Brain Tumor Classification - Final Report

Nicolò Vacis
Department of Computer Science
University of Illinois Chicago
nvaci@uic.edu

Giovanni Vaccarino
Department of Computer Science
University of Illinois Chicago
gvacc@uic.edu

Vittorio Palladino
Department of Computer Science
University of Illinois Chicago
vpall@uic.edu

Diana Gontero
Department of Computer Science
University of Illinois Chicago
dgont@uic.edu

Abstract

This project investigates the application of deep learning techniques for automatic brain tumor classification using magnetic resonance imaging (MRI). We addressed a four-class classification problem involving glioma, meningioma, pituitary tumors, and healthy brain images. A MobileNetV2-based transfer learning model was trained and evaluated against a convolutional neural network trained from scratch. Experimental results show that transfer learning significantly improves generalization, particularly in recall, which is critical in medical diagnosis. Furthermore, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to enhance model interpretability by visualizing anatomically relevant regions driving predictions.

1 Introduction

Brain tumors pose a serious threat to human health, often requiring early detection and accurate diagnosis to enable timely intervention and effective treatment planning. Magnetic Resonance Imaging (MRI) is the most commonly used imaging modality for brain tumor assessment due to its superior contrast resolution and ability to capture fine-grained anatomical details. However, manual analysis of MRI scans is time-consuming, subject to human error, and heavily dependent on expert radiologists.

Recent advances in deep learning have enabled automated systems capable of extracting complex patterns from medical images. Convolutional Neural Networks (CNNs), in particular, have demonstrated strong performance in image classification and segmentation tasks. Nevertheless, training deep models from scratch in the medical domain remains challenging due to limited labeled data and high inter-patient variability.

To address these challenges, this project explored the use of transfer learning by fine-tuning a pretrained MobileNetV2 model. In addition to maximizing classification accuracy, we emphasized **recall** and **interpretability**, ensuring that the resulting model is both clinically useful and transparent in its decision-making process.

2 Dataset

We used the **MRI Brain Tumor Dataset** from Kaggle[1]. The dataset contains MRI scans grouped into four classes: glioma tumor, meningioma tumor, pituitary tumor, and no tumor. Images exhibit variations in resolution, contrast, and tumor morphology, reflecting realistic clinical conditions.

The dataset provides separate training and testing splits stored in NPZ format. The training set was further divided into training and validation subsets using **stratified sampling** to preserve class balance. Figure 1 shows the distribution of training samples per class, indicating a relatively balanced dataset that reduces bias during learning.
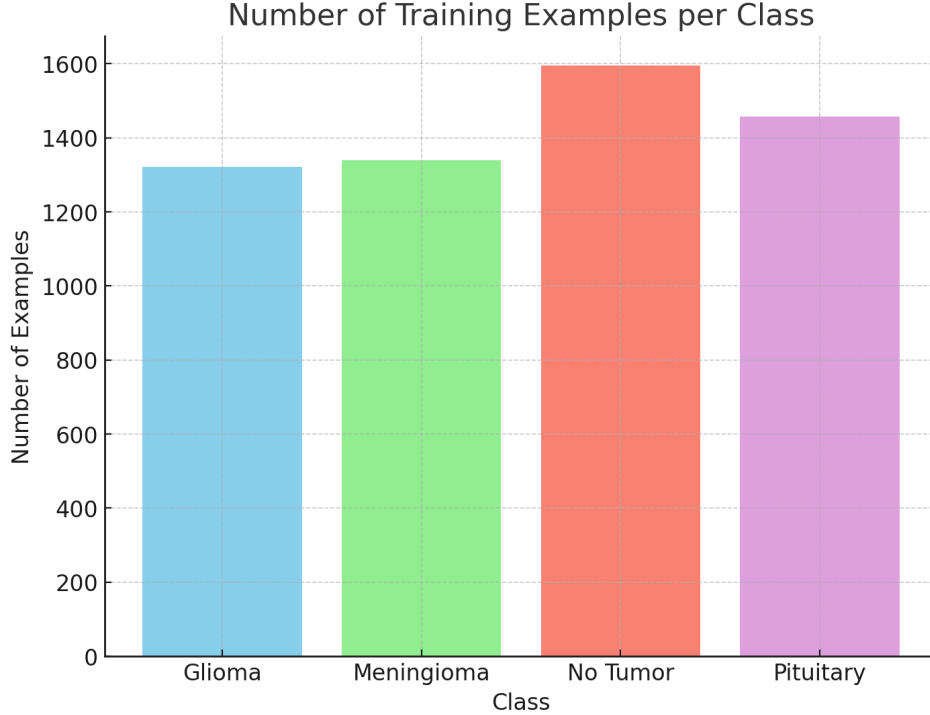

Figure 1: Number of training examples per class in the dataset.

3 Machine Learning Task

The task addressed in this project is a **multi-class image classification** problem. Given an MRI scan of the brain, the model must assign it to one of four categories corresponding to tumor type or healthy tissue. This task is inherently challenging due to overlapping visual characteristics between tumor types and variations in tumor size, shape, and spatial location.

