Group 15 Deep Learning Project Report

Antony Shenouda, Anthony Koulakov, Alex Cohen, Elizabeth Telysheva

Introduction:

Brain tumor detection and classification from MRI scans has become a pivotal application of convolutional neural networks (CNNs) in medical image analysis, enabling advancements in diagnostic precision and early intervention. Existing models predominantly focus on binary classification tasks, identifying the presence or absence of a tumor. For example, datasets available on platforms like Kaggle are primarily geared toward binary tumor detection without further distinction among tumor types. However, differentiating between gliomas, meningiomas, and pituitary tumors is crucial for informing treatment strategies, as these tumor types vary significantly in prognosis and therapeutic approaches.

In this project, we developed a CNN-based model to address this limitation by introducing a two-stage pipeline. The first stage detects the presence of a tumor, while the second stage isolates the tumor region and classifies it into one of three types: glioma, meningioma, or pituitary tumor. By combining detection and classification, our approach goes beyond existing binary detection models and offering a more comprehensive diagnostic tool.

Our model leverages transfer learning to fine-tune a pretrained CNN, adapting it to the specialized requirements of brain tumor detection and classification. PyTorch was used for model development, diverging from the TensorFlow-based approaches commonly seen in related work.

This report details the design, implementation, and evaluation of our model, highlighting its contributions to advancing CNN applications in medical diagnostics. Our two-stage model not only achieves accurate detection but also provides detailed classification, offering significant potential to assist healthcare professionals in clinical decision-making.

Datasets:

The datasets used in this project, “Brain Tumor MRI” by Mostafa Hamada and “Brain Tumor MRI Dataset” by Masoud Nickparvar, comprised MRI scans of brains both with and without tumors. Tumor images were further categorized into three types: glioma, meningioma, and pituitary tumors. The detection model dataset was largely balanced, featuring 5,588 non-tumor images and 24,493 tumor images, which were distributed across 7,928 gliomas, 8,900 meningiomas, and 7,665 pituitary tumors. This comprehensive dataset provided a solid foundation for the detection task. Since we planned to test our final classification model on this dataset we decided to reduce it in size and balance it. We used 11,171 images which consisted of 5,588 non-tumor images and 1861 images for each of the three tumor classes (5,588 non-tumor images and 5583 cumulative tumor images). A smaller similarly balanced dataset was used to train the classification model. For the classification model, only images containing tumors were included.

Data Preprocessing:

Data preprocessing was a critical step in this project to ensure high-quality and consistent input for the models. The MRI datasets were first cleaned and structured, with images organized into separate directories corresponding to their respective classes—non-tumor and tumor for the detection model, and glioma, meningioma, and pituitary tumor for the classification model. The tumor directories were trimmed to ensure each class had an equal amount of data. All images were resized to 240×240 pixels to standardize input dimensions and converted to grayscale, reducing computational complexity while preserving critical details for tumor analysis. To improve convergence and ensure consistent input scales, pixel values were normalized to the range [0, 1]. To address class imbalance and reduce overfitting, data augmentation techniques were applied, including random rotations, flips, brightness adjustments, and resizing.

Detection Model:

The detection model is designed to identify whether a brain MRI scan contains a tumor. This model uses the VGG-16 architecture, a convolutional neural network with 13 convolutional layers and 3 fully connected layers. VGG-16 is widely recognized for its robust feature extraction capabilities, particularly in image classification tasks. In our implementation, the model processes grayscale MRI images with an input resolution of 240×240 pixels.

The feature extraction phase involves sequential convolutional layers interspersed with activation functions (ReLU) and max-pooling layers to progressively downsample the spatial dimensions of the input while retaining critical features. These extracted features are flattened and passed through fully connected layers to make a binary classification: tumor or no tumor.

Data augmentation, including random rotations, flips, and brightness adjustments, was applied to enhance the diversity of the dataset, thereby reducing overfitting. The model is trained using a cross-entropy loss function optimized with the Adam optimizer, which adapts the learning rate dynamically for efficient convergence. We also implemented early stopping to prevent overfitting by halting training when the validation loss stagnated.

Classification Model:

Once the detection model identifies the presence of a tumor, the classification model is employed to categorize the tumor into one of three types: glioma, meningioma, or pituitary tumor. This model also utilizes the VGG-16 architecture, adapted for multi-class classification. In contrast to the detection model, this model operates exclusively on images containing tumors.

