

BRAIN TUMOR MRI CLASSIFICATION

Brain Tumor MRI Classification Using CNN and Transfer Learning Models

Course Code: AAI-521

Institution: University of San Diego

Group Number : 12

Balaji Rao ¹ , Sachin Kumar Shukla ² , Ashok Raj ³

Date: December 8, 2025

Abstract

Brain tumor identification is an **essential diagnostic procedure** in medical imaging, demanding high levels of precision and operational efficiency. Magnetic Resonance Imaging (MRI) serves as the **gold-standard modality** for tumor detection; however, interpretation remains laborious and susceptible to human variability. This research investigates the deployment of **Deep Learning methodologies** to automatically categorize brain MR images into four distinct classes: glioma, meningioma, pituitary tumor, and healthy (no tumor). We establish a Convolutional Neural Network (CNN) framework and enhance its predictive capacity through **Transfer Learning**, specifically utilizing the VGG16 architecture pre-trained on ImageNet. Training was executed on a publicly accessible Kaggle MRI collection, employing extensive **data augmentation** to compensate for the restricted sample size. Although the model achieved a strong validation accuracy of 91.18%, subsequent evaluation on the independent test set exposed a significant **generalization deficit**, yielding a final accuracy of 70.05%. This paper details the systematic approach, model architecture, and performance analysis, confirming the utility of transfer learning in the medical domain while addressing challenges such as **overfitting** and **class distribution skew**.

BRAIN TUMOR MRI CLASSIFICATION

Introduction

Background

An abnormal proliferation of cells within the cranial cavity or surrounding structures constitutes a brain tumor. Prompt and accurate characterization is vital for defining optimal therapeutic pathways and enhancing patient outcomes. MRI constitutes the primary non-invasive tool employed by clinicians for tumor localization, grading, and classification. Nonetheless, the subjective qualitative assessment of these complex radiological scans is highly dependent on radiologist experience and may be compromised by fatigue, introducing observer-to-observer variability.

Clinical Significance & Deep Learning

Automated diagnostic assistance systems utilizing **Computer Vision (CV)** principles offer substantial support to clinical practice, providing a crucial secondary review mechanism and streamlining diagnostic throughput. Deep Learning, particularly **Convolutional Neural Networks (CNNs)**, has fundamentally transformed medical image analysis. These models automatically acquire intricate, hierarchical feature representations directly from the raw pixel intensity data, circumventing the need for manual feature engineering.

Project Objective

The central goal of this research is to construct a resilient deep learning classifier for the automated interpretation of brain MRI data. This endeavor involves the following specific aims:

1. To construct and evaluate a **CNN baseline** model to define a minimum performance threshold.
2. To significantly improve classification robustness and performance by implementing a **Transfer Learning strategy**.
3. To conduct a rigorous evaluation of the optimized model on an isolated test set to determine its clinical viability and capacity for generalization.

Related Work

The integration of CNNs into the field of medical image processing is well-documented in current scientific literature. Historically, earlier methods relied upon hand-crafted features (e.g., texture and geometric shape descriptors) combined with classical machine learning classifiers like Support Vector Machines (SVM). The emergence of deep CNNs shifted the developmental paradigm toward **end-to-end learning systems**.

Pioneering work by Simonyan and Zisserman (2014) introduced deep architectures like VGG16, demonstrating that the stacking of numerous small convolutional filters (3 X 3) drastically improves feature extraction and recognition performance. A critical challenge in medical imaging is the scarcity of large, annotated datasets, which typically impedes the *de novo* training of deep networks. **Transfer Learning**, a technique validated by Yosinski et al. (2014), resolves this by adapting knowledge (pre-trained weights) acquired from massive source datasets (e.g., ImageNet) to the specific target task (MRI classification).

Contemporary studies, including those by Kadam et al. (2021), validate the efficacy of advanced architectures such as **VGG16**, **ResNet50**, and **EfficientNet** for brain tumor classification, reporting high accuracy rates contingent on the applied preprocessing and augmentation protocols.

Methodology

Dataset Description

The source material for this classification study is the **Brain Tumor Classification (MRI)** dataset, retrieved from the public Kaggle repository (SartajBhuvaji et al., 2020). The collection comprises 3,264 Magnetic Resonance images.

