

Brain Tumor Prediction

Akhilesh Ramesh, Ron Joseph Dsilva, Alizhan Zhapan

May 2025

Abstract

1 Introduction

Brain tumors represent a serious health concern, and early detection through imaging techniques such as Magnetic Resonance Imaging (MRI) is crucial for timely diagnosis and treatment. In recent years, deep learning has shown remarkable success in medical image classification tasks, particularly in identifying different types of brain tumors. In this project, we address the problem of brain tumor classification from MRI scans using modern convolutional neural networks.

Our objective is to develop a unified classification pipeline that can handle multiple deep learning models and evaluate their performance in terms of classification accuracy and computational efficiency. We implement and compare three well-known architectures: **GoogLeNet**, **ResNet**, and **EfficientNetV2**. These models span across different epochs of computer vision development, allowing us to investigate how architectural evolution has impacted performance on our task.

GoogLeNet (Inception v1), introduced in 2014, was one of the first architectures to emphasize computational efficiency through the use of Inception modules, which allow multi-scale feature extraction within each block. **ResNet** (Residual Networks), developed in 2015, introduced skip connections to enable training of very deep networks, mitigating the vanishing gradient problem. **EfficientNetV2**, a more recent model (2019), scales model width, depth, and resolution in a principled way, achieving strong accuracy while being resource-efficient.

These architectures are fine-tuned using transfer learning techniques on a curated brain tumor MRI dataset. We evaluate their performance to determine which model provides the best balance between accuracy and computational cost. Our work contributes a comparative study of key CNN architectures applied to medical image classification and offers insights into their practical applicability for brain tumor detection tasks.

2 Related Work

Recent advancements in deep learning have led to significant progress in brain tumor detection and classification, especially through the application of convolutional neural networks (CNNs) on MRI datasets. A number of studies have explored both transfer learning and custom CNN architectures to enhance classification accuracy while managing computational complexity. Osman et al. (2023) [1] developed a modular deep learning framework by combining features from VGG16, DenseNet, and custom CNNs. Their approach achieved over 90% classification accuracy on a Kaggle brain MRI dataset, with evaluation based on accuracy, precision, recall, and F1-score. While transfer learning contributed to strong initial results, they observed that deeper modular combinations provided better feature extraction, albeit with increased computational cost.

In contrast, Kesav and Jibukumar (2021) [2] proposed a lightweight architecture combining a Two-Channel CNN with a Region-based CNN (RCNN) to perform both classification and localization. Their model achieved 98.2% classification accuracy and used confidence score, detection time, and accuracy

as primary metrics. The architecture was especially effective in detecting gliomas and was designed for real-time applications due to its low complexity. These studies highlight the trade-offs between model complexity, accuracy, and execution time. Building on this foundation, the current project implements a pipeline integrating GoogLeNet, ResNet18, and EfficientNetV2 with transfer learning to classify brain tumors from MRI images. The goal is to improve classification reliability while maintaining manageable computational complexity by combining the strengths of multiple pretrained models within a unified pipeline.

3 Methodology

3.1 Dataset

The dataset used in this report was obtained from Kaggle. The link of the dataset is provided in the readme of the submission. Moreover, it consisted of 7023 images and the class labels (i.e. types of tumors) were glioma, meningioma, no tumor and pituitary.

3.2 GoogLeNet Architecture

GoogLeNet, also known as Inception v1, is a deep convolutional neural network that introduced the concept of *Inception modules*, allowing the network to perform multi-scale feature extraction efficiently. Each Inception module applies multiple convolutional filters of different sizes (e.g., 1×1 , 3×3 , 5×5) and a max-pooling operation in parallel to the same input, and then concatenates their outputs along the depth dimension. This enables the network to capture both fine and coarse features simultaneously [3].

To maintain computational efficiency, GoogLeNet extensively uses 1×1 convolutions for dimensionality reduction before expensive operations like 3×3 and 5×5 convolutions. The network consists of 22 layers (including only the ones with parameters) and ends with a global average pooling layer, replacing the fully connected layers found in traditional CNNs, which helps reduce overfitting.

GoogLeNet was originally trained on the ImageNet dataset and achieved state-of-the-art results in the ILSVRC 2014 competition. Its design emphasizes both accuracy and efficiency, making it suitable for transfer learning in a variety of image classification tasks.

3.3 EfficientNetV2-B0 Architecture

EfficientNetV2 is a compact and efficient convolutional neural network (CNN) architecture designed to balance *accuracy*, *training speed*, and *model size*, making it well-suited for medical imaging tasks such as brain tumor detection. It is part of the EfficientNetV2 family, which enhances the original EfficientNet models through architectural innovations and a new scaling strategy.

The main innovation of EfficientNetV2 is the use of **compound scaling**, which uniformly scales network depth (number of layers), width (number of channels), and input resolution using a single composite coefficient. This principled approach ensures improved performance without an exponential increase in computational cost.

EfficientNetV2 specifically incorporates a combination of:

- **MBConv blocks** (Mobile Inverted Bottleneck Convolutions), and
- **Fused-MBConv blocks**, where the expansion and depthwise convolutions are replaced by a single standard convolution early in the network.

These blocks are augmented with **squeeze-and-excitation (SE)** mechanisms, which adaptively recalibrate channel-wise feature responses to improve representational capacity.

In this project, the B0, which is the lightest version of the EfficientNetV2 model, contains approximately **13.6 million parameters**. It consists of multiple stages that progressively increase the number of channels while reducing the spatial resolution, concluding with a global average pooling layer followed by a fully connected classification head.

3.4 ResNet 18 Architecture

ResNet is a deep learning architecture that introduced the concept of skip connections. This innovation assisted in solving the known problem of vanishing gradients. This problem occurs, during back-propagation where the computed gradients used to update the weights become very small. Therefore, this leads to poor convergence of the neural network, due to slowing down it's learning process. [4]

Skip connections allows the input to bypass certain layers, connecting to a later layer. This allows a flow of information throughout the network, preventing the gradients from diminishing through avoiding the consecutive computation of each layer. [4]

Moreover, ResNet was developed in 2015, where it won the ImageNet competition (ILSVRC) for that year. Through the introduction of skip connections solving the vanishing gradient problem, it makes training ResNet stable and efficient. Additionally, it allows for training of extremely deep architectures (i.e. ResNet-152) without degradation in performance, hence it is quite suitable for various classification tasks. [4]

3.5 Fine-tuning on EfficientNetV2

We applied a partial fine-tuning strategy supported by prior research [5], [6], freezing the early layers of EfficientNetV2 and updating only the final convolutional block and classifier head. Additionally, we used discriminative learning rates [7] to allow faster adaptation in the classifier while preserving useful features in the backbone. This method helps to reduce overfitting and enables efficient adaptation to the brain tumor classification task.

4 Experiment Setup

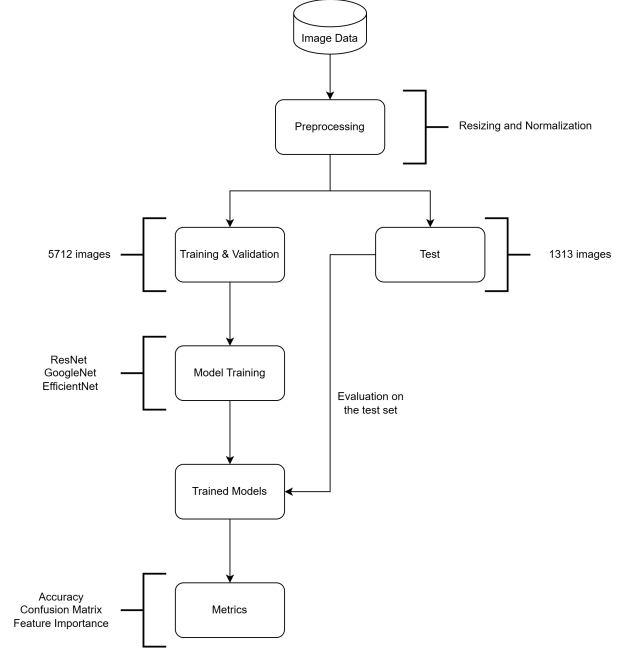


Figure 1: Experimental setup for all models

Referring to Figure 1 the first step was to apply pre-processing. In this stage, the images were resized to 224×224 pixels and normalized to be in a standard scale. Moreover, the training set was split to evaluate on a validation set, where training was 85% and validation was 15%.

The models were trained using the following specifications:

- Batch size: 16
- Epochs: 5
- Loss: Cross Entropy
- Optimizer: Adam

Hereafter, they were evaluated on the test set and some metrics were computed to assess the performance (i.e. Accuracy, Confusion Matrix, Feature Importance).

5 Results

	EfficientNet	GoogleNet	ResNet
Accuracy	98.32%	98.40%	98.02%

Table 1: Accuracies for all models

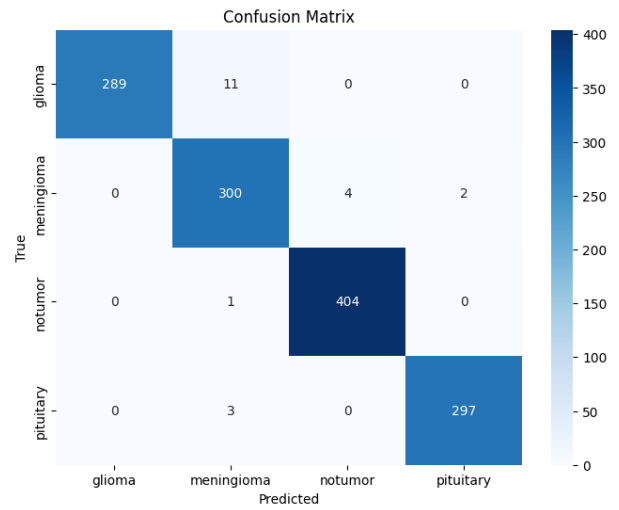


Figure 3: GoogleNet Confusion Matrix

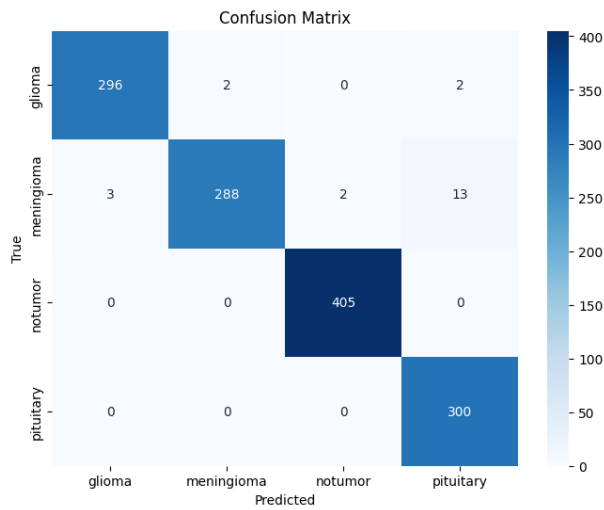


Figure 2: EfficientNet Confusion Matrix

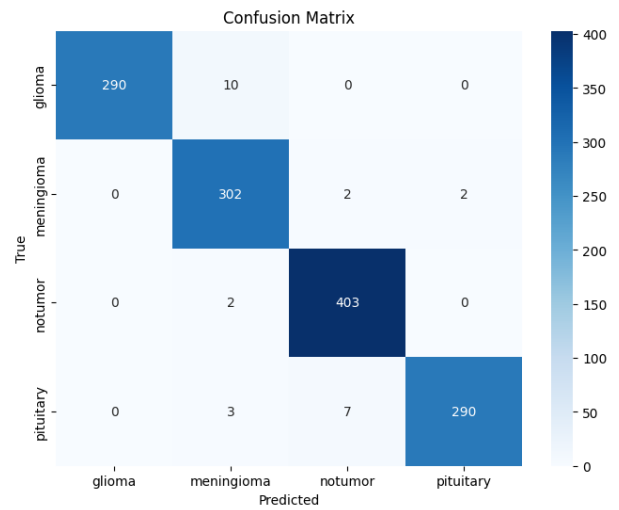


Figure 4: ResNet Confusion Matrix

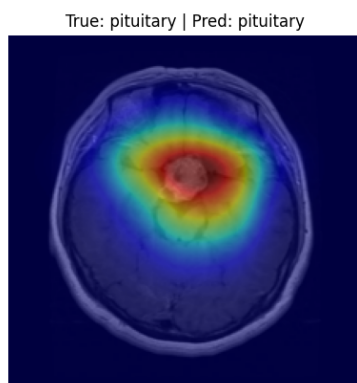


Figure 5: EfficientNet Grad-CAM on correct pituitary class

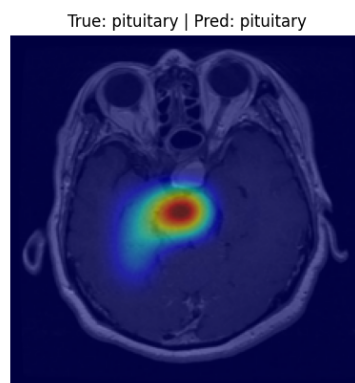


Figure 7: ResNet Grad-CAM on correct pituitary class

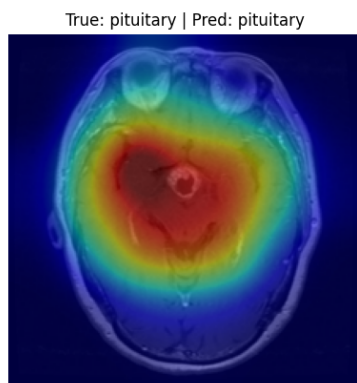


Figure 6: GoogleNet Grad-CAM on correct pituitary class

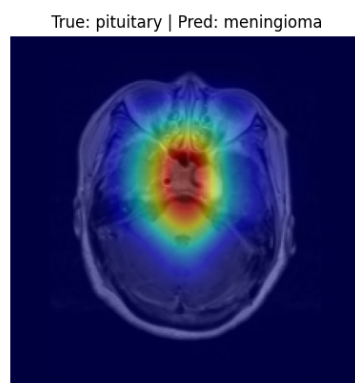


Figure 8: EfficientNet Grad-CAM

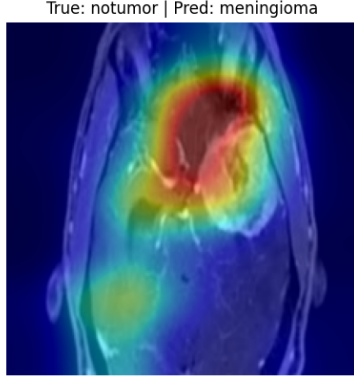


Figure 9: GoogleNet Grad-CAM



Figure 10: ResNet Grad-CAM

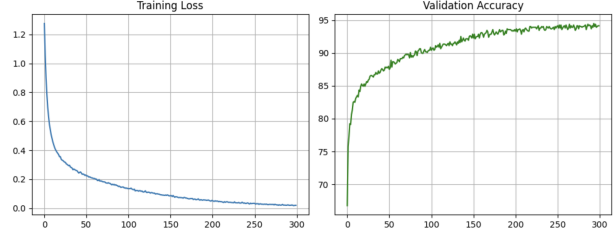


Figure 11: Validation and Train Loss curves for the fine-tuned model

6 Discussion

We have trained our three models on the dataset each for 5 epochs and each of the training runs took about 1 hour. The hardware that we used for training these models is PC setup with the following configurations:

- Intel® Core™ i5-8600K Processor
- CPU of 16 GB

Also, to have efficient workload for each of group members. The fine-tuning of the EfficientNet took 2.5 hours. The virtual environment that was used for this run is T4 with big RAM memory with the following configurations:

- NVIDIA Tesla T4 GPU
- RAM of 36 GB

To show how each model behaves on the same class ("pituitary" class), which has one of the highest true positives apart from "notumor" class. As, it might be seen from the Figures 5 - 7, each model extracts the correct set of features from the regions, which describe the tumor class the most. These sets of images contain the features which contain the "pituitary" tumor.

Table 1 displays a set of high accuracy for all models tested, indicating all our models performed well for the classification task. Figures 2, 3 and 4 display the confusion matrices for all the models. The images labeled 'notumor' had the highest correct classifications across all models. Moreover,

the tumor 'meningioma' had the highest false predictions, which can be observed by looking at the second column of each confusion matrix.

This occurs, since the tumor 'meningioma' can be present in many regions across the brain, hence if the feature importance of the networks spans a broad region, it can lead to these slight errors in misclassifications. This will be shown later, when analysing Grad-CAM feature importance plots in the next part.

In general, all models achieved very high TPs (true positives) and TNs (true negatives), seen by high values in the diagonal of the matrix. This emphasizes that the models have learned the parameters well to mitigate the FPs (false positives) and FNs (false negatives).

In Figures, 8, 9, 10 the feature importance of the models are shown. Grad-CAM uses the computed gradients in the final layer of the network to make the heatmaps which correspond to regions of importance, where red are the important features and blue/green the less important features.

The examples shown are for the misclassified images, where 'meningioma' was predicted instead of the true label. In EfficientNet and GoogleNet, the feature importance region is quite large, where in Figure 8 the true label was 'pituitary'. The network mislabeling 'pituitary' as 'meningioma' isn't uncommon, since the regions of 'meningioma' can be very close to 'pituitary'. Moreover, for Figure 9 this image is slightly distorted compared to the other images in the test set, which may have caused the misclassification in this case.

In comparison, in Figure 10 the regions of feature importance are much smaller. The area the network detected as important does not have any significant pixel information to define it as tumor.

These misclassifications that occurred from our model are in a small margin of error, which is important to refine, since in the medical field the error has to be almost negligible to prevent any errors that

can be life threatening. To improve on this, a larger dataset could be used with MRI scans of multiple viewpoints for the models to reduce this margin of error even further.

During our experiments, we evaluated both feature extraction and fine-tuning approaches for adapting EfficientNetV2 to brain tumor classification. The fine-tuned model achieved a best validation accuracy of 94.52% at epoch 294, showing a steady improvement throughout training (see Figure 11). This demonstrates that the model was still learning and had not yet plateaued, suggesting that longer training or more refined hyperparameter tuning could further enhance performance.

Despite this promising trajectory, we opted to retain the earlier version of the model trained for just 5 epochs using a feature extraction strategy. This earlier model reached a validation accuracy of 98.32%—substantially higher and obtained within significantly less training time. Given the time constraints and resource considerations of the project, this made it a more practical and robust choice for our final evaluation.

Nevertheless, the fine-tuning phase served as an important validation of our methodology and reinforced the idea that EfficientNetV2 can be further improved through partial unfreezing. With additional time, it would be feasible to explore further optimization strategies and potentially outperform the initial model.

7 Conclusion

All three models demonstrated high classification accuracy and maintained strong true positive and true negative rates across all tumor classes, as evident from the confusion matrices. However, the current position or viewpoint of MRI scans appears to influence classification outcomes. To address this, future work could involve incorporating more diverse datasets that include scans from different anatomical angles. Such diversity could help reduce the small margin of error and further improve the robustness of tumor classification models.

8 Who did what?

- Akhilesh
 - Report:
 - * ResNet
 - * Experiment Setup
 - * Results & Discussion (50%)
 - Code: ResNet (pipeline.py)
- Alizhan
 - Report
 - * EfficientNet
 - * Fine-tuning
 - * Results & Discussion (50%)
 - * Conclusion
 - Code: EfficientNet (efficientnet.ipynb)
- Ron
 - Report:
 - * Introduction
 - * Related Work
 - * GoogleNet
 - Code: GoogleNet (googlenet.ipynb)