistency since the system reduces manual efforts in segmentation and automates classification. Future work will look at real-time inference that enables integrating this pipeline into clinical workflows.

V. Conclusions and Future work

This project is a step toward the use of deep learning for automated brain tumor analysis. By incorporating segmentation and classification tasks, the proposed pipeline was able to localize and classify brain tumors with high efficiency. The U-Net model showed a moderately successful delineation of tumor boundaries, as reflected by an average Dice score on the test dataset of 75%. Meanwhile, the ResNet-50 classification model performed quite well by achieving a final accuracy on the test dataset of 96% in differentiating between meningioma, glioma, and pituitary tumors. These results can demonstrate how these deep learning methods will perform when integrated to develop a better diagnosis workflow for medical imaging.

However, applying deep learning in such complex areas raises many challenges. Throughout this project it clearly shows that the segmentation model struggled with tumors that had irregular or ambiguous boundaries, while the classification model faced difficulty differentiating visually similar tumor types, such as meningiomas and gliomas. Future work and iterations of such a concept can use larger and more diverse datasets that will enhance the overall performance of the model. In addition, experimenting with advanced architecture and models may help capture and extract more nuanced features to improve the performance of both the segmentation and classification tasks.

Finally, the extension of the pipeline to multi-modal imaging data, like MRI and CT scans, for a comprehensive analysis of brain tumors can be done [16] . Moreover, interpretability features in the models will increase transparency and trust of clinicians who make critical decisions with such systems. Limitations notwithstanding, this work indicates the potential that deep learning may bring into medical imaging and lays the ground for further studies in analyzing brain tumors.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Dr. Omar the professor of the CPS843 course (Intro To Computer Vision) for the guidance and support throughout this project. We also acknowledge the creators of the publicly available dataset used in this study for making their data accessible, enabling advancements in brain tumor research

REFERENCES

[1]J. Cheng *et al.*, “Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition,” *PLOS ONE*, vol. 10, no. 10, p. e0140381, Oct. 2015, doi: https://doi.org/10.1371/journal.pone.0140381.

[2]A. Mire, Vinayak Elangovan, and S. Patil, *Advances in Deep Learning for Medical Image Analysis*. Informa, 2022. doi: https://doi.org/10.1201/9781003230540.

[3] Papers with Code. "Brain Tumor Segmentation Dataset." Available at: https://paperswithcode.com/.

[4]“HDF5 for Python — h5py 3.12.1 documentation,” *H5py.org*, 2014. https://docs.h5py.org/en/stable/index.html (accessed Dec. 01, 2024).

[5] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, “Differential Data Augmentation Techniques for Medical Imaging Classification Tasks,” *AMIA ... Annual Symposium proceedings*, vol. 2017, pp. 979–984, 2017.

[6] P. Pramanik, R. Pramanik, F. Schwenker, R. Sarkar, and S. H. Khan, “DBU-Net: Dual branch U-Net for tumor segmentation in breast ultrasound images,” *PloS one*, vol. 18, no. 11, pp. e0293615-, 2023, doi: 10.1371/journal.pone.0293615

[7] M. N. Musa, “MRI-Based Brain Tumor Classification using ResNet-50 and Optimized Softmax Regression,” *Jurnal infotel (Online)*, vol. 16, no. 3, 2024, doi: 10.20895/infotel.v16i3.1175

[8] M. Abadi, A. Agarwal, P. Barham, et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: https://www.tensorflow.org.

[9] A. Paszke, S. Gross, F. Massa, et al., "PyTorch: An imperative style, high-performance deep learning library," Advances in Neural Information Processing Systems, vol. 32, pp. 8024–8035, 2019.

[10] S. Harris, K. Millman, S. van der Walt, et al., "Array programming with NumPy," Nature, vol. 585, pp. 357–362, 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2.

[11] G. Bradski, "The OpenCV Library," Dr. Dobb’s Journal of Software Tools, 2000. [Online]. Available: https://opencv.org.

[12]I. G. L. N. A. Artha Wiguna *et al.*, “A deep learning approach for cervical cord injury severity determination through axial and sagittal magnetic resonance imaging segmentation and classification,” *European spine journal*, vol. 33, no. 11, pp. 4204–4213, 2024, doi: 10.1007/s00586-024-08464-7

[13]A. Humphrey *et al.*, “Machine-learning classification of astronomical sources: estimating F1-score in the absence of ground truth,” *Monthly notices of the Royal Astronomical Society. Letters*, vol. 517, no. 1, pp. L116–L120, 2022, doi: 10.1093/mnrasl/slac120

[14]G. Litjens *et al.*, “A Survey on Deep Learning in Medical Image Analysis,” *Medical Image Analysis*, vol. 42, no. 1, pp. 60–88, Dec. 2017, doi: https://doi.org/10.1016/j.media.2017.07.005.

[15]K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *arXiv.org*, Dec. 10, 2015. https://arxiv.org/abs/1512.03385

[16]T. Baltrusaitis, C. Ahuja, and L.-P. Morency, “Multimodal Machine Learning: A Survey and Taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, Feb. 2019, doi: https://doi.org/10.1109/tpami.2018.2798607.