In clinical applications, false negatives are particularly dangerous, as failing to detect a tumor may delay diagnosis and treatment. Therefore, recall was treated as a primary evaluation metric, complemented by precision, F1-score, and overall accuracy.

4 Techniques and Methods

4.1 Data Preprocessing and Augmentation

All MRI images were converted into tensors and normalized using ImageNet mean and standard deviation values to ensure compatibility with pretrained models. Normalization was essential to

---

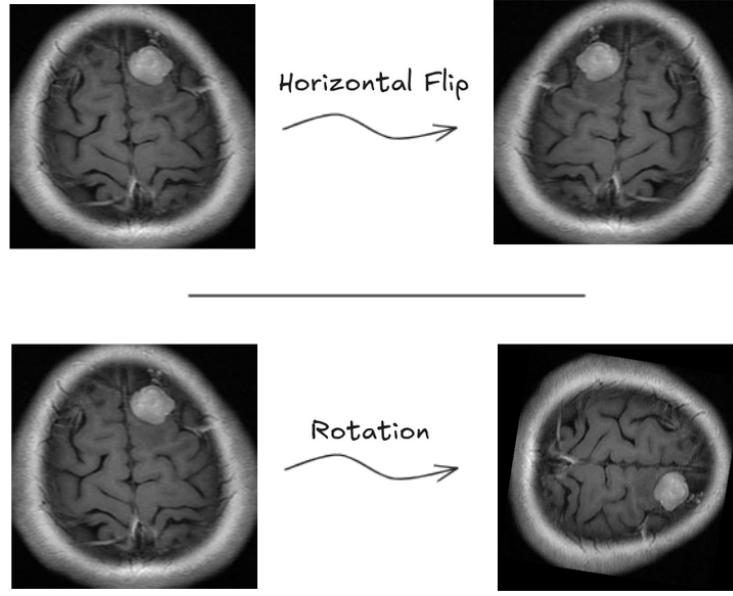[1]`https://www.kaggle.com/datasets/alaminbhuyan/mri-image-data/data`

Figure 2: Examples of data augmentation applied to brain MRI images. Top: horizontal flipping. Bottom: random rotation. These transformations preserve anatomical structure while increasing training data variability.

stabilize training and allow the transfer learning model to effectively reuse learned feature representations.

To improve generalization and reduce overfitting, data augmentation techniques were applied exclusively during training. These included random horizontal flipping and random rotations of up to 20 degrees, as illustrated in Figure 2. Horizontal flipping helps the model become invariant to left–right orientation differences, while rotation simulates variations in patient positioning and scanner alignment. Importantly, these transformations preserve the anatomical structure of the brain and do not alter the semantic meaning of the tumor regions.

By artificially increasing the diversity of the training data, augmentation encouraged the model to learn more robust and spatially invariant features, leading to improved performance on unseen test images.

4.2 Model Architecture

Two architectures were implemented:

- A CNN trained from scratch, consisting of stacked convolutional layers with ReLU activations and max-pooling operations. In Fig. 3 you can see the architecture of the model.
- A MobileNetV2-based transfer learning model, where the convolutional feature extractor pretrained on ImageNet was used as a backbone.

MobileNetV2 was chosen for its computational efficiency and use of depthwise separable convolutions, which significantly reduce parameter count while preserving representational capacity.

4.3 Optimization Strategy and Learning Rate Scheduling

Training was performed using the Adam optimizer with categorical cross-entropy loss. The initial learning rate was set to 1×10^{-5} during fine-tuning, as shown in the implementation. A StepLR scheduler was applied with a step size of 7 epochs and a decay factor of 0.1, allowing coarse updates early in training and finer adjustments later.

3

```
self.features = nn.Sequential(
    nn.Conv2d(3, 32, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(2),

    nn.Conv2d(32, 64, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(2),

    nn.Conv2d(64, 128, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(2),

    nn.Conv2d(128, 256, kernel_size=3, padding=1),
    nn.ReLU(),
    nn.MaxPool2d(2)
)

self.pool = nn.AdaptiveAvgPool2d((1,1))

self.classifier = nn.Sequential(
    nn.Flatten(),
    nn.Linear(256*1*1, 256),
    nn.ReLU(),
    nn.Dropout(0.3),
    nn.Linear(256, num_classes)
)
```

Figure 3: CNN from scratch architecture.

Early stopping was implemented to monitor validation loss and terminate training when no improvement was observed over multiple epochs. This strategy helped prevent overfitting and ensured stable convergence.

5   Experiments

All experiments were conducted using identical training, validation, and test splits to ensure fair comparisons. Model training included detailed tracking of loss and accuracy curves for both training and validation phases.

After training, the best-performing model was selected based on validation accuracy and evaluated on the held-out test set. Evaluation metrics included accuracy, precision, recall, F1-score, and confusion matrix analysis.

6   Results

The MobileNetV2 transfer learning model substantially outperformed the CNN trained from scratch across all metrics. Figure 4 provides class-wise precision, recall, and F1-score.

