Distributed Deep Learning Framework for Brain Tumor Classification

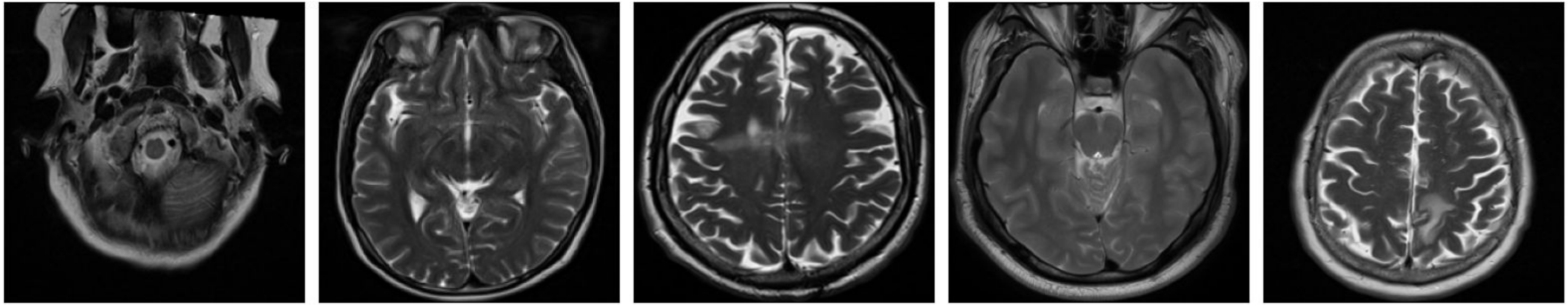
Anna Huang, Xinyi Lyu, Letitia Su, Vicky Yeh

Introduction

- The motivation behind this project stems from the critical need for early and accurate diagnosis of brain tumors, which is essential for optimizing treatment outcomes.
- Magnetic Resonance Imaging (MRI) is a pivotal tool in diagnosing brain tumors due to its ability to provide detailed images of brain structures.
- This project leverages advanced machine learning technologies, specifically Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), to automate and enhance the accuracy of brain tumor classification from MRI scans.
- We aim to identify the most effective pre-trained models for classifying the different types of brain tumors, with an emphasis on speed and accuracy.

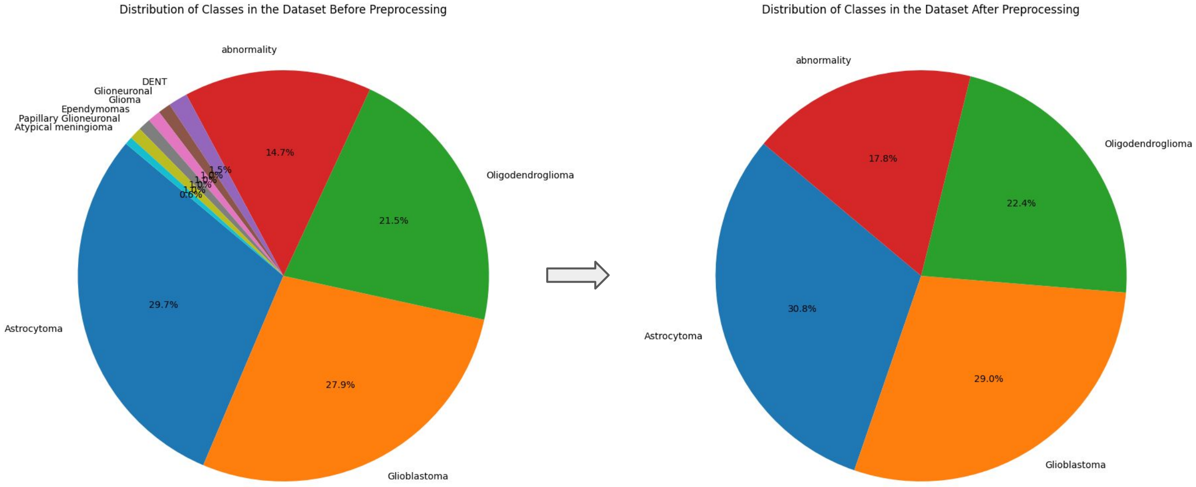
Dataset

- ReMIND dataset from NIH
  - 85,057 high-resolution 2D brain MRIs from 114 subjects featuring various types of brain tumors
  - Enables training of deep learning models for tumor recognition and classification
  - https://www.cancerimagingarchive.net/collection/remind