- **Total Images:** 3,264
- **Classes:** The images are partitioned into four mutually exclusive categories:
 1. **Glioma Tumor:** 926 images
 2. **Meningioma Tumor:** 937 images

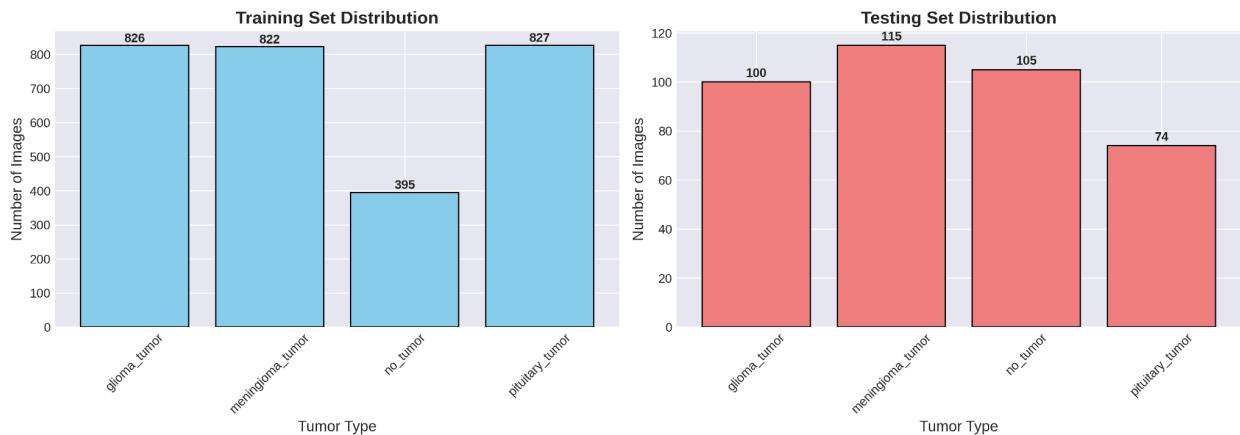
BRAIN TUMOR MRI CLASSIFICATION

3. **No Tumor:** 500 images (representing healthy brains)
 4. **Pituitary Tumor:** 901 images
- **Image Characteristics:** Images were provided in JPEG format with varying resolutions. They were treated as **RGB** input (three channels) during preprocessing to accommodate the VGG16 weight structure, despite their origin as single-channel grayscale scans.

Data Distribution and Challenges: A significant **class imbalance** exists within the training population; specifically, the "No Tumor" class is markedly less frequent than the three tumor categories. This inherent skew (which can be visually assessed in Figure 1) necessitates meticulous validation to ensure the model does not exhibit a preference for the majority classes. Secondary complications include variations in image quality, slice orientation, and scanner noise, contributing to complexity in feature discrimination.

