

Department of Electrical and Computer Engineering
North South University (NSU)

CSE 445: Machine Learning
Section 6

Project Report

Instructor: Dr. Mohammad Mahmudul Alam

Semester: Summer 2025

Title: Brain Tumor Detection using MRI Images	
Name: Md. Minhajul Islam	Name: Nur Ibne Kawsar Zitu
NSU ID: 2211022042	NSU ID: 2212021042

Brain Tumor Detection using MRI Images

Abstract:

This study presents the development of a deep learning-based framework for classifying brain tumor types using Convolutional Neural Networks (CNNs). The project began by training a standard CNN and exploring ensemble learning and XGBoost techniques to enhance model performance. However, it was found that XGBoost did not perform as well as expected, lowering the accuracy of the ensemble of baseline CNN from 92.47% to 86.19%. This highlighted the limitations of using XGBoost with CNN features for this task. We implemented a separable CNN architecture to improve classification performance, incorporating a focal loss function to address class imbalance. This model outperformed the previous ones, achieving an accuracy of 93.31%. The dataset used in this project consists of 3064 MRI images from the Figshare Brain MRI Dataset, categorized into three brain tumor types: glioma, meningioma, and pituitary tumors. Various techniques such as data augmentation, dropout, and batch normalization were employed during training to optimize model performance. The model's efficacy was evaluated using key metrics including accuracy, F1-score, and Receiver Operating Characteristic (ROC) curves. The ensemble of three separable CNNs, utilizing soft-vote aggregation, achieved the highest performance with an accuracy of 94.14%, highlighting the effectiveness of focal loss in enhancing classification accuracy. These results demonstrate that CNN-based models offer a highly effective approach to medical image classification, especially when combined in an ensemble.

Introduction:

Brain tumors are one of the most significant medical conditions that impact individuals worldwide. Early and accurate detection is critical for effective treatment, yet the process remains challenging. Diagnosing brain tumors typically involves radiologists manually reviewing MRI images, which is not only time-consuming but also susceptible to human error. Integrating automated systems to assist with medical image analysis has gained significant attention. Deep learning, particularly Convolutional Neural Networks (CNNs), has proven to be highly effective in image classification tasks, offering a promising solution for the automated classification of brain tumor types from MRI scans.

This project aims to develop an automated system for classifying brain tumors using deep learning. Specifically, the study explores different CNN architectures and leverages techniques like separable convolutions, ensemble learning, and focal loss to enhance the model's accuracy and robustness. The focus of the project is to classify MRI images into three brain tumor categories: **glioma**, **meningioma**, and **pituitary** tumors. The approach aims to reduce the reliance on manual labor while improving diagnostic accuracy. This work ultimately seeks to create a solution that can support radiologists and clinicians, aiding in the decision-making process to diagnose and treat brain tumors.

Problem Statement:

The accurate classification of brain tumors is crucial for effective treatment planning. However, current methods still rely heavily on manual analysis of MRI images by clinicians, which is both time-consuming and prone to human error. In many cases, it is difficult to identify subtle patterns

within medical images, especially when dealing with large datasets. Consequently, there is a need for automated and reliable systems that can assist in classifying brain tumor types with high accuracy and efficiency.

This project addresses the problem by leveraging deep learning techniques, particularly Convolutional Neural Networks (CNNs), to automate the classification process. The goal is to design a model that can accurately categorize MRI scans into three tumor types (glioma, meningioma, and pituitary) and potentially assist clinicians in making faster and more accurate diagnoses. The problem is further complicated by class imbalance in the dataset, where certain tumor types are underrepresented compared to others, making it challenging to train a model that generalizes well. This project also tackles the challenge of class imbalance by implementing techniques like focal loss and class weighting.

Methodology

This project followed a systematic approach like data preprocessing, model development and evaluation to overcome the brain tumor classification problem.

A. Dataset

The dataset we used in this project is the Figshare Brain MRI Dataset which consists of 3064 T1 weighted contrast enhanced MRI slices collected from 233 patients. These images belong to three tumor types: glioma (1426 slices), meningioma (708 slices), and pituitary tumors (930 slices). The dataset was initially provided as a single compressed file which has four sub-zip files with 766 slices each. The dataset also has the 5-fold cross-validation indices (cvind.mat) to facilitate model evaluation. The original dataset is provided in MATLAB .mat format, where each file contains an image along with its class label (1: meningioma, 2: glioma, 3: pituitary tumor), patient ID, the raw MRI image, a manually annotated tumor border, and a binary tumor mask indicating the tumor region.