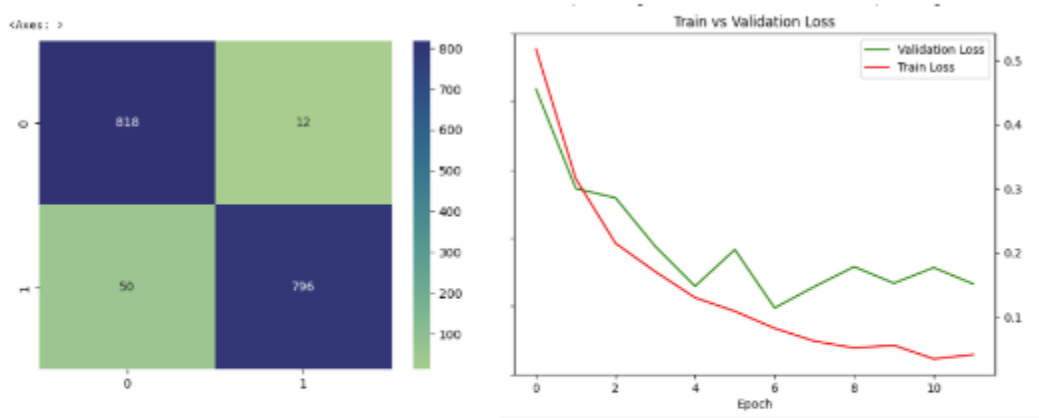
The classification model follows a similar structure to the detection model, with convolutional layers extracting intricate spatial features from the input images. The fully connected layers at the end of the network are tailored for three output classes, corresponding to the three tumor types. We applied a softmax activation function to the output layer to generate probabilities for each class, enabling the selection of the class with the highest likelihood.

Data augmentation was again critical, ensuring balanced training across tumor types despite the relatively smaller dataset. The classification model was trained using a cross-entropy loss function and the Adam optimizer, with additional measures such as early stopping to mitigate overfitting.

Results

Detection Model:

```
Total number of correct predictions: 1614
Accuracy Score:% 96.30071599045345
```

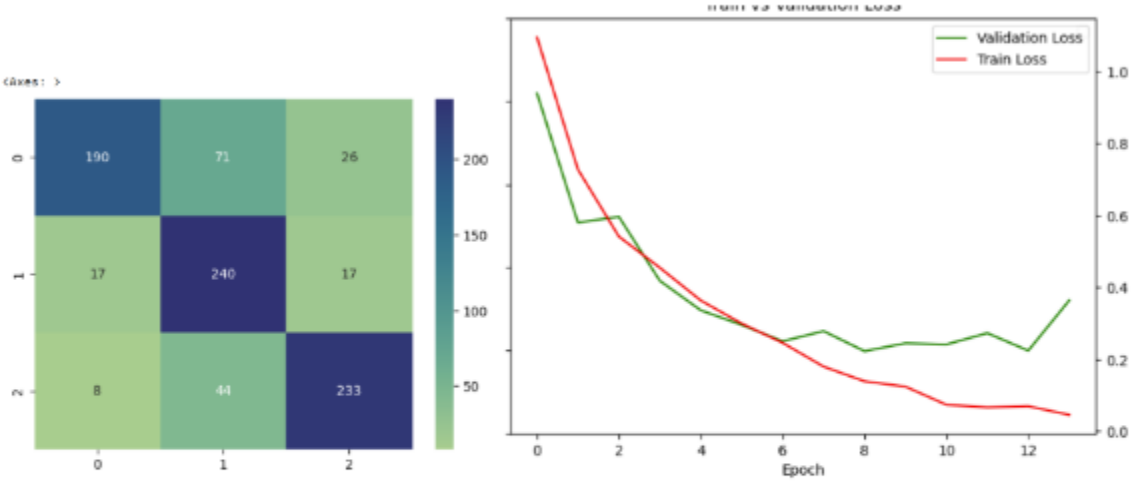


	precision	recall	f1-score	support
0	0.94	0.98	0.96	830
1	0.98	0.93	0.96	846
accuracy			0.96	1676
macro avg	0.96	0.96	0.96	1676
weighted avg	0.96	0.96	0.96	1676

The detection model achieved a performance with an overall accuracy of 96.3%. It demonstrated high precision (0.94 for no tumor and 0.98 for tumors) and recall (0.98 for no tumor and 0.93 for tumors), resulting in strong F1-scores for both classes (0.96 and 0.96, respectively). During the training we found that the validation loss of the model stabilized after epoch 8. This indicates effective learning and good generalization without significant overfitting.