Class	MobileNetV2 (TL)			CNN From Scratch		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Glioma	0.96	0.94	0.95	0.89	0.82	0.85
Meningioma	0.91	0.91	0.91	0.77	0.79	0.78
Notumor	0.95	0.99	0.97	0.92	0.95	0.94
Pituitary	0.98	0.96	0.97	0.93	0.95	0.94
Accuracy		0.95			0.88	
Macro Avg	0.95	0.95	0.95	0.88	0.88	0.88
Weighted Avg	0.95	0.95	0.95	0.88	0.88	0.88

Figure 4: Overall performance comparison between transfer learning and training from scratch.

The transfer learning model achieved an overall accuracy of 95%, compared to 88% for the CNN trained from scratch. The largest improvements were observed for glioma and meningioma tumors, which are visually similar and harder to distinguish.

Data augmentation played a key role in improving generalization, particularly by reducing sensitivity to tumor orientation and spatial location variations across patients.

7 Model Interpretability via Grad-CAM

To give more interpretability, Grad-CAM was applied to visualize the regions of MRI scans that contributed most to model predictions. Grad-CAM computes gradients of the predicted class with respect to the final convolutional layer, weighting feature maps according to their importance.

Figure 5 shows an example MRI scan and its corresponding Grad-CAM heatmap. The highlighted regions align with tumor locations, demonstrating that the model focuses on clinically meaningful features rather than background artifacts. Additional analyses compared attention patterns for correct and incorrect predictions, further validating the model's behavior.
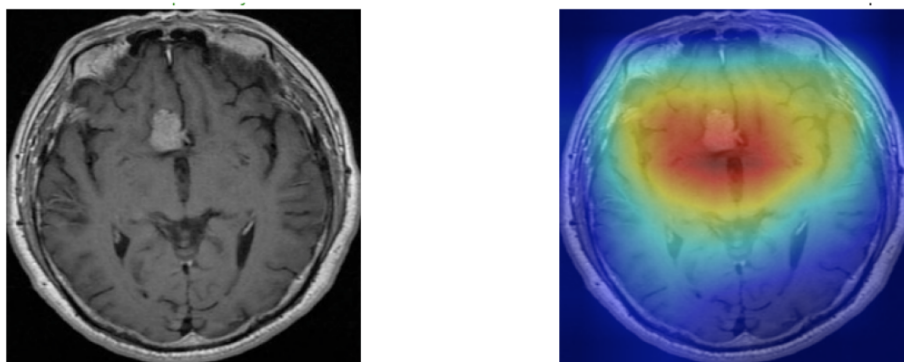

Figure 5: Original MRI image (left) and Grad-CAM visualization (right).

8 Lessons Learned

Through this project, we acquired a better idea of the challenges and best practices of applying deep learning to medical imaging tasks. One key lesson was the importance of transfer learning when working with limited and highly specialized datasets. Training a convolutional neural network from scratch resulted in inferior performance and slower convergence, whereas fine-tuning a pretrained MobileNetV2 model significantly improved generalization and stability.

We also learned that optimization choices, such as learning rate scheduling and early stopping, play a critical role in preventing overfitting and ensuring reliable performance. Careful monitoring of validation loss proved essential, particularly when fine-tuning deep models on medical data. Additionally, implementing Grad-CAM highlighted the value of model interpretability: visualizing attention maps helped verify that the network focused on clinically meaningful regions rather than spurious patterns. Overall, this project reinforced the importance of combining strong quantitative performance with transparency and interpretability when developing machine learning models for healthcare applications.

9   Conclusion

This project demonstrated that deep learning models, particularly those leveraging transfer learning, are highly effective for brain tumor classification using MRI images. MobileNetV2 significantly improved generalization and recall compared to a CNN trained from scratch.

Furthermore, Grad-CAM visualizations provided meaningful insights into the model's decision-making process, increasing interpretability and trust. Overall, this work highlights the potential of interpretable deep learning systems as reliable clinical decision-support tools for medical imaging applications.

References

[1] G. Litjens et al., *A survey on deep learning in medical image analysis*, Medical Image Analysis, vol. 42, pp. 60–88, 2017.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet classification with deep convolutional neural networks*, Advances in Neural Information Processing Systems (NeurIPS), 2012.

[3] A. G. Howard et al., *MobileNets: Efficient convolutional neural networks for mobile vision applications*, arXiv preprint arXiv:1704.04861, 2017.

[4] M. Sandler et al., *MobileNetV2: Inverted residuals and linear bottlenecks*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[5] R. R. Selvaraju et al., *Grad-CAM: Visual explanations from deep networks via gradient-based localization*, Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.

[6] C. Shorten and T. M. Khoshgoftaar, *A survey on image data augmentation for deep learning*, Journal of Big Data, vol. 6, no. 60, 2019.

[7] A. Bhuyan, *MRI Brain Tumor Dataset*, Kaggle, 2021. `https://www.kaggle.com/datasets/alaminbhuyan/mri-image-data/data`