To prevent patient-level data leakage, To prevent patient-level data leakage, we split the dataset by patient ID into training (70%), validation (15%), and test (15%) sets using a group-aware stratified split, ensuring that images from the same patient did not appear in multiple splits.

B. Data Preprocessing

We convert the .mat files into .png format and resized all images to 224x224 pixels. The labels were one-hot encoded for categorical classification. The training set consists of 2048 images while the validation and test sets contained 538 and 478 images respectively. We calculated the class weights to address the class imbalance in training data {0: 0.785, 1: 1.346, 2: 1.017} for glioma, meningioma, and pituitary tumors, respectively. We applied a data augmentation pipeline including horizontal flips, rotations, zooms and translations to the training set. Finally, we normalized all images to the [0,1] range to improve model generalization. For some experiments we created a balanced training dataset by oversampling the smaller classes to match the largest class. This ensured that the model could learn more effectively from underrepresented tumor types, such as meningioma.

C. Hyperparameter Tuning and Loss Function

We carefully adjusted several hyperparameter to optimize the model's performance. The Adam optimizer was used with a learning rate 0.0003 for ensuring stable convergence through the experimentation. we applied the dropout to the fully connected layers with rates of 0.4 and 0.3 to reduce the risk of overfitting while L2 regularization with a factor of $1e-4$ applied on the dense layer to penalize large weights. we selected a batch size of 32 to balance training efficiency with model convergence. All of these hyperparameter choices were determined manually based on performance on the validation set, without performing an exhaustive grid or Bayesian search.

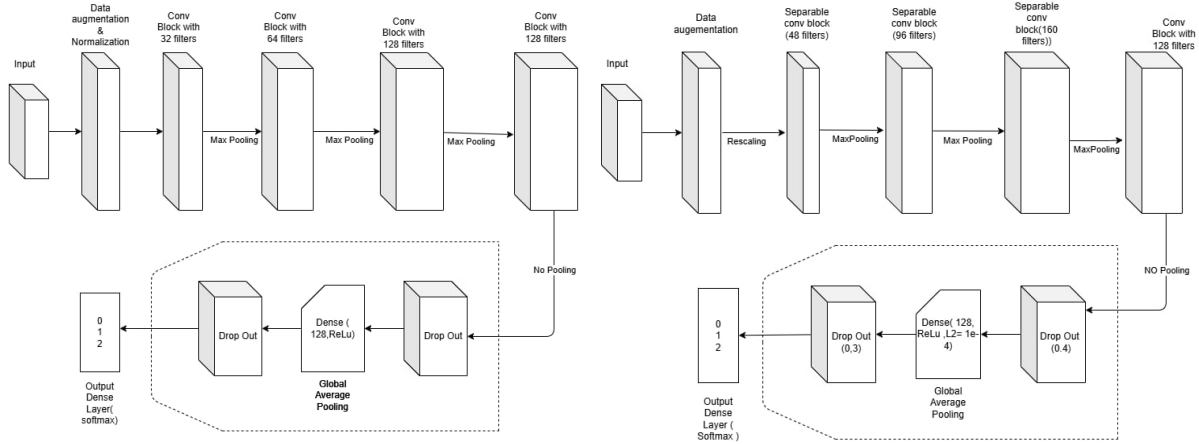
To address class imbalance more effectively, **sparse categorical focal loss** was used as the model's loss function. The focusing parameter gamma was set to 2.0, and class weights (alpha) were derived from the inverse frequency of each class in the training data. This allowed the model to pay more attention to harder-to-classify examples, improving overall performance on minority classes.

D.CNN Architectures

We designed and trained two distinct convolutional neural networks (CNNs) architecture. Both model are designed to extract hierarchical features from images but they differ in their type and block design .