Classification Model:

```
Total number of correct predictions: 663
Accuracy Score:% 78.36879432624113
```



	precision	recall	f1-score	support
1	0.88	0.66	0.76	287
2	0.68	0.88	0.76	274
3	0.84	0.82	0.83	285
accuracy			0.78	846
macro avg	0.80	0.79	0.78	846
weighted avg	0.80	0.78	0.78	846

The classification model achieved an overall accuracy of 78%. Precision, recall, and F1-scores for the three classes were as follows: 0.88, 0.66, and 0.76 for glioma; 0.68, 0.88, and 0.76 for meningioma; and 0.84, 0.82, and 0.83 for pituitary tumors, respectively. The training and validation loss curves for the classification model reach a minimum around epoch 8 but then the validation loss starts to increase, indicating that the model starts to overfit.

Challenges and Future directions:

While our model demonstrates high accuracy, there are specific scenarios where it loses precision. For example, the classification model assumes the presence of a tumor even in cases where the detection model produces a false positive. As a result, it assigns a tumor class to an image that may not belong to any of the predefined categories. Although MRI scans without tumors were excluded during training, such cases may still arise during real-world use, potentially leading to misclassifications. While one potential solution could involve incorporating a 'no tumor' category into the classification stage, this approach may conflict with the streamlined two-stage process. As discussed earlier, such adjustments must be carefully considered to avoid diluting the model's focus.

Another potential issue lies in the dataset distribution for the detection model. While the dataset includes four nearly equal parts—no-tumor, glioma, meningioma, and pituitary scans—the detection model treats all tumor types as a single category. This creates an imbalance in the data, with a 3:1 ratio of tumor images to no-tumor images. As a result, the no-tumor set becomes a minority, which could skew the model's performance by prioritizing tumor detection over non-tumor cases. This imbalance was potentially reflected in the results, as the detection model showed a 3% lower accuracy for the no-tumor case, after a trim of the dataset to where the amount of tumor images was roughly equal to the amount of non-tumor images this difference seems to disappear. Addressing this issue, perhaps through weighted loss functions or data augmentation for no-tumor cases, could improve the model's accuracy and generalization.

While we saw very high accuracy with the detection model there was a drop in the results for the classification model. Potential improvements that could be made to increase accuracy could include increasing brightness and contrast. This would make the data more readable and increase its quality potentially helping with identifying important features. Additionally, adjustments to the size and distribution of the data set would help with any overfitting problems and finetune the results to become more accurate.

What we Learned:

Throughout this project we learned several hard and soft skills. From a technical perspective, we learned how convolutional neural networks (CNNs) can be utilized to analyze and interpret medical imaging data, specifically MRI scans, for diagnostic purposes. Also, we learned how to design and implement a two-stage deep learning pipeline that combines object detection (tumor localization) with multi-class classification (tumor type identification). Lastly, we learned how to evaluate model performance in medical diagnostics using metrics like precision, recall, and F1-score. We also gained several soft skills such as how to work together with a team and coordinate deep learning projects. This includes communicating with one another through messages and meetings, as well as offering constructive feedback to help improve the outcome of the project. Both the hard and soft skills we gained from this project are invaluable in academia and industry and were grateful to have had this experience.

References:

RuslanKL. (n.d.). Brain Tumor Detection v1.0: CNN VGG-16. Kaggle. Retrieved from https://www.kaggle.com/code/ruslankl/brain-tumor-detection-v1-0-cnn-vgg-16#1.1.-Data-Set-Description

Mohamada2274. (n.d.). Brain Tumor MRI. Kaggle. Retrieved from https://www.kaggle.com/datasets/mohamada2274/brain-tumor-mri

National Brain Tumor Society. (n.d.). Brain Tumor Information. Retrieved from https://braintumor.org/

Penn Medicine. (2018, November). Explaining a Brain Tumor Diagnosis. Neuroscience Blog. Retrieved from https://www.pennmedicine.org/updates/blogs/neuroscience-blog/2018/november/explaining-a-brain-tumor-diagnosis

Scientific Animations. (n.d.). Glioblastoma Cancer: Brain. Retrieved from https://www.scientificanimations.com/glioblastoma-cancer-brain/uncategorized/

Masoud Nickparvar. (n.d.). Brain Tumor MRI Dataset. Kaggle. Retrieved from https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset