Project Milestone: TumorTrace

William Akins, Jimmy Vu, Chen Zhang
Georgia Institute of Technology
Atlanta, GA 30332

Abstract

We present brain tumor classification of medical images through 3D convolutional neural networks, using reconstructions of a patient's brain. We survey methods for this domain, looking for the architecture that gives the highest accuracy. We fine-tune models like ResNet18, InceptionNet, and finally we attempt to construct our own deep convolutional network. At this point in the project, we have started to build our original architecture, achieving only sub-par results.

1. Introduction

Magnetic Resonance Imaging (MRI) is a critical diagnostic tool in modern medicine. It provides high-resolution images of internal body structures [3]. However, analyzing MRI scans for accurate diagnosis remains to be a challenging task that requires extensive expertise and time. In particular, detecting and classifying abnormalities such as tumors, infections, or degenerative diseases from MRI data can be complicated due to the variations in image quality, patient anatomy, and disease presentation [7]. Misdiagnoses or delays in interpretation can lead to severe consequences, such as ineffective treatments and worsened patient outcomes. This project aims to develop a deep learning-based approach to assist in diagnosing medical conditions from MRI scans. The goal is to improve the efficiency and accuracy of medical image analysis, potentially reducing diagnostic errors and assisting radiologists in clinical decision-making. Our focus will be on classifying MRI scans to detect brain tumors, taking multiple 2D images as input and outputting a classification. By utilizing publicly available MRI datasets, we seek to develop a model that generalizes well across different patient cases.

1.1. Related Works

Previous works have looked into utilizing deep learning as a means to identify disease in patients via MRI imaging, and thus we have taken inspiration from some to see how we could further elevate their processes and methodologies. Other works will often go through with more mining and search for MRI scans and data collection [2], and while this does have the benefit of widening the available set, it also results in drawbacks such as additional and nonstandardized preprocessing techniques having to be performed. Instead we choose to go with a synergized dataset from an online dataset repository, which standardizes the preprocessing step. Exisiting models like YOLOv7 have been applied to this domain with much success [1]. Abdusalomov et. al [1], have shown how combining an object detection model (YOLOv7) with convolutional block attention modules (CBAMs) can yield accuracies above .995. This hints at transformers and attention based modules being a future route of improvement for our domain, but our research is focused on applying purely convolutional networks. Furthermore, our task is classification which is unique from their work on object detection and segmentation of tumors within an image. The work of Horasan and Güneş [5] illustrated the effectiveness of ensemble methods for MRI classification. Using convolutional methods, three models were produced at which point ensemble was employed to make the final classification. Achieving an accuracy greater than 0.9, ensemble is clearly an effective technique that we can consider adding to our research.

2. Technical Approach

To address the classification of MRI scan images for classifying tumors, we will implement a structured, multi-model approach integrating advanced preprocessing and validation techniques.

Our baseline model is a 3D Convolutional Neural Network (CNN), selected for its ability to handle higher-dimensional data [4]. Preprocessing steps will include normalization, 3D scaling, and data augmentation to enhance model robustness. This baseline provides a foundational performance benchmark against which other models will be compared. Similar research has reported baseline F1-scores of approximately 90% for 3D CNNs in medical image classification tasks [5], which we aim to match or exceed.

The second model is the Inception model, chosen for its computational efficiency and scalability. This model

is expected to utilize resources more effectively, reducing training time while maintaining performance. Preprocessing will mirror that of the baseline for consistency.

The third model is a ResNet, ideal for addressing vanishing gradient issues through skip connections [6]. This model is particularly suited for deep architectures common in complex datasets. Similar preprocessing techniques will be applied to maintain uniformity in data preparation.

The first model (3D CNN) will be built from scratch, adding convolutional layers, max pooling layers, etc. as we see fit to increase model accuracy. The latter two models will be fine-tuned; many of the initial layers will be frozen, while we only train the last few fully connected layers of the model for our domain, utilizing the feature extractors already learned by the two models.

We will measure performance using two key metrics: F1-score and accuracy. Accuracy offers an overall performance measure but may be unreliable with class imbalance. Class imbalance is a necessary issue to address for our domain as the majority of MRI's are likely to not have brain tumors in them. F1-score provides a balanced assessment by combining precision and recall, making it robust against skewed class distributions. A confusion matrix will further detail true positive (TP), false positive (FP), true negative (TN), and false negative (FN) outcomes.

To ensure comprehensive evaluation, we will conduct ablation studies to examine the impact of preprocessing techniques, architectural modifications, and hyperparameter tuning. Cross-validation will prevent bias from specific data splits, and analysis of misclassified cases from a confusion matrix will help identify patterns of failure, such as tumor types prone to misclassification. Additionally, we will compare our results against existing benchmarks to assess the effectiveness of our approach in improving model performance.

3. Preliminary Results

The dataset used comes from the Kaggle source as previously mentioned. It is comprised of roughly 3000 data entries, with each being a 512x512 image split across three image planes that represent their respective RGB values. The data set is labeled and each label categorizes whether a given data entry illustrates a magnetic resonance scan of Pituitary, Glioma, or Meningioma tumors. The data set was already compiled into .pickle file formatting, which allowed a simplified loading of data into associated data loader classes. It was also split into an 80-20 percentage to distinguish between the training set and the validation set. Furthermore, each entry had a randomized transformation applied to it in the form of a rotation or reflection across its axis. This gives the data set more variance in its entries in order to give a broader range of samples to select from.

Unfortunately, the way the data set is structured does not allow for the remodeling of individual MRI slices into full three-dimensional representations of a given brain. While some data sets do include individual data entries that are the full three-dimensional representation of an MRI scan, those lack categorical tumor labeling and instead focus on the segmentation positioning, which is not the focus of this research. So instead, we will treat each individual MRI slice as its own data entry, which should still provide enough information to extract an accurate model from.

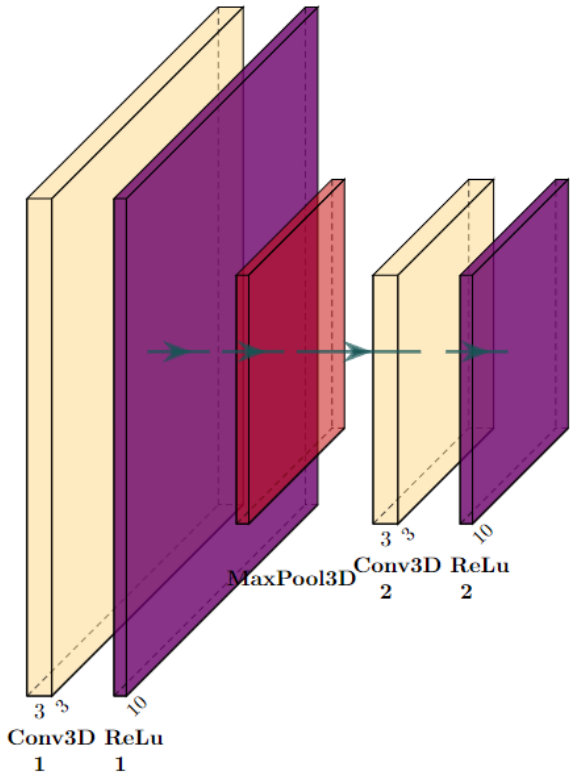

Figure 1. 3D CNN Architecture

As of writing, the model currently being employed is the 3D CNN. This was chosen as the first model to develop as it should provide a solid baseline to evaluate future models from, layering 3D CNNs, max pooling, and ReLU activation functions for a simplified approach. 3D CNNs are also known to be computationally intensive, so addressing it first would allow for more time for experimentation with the model.

Although some aspects of the current implementation show a good starting point for future iterations, there are several adjustments needed to make the 3D CNN reach its potential.

The model was configured with ten total epochs and a batch size of sixteen, with the average running time coming to approximately two hours.

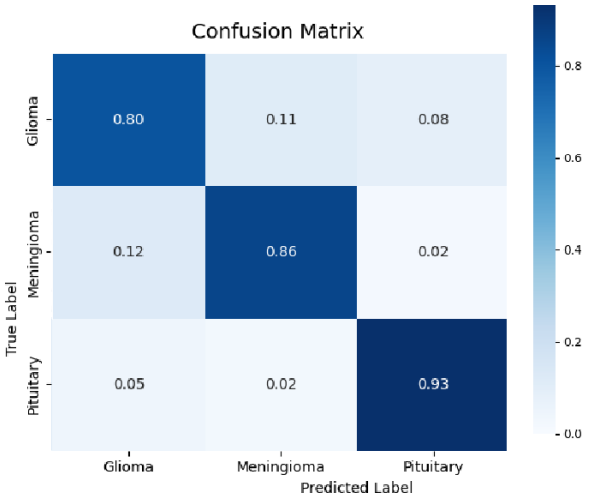

Figure 2. Confusion Matrix

Classification Report:

	precision	recall	f1-score	support
Glioma	0.72	0.80	0.76	143
Meningioma	0.93	0.86	0.89	293
Pituitary	0.90	0.93	0.92	177
accuracy			0.87	613
macro avg	0.85	0.86	0.86	613
weighted avg	0.87	0.87	0.87	613

Figure 3. Classification Report

Our evaluation matrices give insight to the current performance of the model. The Pituitary label comes out with the best recall and F1-score, with Meningioma having the best precision. The margin between the two labels is generally slim, however the results of Glioma has a much lower report, indicating that this label is somewhat difficult to identify for the model. Additionally, most misclassifications stem from labels that relate Glioma.

As seen in the graphs across epochs, training accuracy shows a gradual trend up with losses correspondingly showing a trend downwards, which in isolation would indicate a precise model. However, the validation accuracy is astronomically low with no pattern of increase nor decrease, meaning that our model's weights does not fare well against external data entries. We can draw the conclusion that the current model is suffering from massive overfitting, being hyper fixated on its current data set rather than overarching trends.

This can result from a variety of reasons, from limited data size to hyper-specific features that model fixates on. Either way, the approach to the model must be reexamined thoroughly, as low validation accuracy indicates the the model would not make accurate real-world determinations of tumor classification.

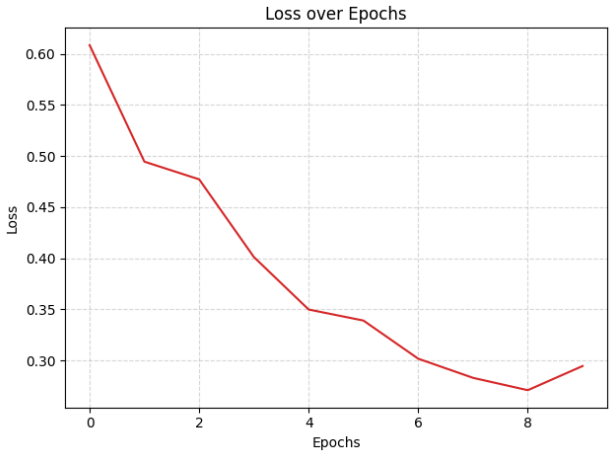

Figure 4. Losses

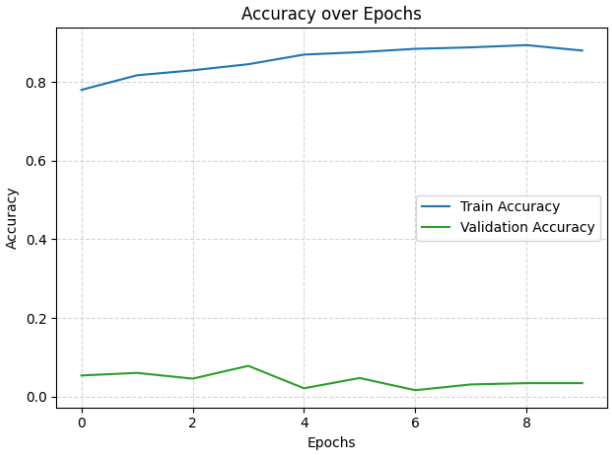

Figure 5. Accuracies

4. Next Steps

We will primarily focus on improving the performance of our 3D convolutional neural network architecture. While the current model offers a promising foundation, we plan to enhance its accuracy by experimenting with the depth of the network, the arrangement of layers, and the configurations of key components like convolutional and pooling layers. We might adjust kernel sizes, change stride values, and introduce more complex layer arrangements to help the model learn richer representations of tumor features across different MRI slices. We will incorporate regularization techniques such as batch normalization, dropout, and weight decay to stabilize training and prevent overfitting.

Meanwhile, we will explore how well ResNet and Incep-

tion architectures perform in this domain. These models are known for their efficiency and strong performance on a variety of image classification tasks. Our goal is to understand how effectively these architectures transfer to MRI-based tumor classification and gain insights into the strengths and limitations of each approach.

All models will be evaluated using multiple metrics. We will track F1-score, precision, recall, and AUC and conduct five-fold cross-validation to get a more reliable picture of performance across the dataset. We will also examine confusion matrices to identify recurring misclassifications and understand where the models are struggling.

Some potential challenges might be the limited dataset size, which could lead to overfitting, and computational constraints during training, especially with deep architectures. To address these, we will rely on data augmentation and regularization to improve generalization. We will manage training time by using cloud resources.

In the next two weeks, we aim to finalize the structure of our 3D CNN and begin testing different configurations. By mid-April, we plan to complete the training and evaluation of the ResNet and Inception models. In the last two weeks of the semester, our focus will shift to performance comparisons, refining the best-performing model, and generating visualizations.

Ultimately, success in our research could be achieving an F1-score higher than 0.90 with at least one of the models, ideally with our custom 3D CNN. We will also deliver clear, reproducible code, well-documented results, and meaningful visualizations.

References

[1] A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo. Brain tumor detection based on deep learning approaches and magnetic resonance imaging. *Cancers*, 2023. 1

[2] I. Ahmad and F. Alqurashi. Early cancer detection using deep learning and medical imaging: A survey. *Critical Reviews in Oncology/Hematology*, 2024. 1

[3] J. Butman. Neuroradiologist offers insight on mri scans. NCI, 2022. Accessed: 2025-03-26. 1

[4] F. J. Dorfner, J. B. Patel, J. Kalpathy-Cramer, E. R. Gerstner, and C. P. Bridge. A review of deep learning for brain tumor analysis in mri. *Nature News*, 2025. 1

[5] A. Horasan and A. Güneş. Advancing prostate cancer diagnosis: A deep learning approach for enhanced detection in mri images. *MDPI*, 2024. 1

[6] D. Lee, J. Lee, J. Ko, J. Yoon, K. Ryu, and Y. Nam. Deep learning in mr image processing. *Investig Magn Reson Imaging*, 23(2):81–99, 2019. 2

[7] L. Pinto-Coelho. How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications. *Bioengineering (Basel, Switzerland)*, 2023. 1