1. Deeper Custom CNN (Standard Conv2D)

We uses a standard convolutional layers organized into back to back convolutional blocks with light residual connections. Each block consists of two 3×3 convolutional layers followed by batch normalization and ReLU activation. We also applied Residual connections within each block, and if the input and output channel dimensions differ, a 1×1 convolution adapts the residual to match. We applied Max-pooling in most blocks to reduce spatial dimensions while preserving essential features. The network begins with data augmentation and normalization to improve generalization, then progresses through convolutional blocks with increasing filter sizes (32, 64, 128), extracting features ranging from low-level edges and textures to high-level patterns. After the convolutional layers, global average pooling condenses the feature maps, and fully connected layers with dropout are applied to reduce overfitting before the final softmax output predicts the tumor class.



Deeper Custom CNN (Standard Conv2D)

Depthwise Separable CNN

2. Depthwise Separable CNN

We implemented a Depthwise separable convolutions to achieve efficient feature extraction with fewer parameters than before. In each separable block spatial filtering applied independently to each input channel. After that a 1×1 convolution the per channel outputs into a better feature representation. After each convolution we applied Batch normalization and ReLU activation. also when the input and output channel dimensions don not match residual connection applied. This architecture uses wider blocks with increasing filter sizes (48, 96, 160) to capture more complex features, and max-pooling is applied in the first three blocks to reduce spatial dimensions. After the convolutional blocks, global average pooling is performed followed by dropout with rate 0.4, then fully connected layers with 128 neurons and ReLU activation followed by dropout with rate 0.3. The network concludes with a softmax layer to predict the tumor category. The use of separable convolutions allows this network to achieve computational efficiency and lower memory usage, while the residual connections maintain feature integrity and facilitate gradient flow during training.

E. Ensemble Methods

We explored ensemble strategies for further improvement of classification performance. First, we implemented a standard CNN and trained the model with three different seeds, aggregating their predictions. We got a good accuracy of 92.47%. We then combined the ensemble with XGBoost for better improvement, but this approach did not improve performance and overall decreased the accuracy to 86.19%.

As the accuracy decreased, we focused on the Depthwise Separable CNN (CNN2). While the single CNN2 performed a good accuracy of 93.31%, we trained the same model with three different seeds and applied soft voting across their predictions. This approach gave a notable outcome of 94.14% accuracy. The separable convolution architecture, combined with ensemble averaging, contributed to better generalization and stability in predictions compared to single models or the standard CNN ensemble.

Results

We evaluated the performance of our models by using accuracy, macro F1-score, weighted F1-score, confusion matrices and ROC curves. The evaluation was performed on the test set containing 478 images across three classes: glioma, meningioma, and pituitary. The single separable CNN achieved a good accuracy of **93.31%** with strong performance on glioma and pituitary classes. But the meningioma showed slightly lower precision in single separable CNN. The corresponding confusion matrix is shown in Figure 1.

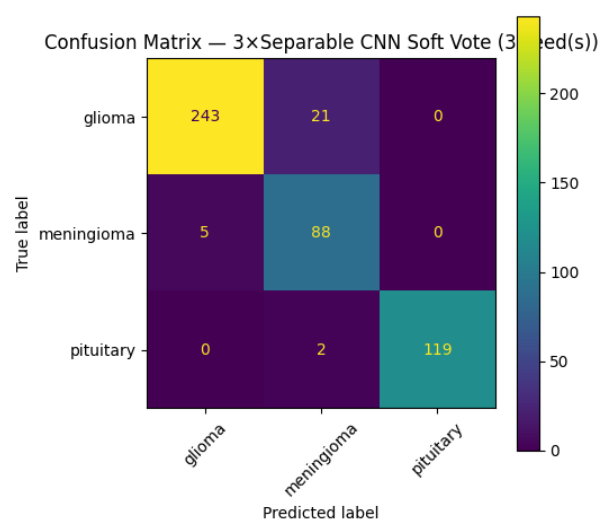
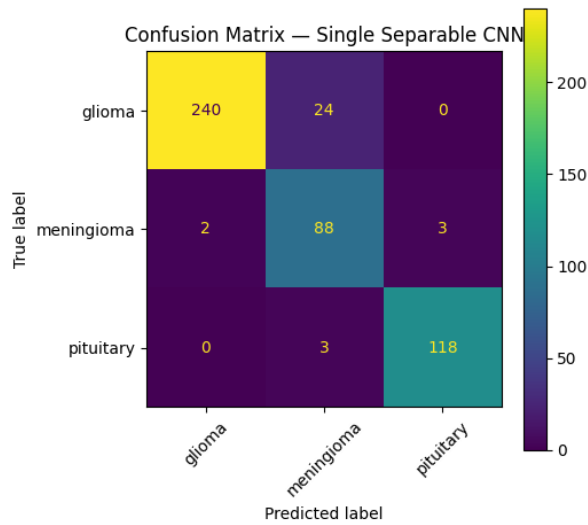