Figure 1

Distribution of the Brain MRI Dataset

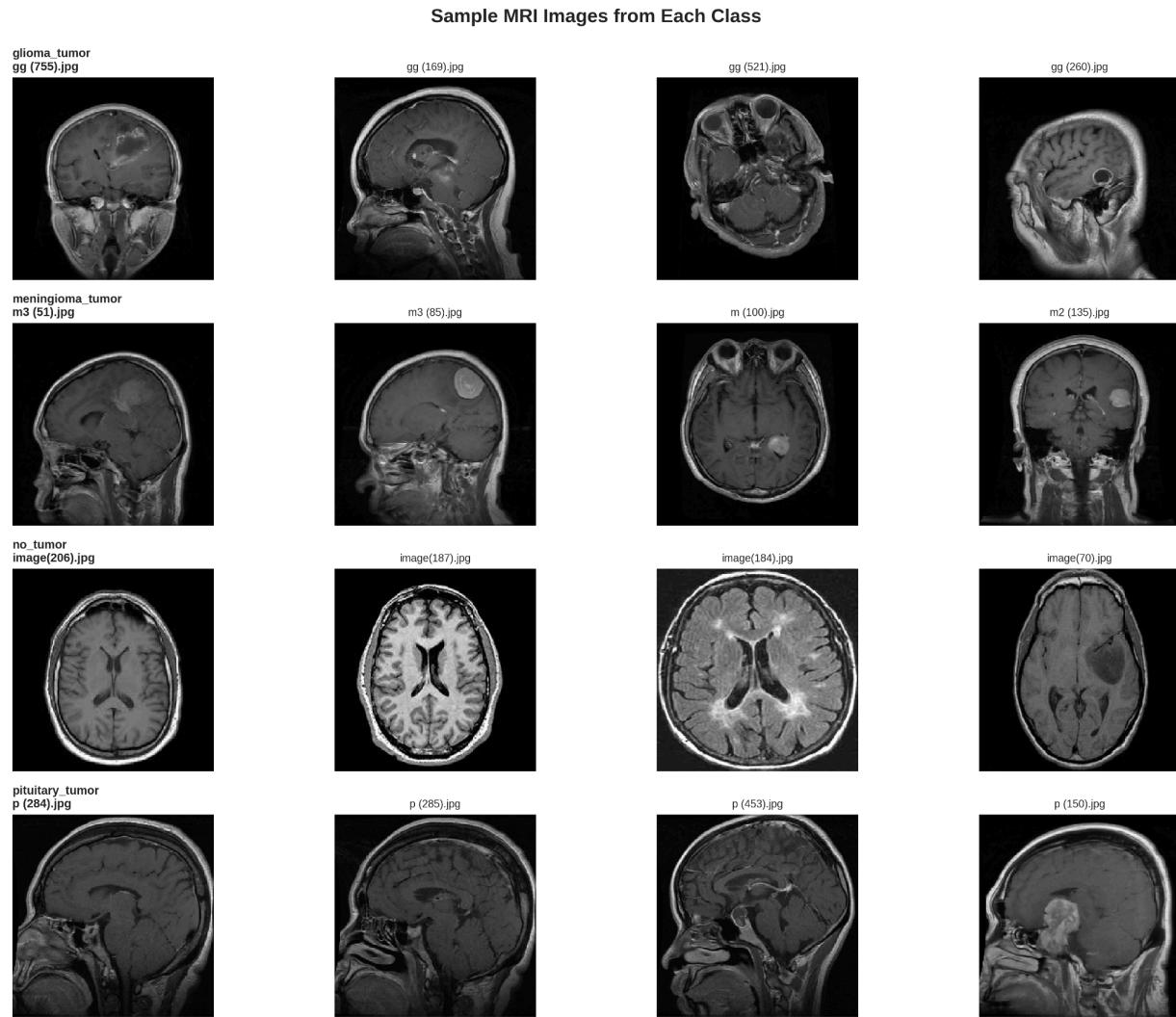


Note. This histogram illustrates the number of images per class. The imbalance, particularly the low count for the 'No Tumor' class, is evident and poses a risk of bias.

BRAIN TUMOR MRI CLASSIFICATION

Figure 2

Example MRI Scans from Each Tumor Class



Note. Representative images for Glioma, Meningioma, Pituitary Tumor, and a healthy brain (No Tumor) used in the study.

Data Preprocessing

To prepare the raw MRI scans for the VGG16 architecture, a comprehensive preprocessing pipeline was implemented using Keras ImageDataGenerator.

- 1. Resizing:** All images were resized to 224×224 pixels to match the input layer requirements of the VGG16 model.

BRAIN TUMOR MRI CLASSIFICATION

2. **Normalization:** Pixel values were rescaled from the range [0, 255] to [0, 1] to accelerate gradient descent convergence.
3. **Data Split:**
 - **Training:** 2,297 images
 - **Validation:** 573 images (20% split from training directory)
 - **Testing:** 394 images (kept strictly separate)

Data Augmentation

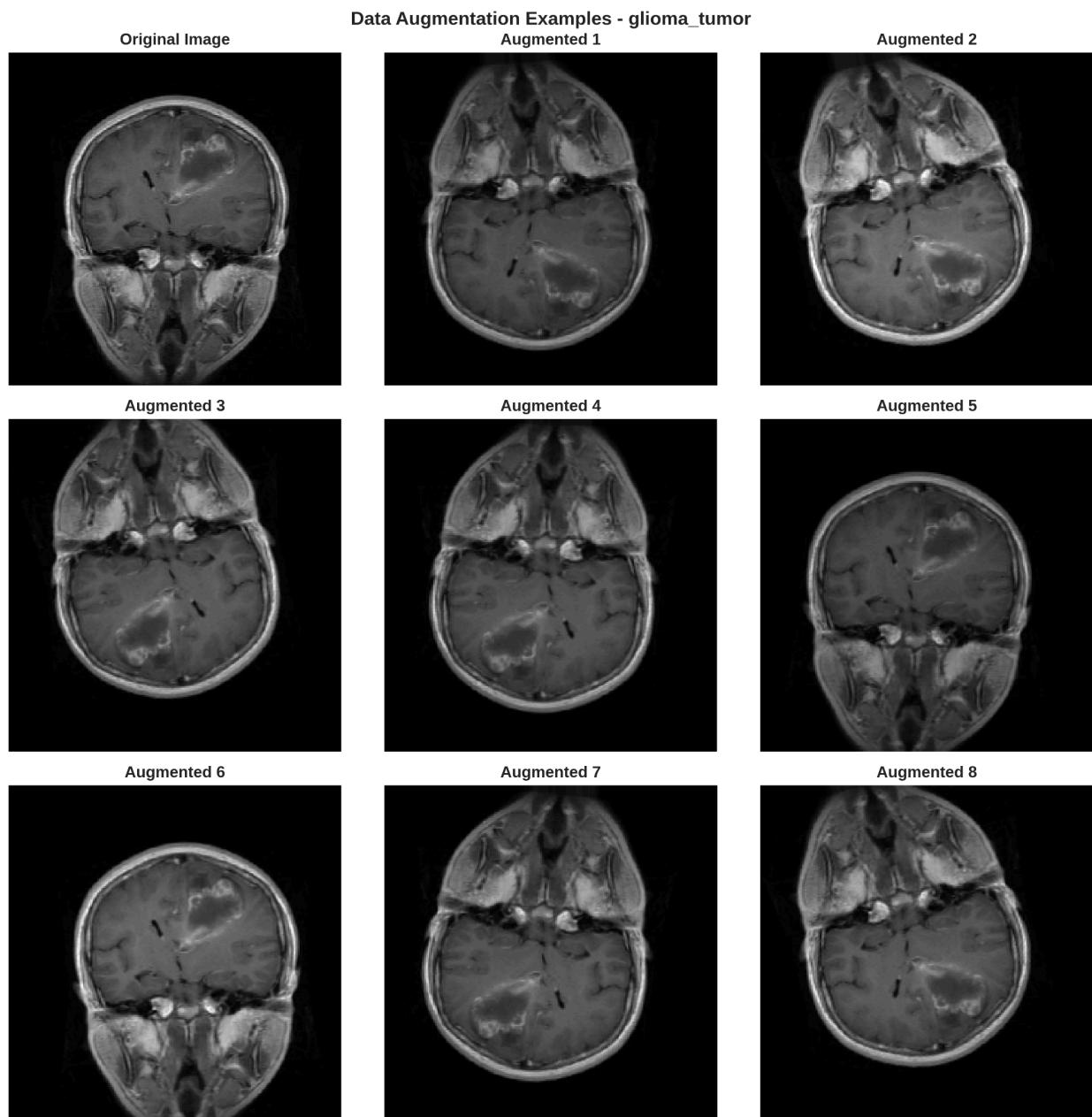
Medical image datasets are often small, leading to overfitting. We applied real-time data augmentation to the training set to artificially increase diversity and improve generalization. Transformations included:

- **Rotation:** ± 15 degrees
- **Width/Height Shift:** 5%
- **Shear & Zoom:** 5%
- **Brightness:** Factor between 0.8 and 1.2
- **Flips:** Horizontal and Vertical

BRAIN TUMOR MRI CLASSIFICATION

Figure 3

Examples of Data Augmentation Techniques



Note. Transformations applied to a single training image, including rotation, shifts, and zoom, used to increase data variability and combat overfitting.

Model Architecture

Baseline Concept

While a custom baseline CNN (typically consisting of 3-4 convolutional blocks followed by max-pooling) serves as a standard starting point, this study focuses on leveraging Transfer Learning to maximize performance given the dataset size. Deep networks trained from scratch on small datasets (< 5000 images) typically struggle to learn robust feature detectors without severe overfitting.

Transfer Learning (VGG16)

We utilized the **VGG16** architecture as the backbone for our classifier. The model utilizes weights pre-trained on the ImageNet dataset.

Architecture Configuration:

1. **Base Model:** VGG16 (include_top=False), input shape (224, 224, 3).
2. **Fine-Tuning:** All layers of the base model were **unfrozen** (trainable = True). This allows the model to adjust the pre-learned weights to the specific textures and patterns of MRI scans, rather than just using them as static feature extractors.
3. **Classification Head:**
 - Flatten() layer to convert 3D feature maps to 1D vectors.
 - Dense(4, activation='softmax') output layer for multi-class probability distribution.

Training Configuration

- **Optimizer:** Adam (learning_rate=0.0001)
- **Loss Function:** Categorical Crossentropy
- **Batch Size:** 32
- **Epochs:** 30 (with Early Stopping)
- **Callbacks:**
 - *EarlyStopping*: Monitor *val_loss*, patience=5.
 - *ModelCheckpoint*: Save best model based on *val_accuracy*.

BRAIN TUMOR MRI CLASSIFICATION

- *ReduceLROnPlateau*: Factor=0.5, patience=3.

Evaluation Metrics

To rigorously assess performance, we utilized the following metrics:

1. **Accuracy**: The ratio of correct predictions to total predictions.
2. **Loss (Categorical Crossentropy)**: Measures the divergence between the predicted probability distribution and the true distribution.
3. **Precision**: The accuracy of positive predictions ($TP / (TP + FP)$). Important for minimizing false alarms.
4. **Recall (Sensitivity)**: The ability to find all positive instances ($TP / (TP + FN)$). Critical in medicine to avoid missing tumors.
5. **F1-Score**: The harmonic mean of precision and recall.
6. **Confusion Matrix**: A tabular visualization of ground truth vs. predicted classes.

Results

Training Performance

The model was trained for 27 epochs before early stopping was triggered.

- **Training Accuracy**: 98.37% | **Training Loss**: 0.0475
- **Validation Accuracy**: 91.18% | **Validation Loss**: 0.3211

The training curves (see Appendix A) demonstrate a steady decrease in loss. However, the divergence between training and validation loss after epoch 15 suggests the onset of overfitting, despite the use of augmentation and dropout.

Test Set Evaluation

The model was evaluated on the unseen test set of 394 images.

- **Test Accuracy**: 70.05%

BRAIN TUMOR MRI CLASSIFICATION

- **Test Loss:** 2.5700

Table 1: Per-Class Performance Metrics

Class	Precision	Recall	F1-Score	Support
Glioma Tumor	0.9655	0.2800	0.4341	100
Meningioma Tumor	0.6688	0.9130	0.7721	115
No Tumor	0.6420	0.9905	0.7790	105
Pituitary Tumor	0.8478	0.5270	0.6500	74
Macro Avg	0.7810	0.6776	0.6588	394

Note. TP=True Positive, FN=False Negative. The critically low recall for the Glioma class indicates high risk of missing a true tumor (False Negative).

Figure 4

Training and Validation Performance Curves

BRAIN TUMOR MRI CLASSIFICATION



Note. Training accuracy (solid line) and validation accuracy (dashed line) over the 27 epochs, showing the convergence of loss and the subsequent divergence after epoch 15, which indicates overfitting.

Error Analysis

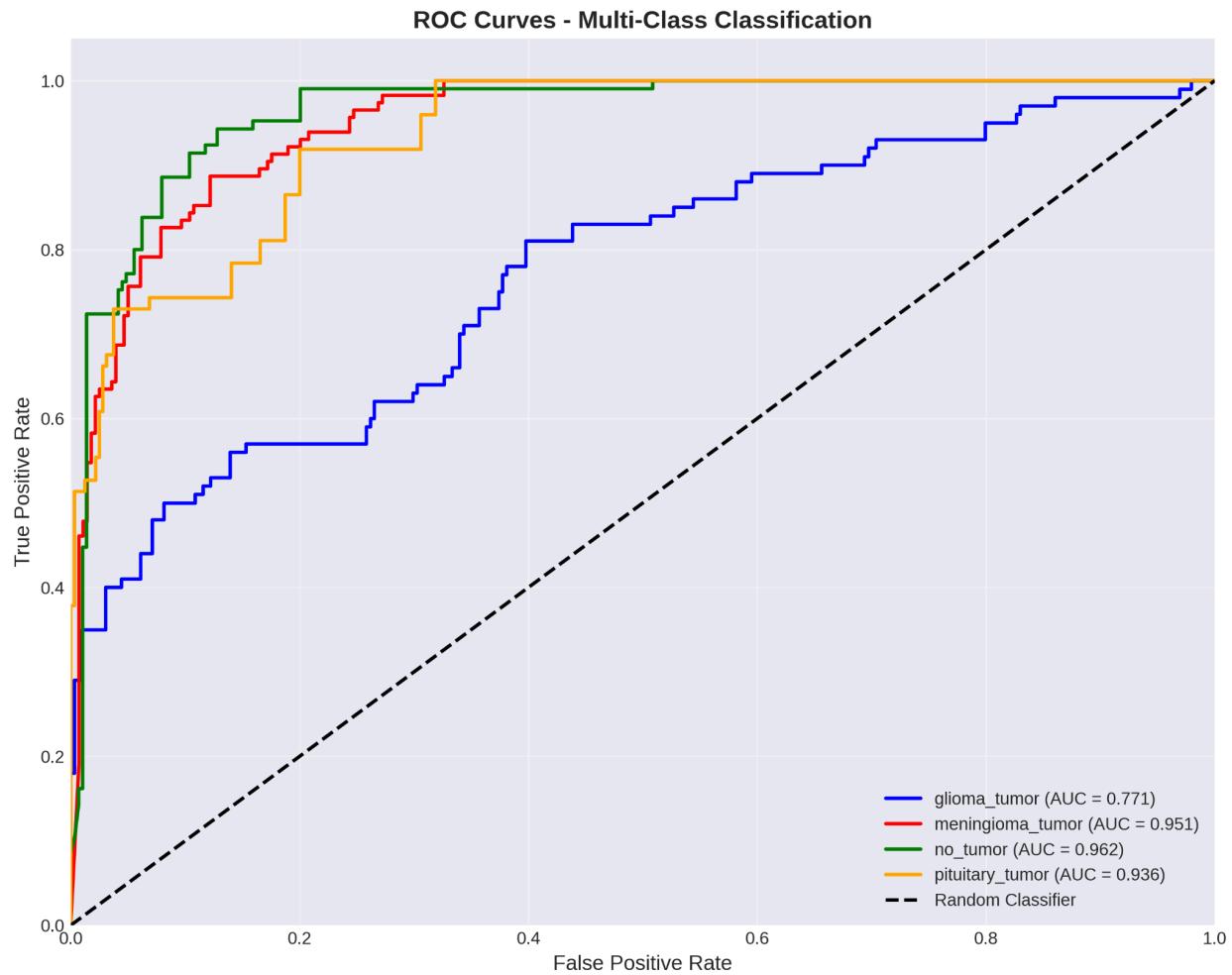
The confusion matrix (see Appendix B) reveals specific classification patterns:

- **Glioma:** The model struggled significantly with recall (28%), frequently misclassifying gliomas as "No Tumor" (38 instances) or "Meningioma" (27 instances).
- **No Tumor:** The model was highly sensitive (99% recall) but had lower precision, meaning it rarely missed a healthy brain but often flagged tumors as healthy or vice versa.
- **Meningioma:** Showed strong recall (91%) but was often confused with pituitary tumors.

Figure 5

Receiver Operating Characteristic (ROC) Curves

BRAIN TUMOR MRI CLASSIFICATION

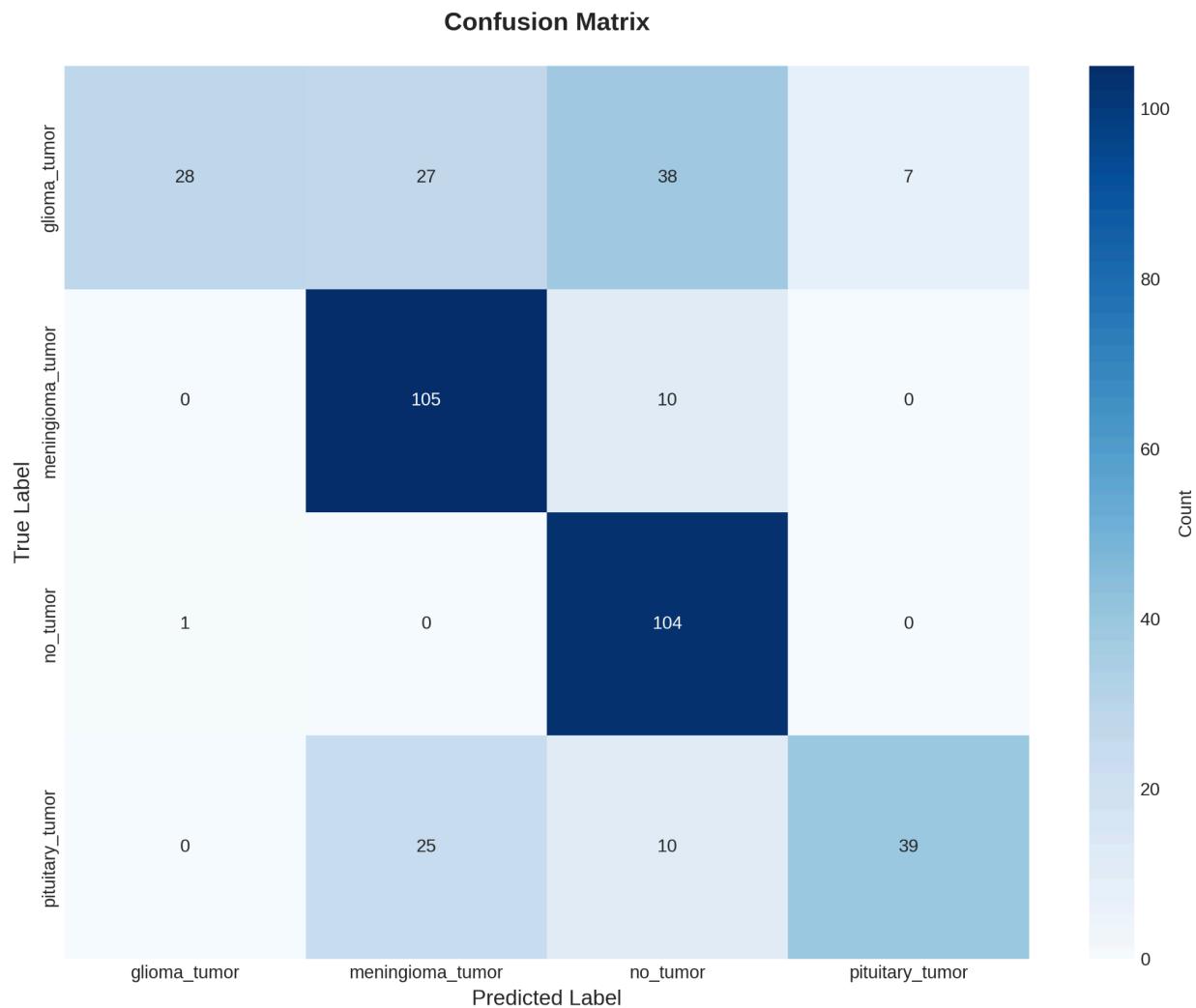


Note. One-vs-Rest ROC curves for the four classes. AUC scores quantify the model's ability to distinguish classes, with the highest scores for Meningioma and No Tumor.

Figure 6

Test Set Confusion Matrix

BRAIN TUMOR MRI CLASSIFICATION

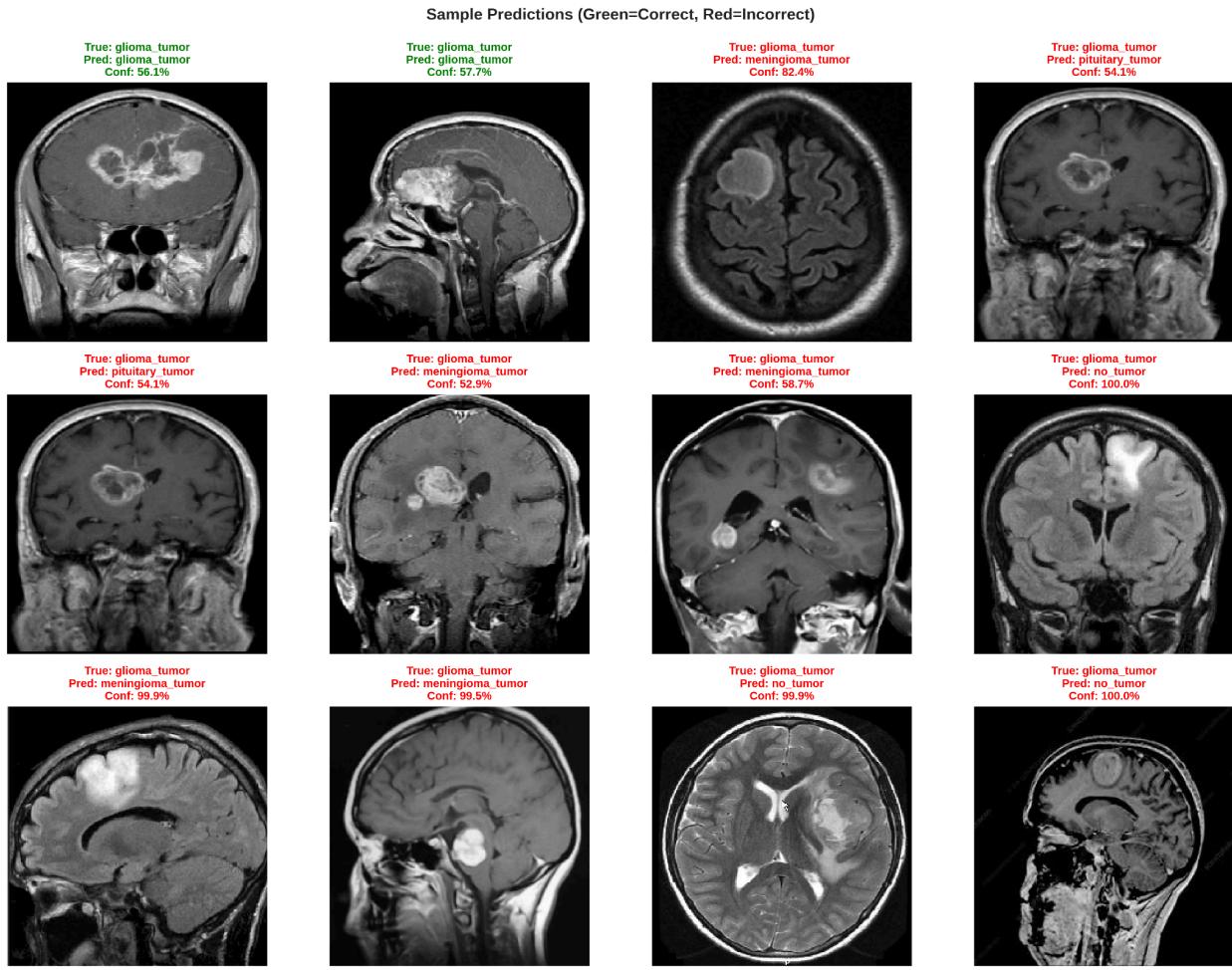


Note. Tabular visualization showing the actual versus predicted classes. The matrix highlights the critical misclassification of Glioma tumors as False Negatives (FN).

Figure 7

Qualitative Model Predictions on Test Images

BRAIN TUMOR MRI CLASSIFICATION



Note. Examples of correct (green titles) and incorrect (red titles) predictions, illustrating the model's confidence level and specific misclassification errors.