Figure 1: Confusion Matrix for Single Separable CNN

Figure 2: Confusion Matrix for 3-Seed Separable CNN Ensemble

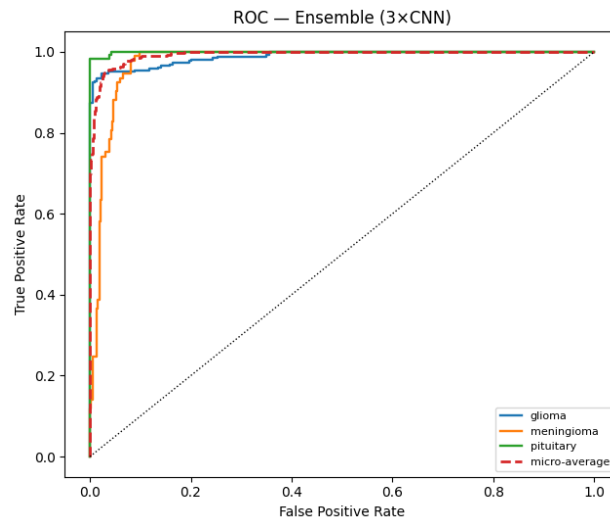


Figure 3: ROC Curve for 3-Seed Separable CNN Ensemble

The standard CNN (Depthwise CNN) trained with three different seeds achieved an accuracy of 92.47%. Though it performed well overall but the meningioma class had lower recall (0.7957) compared to glioma (0.9545) and pituitary (0.9587). This indicates that the standard CNN

ensemble struggled with this class. Also adding XGBoost as a stacked ensemble decreased performance to **86.19%**, demonstrating that naive stacking did not improve the model.

We trained the separable model with three different seeds and applied soft voting across their prediction. This ensemble achieved the highest accuracy of **94.14%**. This also effectively improve predictions for meningioma while maintaining excellent results on glioma and pituitary. The confusion matrix for this ensemble is shown in Figure 2, and the ROC curves illustrating class separability are presented in Figure-3. The summary of model performance is shown in Table-I.

Table-I

Model	Accuracy	Macro-F1	Weighted-F1
Single Separable CNN	0.9331	0.9233	0.9354
Deeper Custom CNN Ensemble (3 seeds)	0.9247	0.9111	0.9243
Deeper custom Stacked CNN + XGBoost	0.8619	0.8292	0.8618
Separable CNN Ensemble (3 seeds, soft voting)	0.9414	0.9345	0.9431

These results indicate that the soft-voting ensemble of separable CNNs provides better generalization and stability compared to single models or standard CNN ensembles. This also overcomes the gaps observed in meningioma predictions. The combination of depthwise separable convolutions and ensemble averaging contributed to efficient feature extraction and robust classification performance.

Conclusion

We developed and evaluated multiple CNN-based models such as a standard CNN, a deeper CNN, and a Depthwise Separable CNN. Our results show that the Depthwise Separable CNN performs best, especially when trained with three different seeds and combined using soft voting. This approach achieved the highest performance with an accuracy of 94.14%, outperforming both the standard CNN and the stacked XGBoost ensemble. The separable convolution architecture, along with ensemble averaging, improved generalization, stabilized predictions, and reduced computational cost.

Despite this success, challenges in accurately classifying the meningioma class remained due to its relatively lower recall in some experiments. Future improvements could involve exploring advanced attention mechanisms, more diverse data augmentation, or incorporating multimodal data such as MRI metadata to further enhance classification performance and robustness.

Dataset: <https://doi.org/10.6084/m9.figshare.1512427.v8>

Github: <https://github.com/Minhajul-Islam-Rimon/Brain-Tumor-Detection-using-MRI-Images-.git>