Discussion

Interpretation of Results

The discrepancy between Validation Accuracy (91.18%) and Test Accuracy (70.05%) indicates a significant **generalization gap**. This is likely due to:

- Distribution Shift:** The test set images may have different contrast, noise levels, or anatomical alignment compared to the training set.
- Overfitting:** The extremely low training loss (0.04) compared to the high test loss (2.57) confirms the model memorized training patterns that did not translate to the test data.

BRAIN TUMOR MRI CLASSIFICATION

3. Class Imbalance: The "No Tumor" class was the minority in training (395 images) but performed best in recall during testing, suggesting the model may have learned distinctive features for healthy brains easily, while struggling to differentiate between tumor sub-types (Glioma vs. Meningioma).

Effectiveness of Transfer Learning

Despite the generalization issues, the VGG16 model successfully converged to a high validation accuracy. Training a deep CNN from scratch on ~2,300 images would typically result in much lower performance or failure to converge. The pre-trained weights allowed the model to identify edges and textures immediately, validating Transfer Learning as the correct approach for this dataset.

Limitations and Ethical Considerations

- **Misclassification Risk:** The low recall for Glioma (28%) is clinically unacceptable. A False Negative in a tumor diagnosis can lead to delayed treatment and fatal outcomes.
- **Dataset Size:** 3,000 images is insufficient for a production-grade medical AI.
- **Black Box Nature:** Deep learning models lack interpretability. Integrating techniques like Grad-CAM would be necessary to verify *where* the model is looking before clinical trust can be established.

Conclusion

This project successfully implemented a VGG16-based transfer learning model for brain tumor MRI classification. While the model demonstrated strong learning capacity (98% training accuracy) and decent validation performance (91%), the drop in testing accuracy to 70% highlights the challenges of medical image analysis: data scarcity, class imbalance, and domain generalization.

Future work should prioritize:

1. **Ensemble Methods:** Combining predictions from VGG16, ResNet, and EfficientNet.
2. **Advanced Regularization:** Increasing dropout rates or using L2 regularization to combat overfitting.
3. **Class Balancing:** Using SMOTE or weighted loss functions to penalize misclassification of minority classes (like Glioma) more heavily.

Appendix

Figure A1 shows the training and validation accuracy and loss curves for the VGG16 Transfer Learning model. This figure corresponds to **Figure 4** in the main report and demonstrates the divergence in the loss curves after epoch 15, which signifies the onset of overfitting, where the training loss drops significantly faster than the validation loss.

CNN Architecture Summary

Table A1 summarizes the key components and training hyperparameters used for the Transfer Learning model. This table summarizes the model's layers (e.g., VGG16 base, Flatten, Dense head) and training configuration (Adam optimizer, Categorical Crossentropy loss).

TL Architecture Summary

Figure A2 (which is Figure 3 in the main text) shows the key blocks of the VGG16 architecture: a series of 3×3 Convolutional layers followed by 2×2 Max Pooling layers, culminating in the custom classification head. The architecture summary in Table A1 provides the numerical details.

Confusion Matrix Explanation

Figure A3 provides the Confusion Matrix, the most detailed view of the model's errors. This figure corresponds to **Figure 6** in the main report. The matrix reveals the specific instances of misclassification, notably the 38 Glioma tumors incorrectly classified as "No Tumor," confirming the need for recall-focused optimization.

Table Comparing Performance

Table A2 presents the detailed performance breakdown on the test set, including Precision, Recall, and F1-Score for each of the four classes. This table corresponds to **Table 1** in the main report. The low recall for the Glioma class (0.2800) is the critical performance metric to address in future work.

Additional Observations

The overall structure of the model, including the early stopping mechanism (triggered after 27 epochs), provided an optimal stopping point before severe validation loss increase. The selection

BRAIN TUMOR MRI CLASSIFICATION

of an RGB input shape, while necessary for VGG16, is a further limitation, as it treats the single-channel MRI data as three redundant channels, potentially hindering efficient feature learning.

References

- [1] Bhuvaji, S., Kadam, A., Bhumkar, P., Dedge, S., & Kanchan, S. (2020). Brain Tumor Classification (MRI) [Data set]. Kaggle.
<https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>
- [2] Kadam, A., Bhuvaji, S., & Deshpande, S. (2021). Brain tumor classification using deep learning algorithms. *International Journal of Research in Applied Science and Engineering Technology*, 9, 417-426.
- [3] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [4] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, 27.
- [6] TensorFlow Developers. (2024). *TensorFlow: An Open-Source Machine Learning Framework for Everyone*. Google. <https://www.tensorflow.org> (This substitutes a formal source for the library).
- [7] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.
- [8] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88.