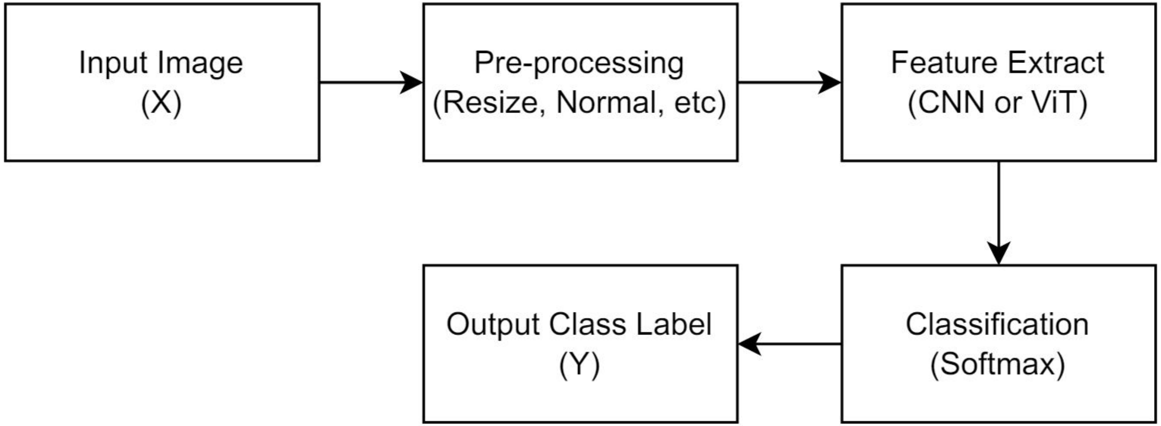
Dataset Preparation

- 78560 samples of 10 classes at beginning
- Remove rare classes
- Downsample by patient: 1200 -> 399
- Split randomly by patient into 70% train, 15% val, 15% test
- 42693 samples used in our experiments



Problem Setup - **Multi-Class Classification**

- Objective:
    - The goal is to classify brain tumors into multiple classes based on MRI images.



Models

Pre-trained CNNs

- VGG19
- InceptionV3
- DenseNet121
- Xception
- ResNet50

Pre-train ViTs

- ViT

Transfer Learning Implementation

Pre-trained CNNs

- Freeze the early layers and fine-tune only the last 10 layers
- Add GlobalAveragePooling2D layer and a Dense layer to transform output into four classes
- Four-dimensional one-hot encoding
- Adam optimizer
- Categorical cross-entropy loss function
- ReduceLROnPlateau learning rate scheduler

Pre-train ViTs

- Used only one-third of the dataset by downsampling the number of images per patient
- AdamW optimizer
- Categorical cross-entropy loss function
- ReduceLROnPlateau learning rate scheduler

Evaluation Metrics

- **Accuracy** calculates the proportion of correctly classified observations over the total number of observations

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

- **Recall** measures the sensitivity by calculating the number of observations correctly classified as true positives over the total number of true positives and false negatives

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **Precision** quantifies the proportion of true positives among all positive predictions, computed as the number of true positives divided by the sum of true positives and false positives.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **F2 Score** is similar to the F1 Score but emphasizes recall with parameter Beta

$$F_2 = \frac{(1+\beta^2) * \text{precision} * \text{recall}}{(\beta^2 * \text{precision} + \text{recall})}$$

Additional Evaluation Metrics

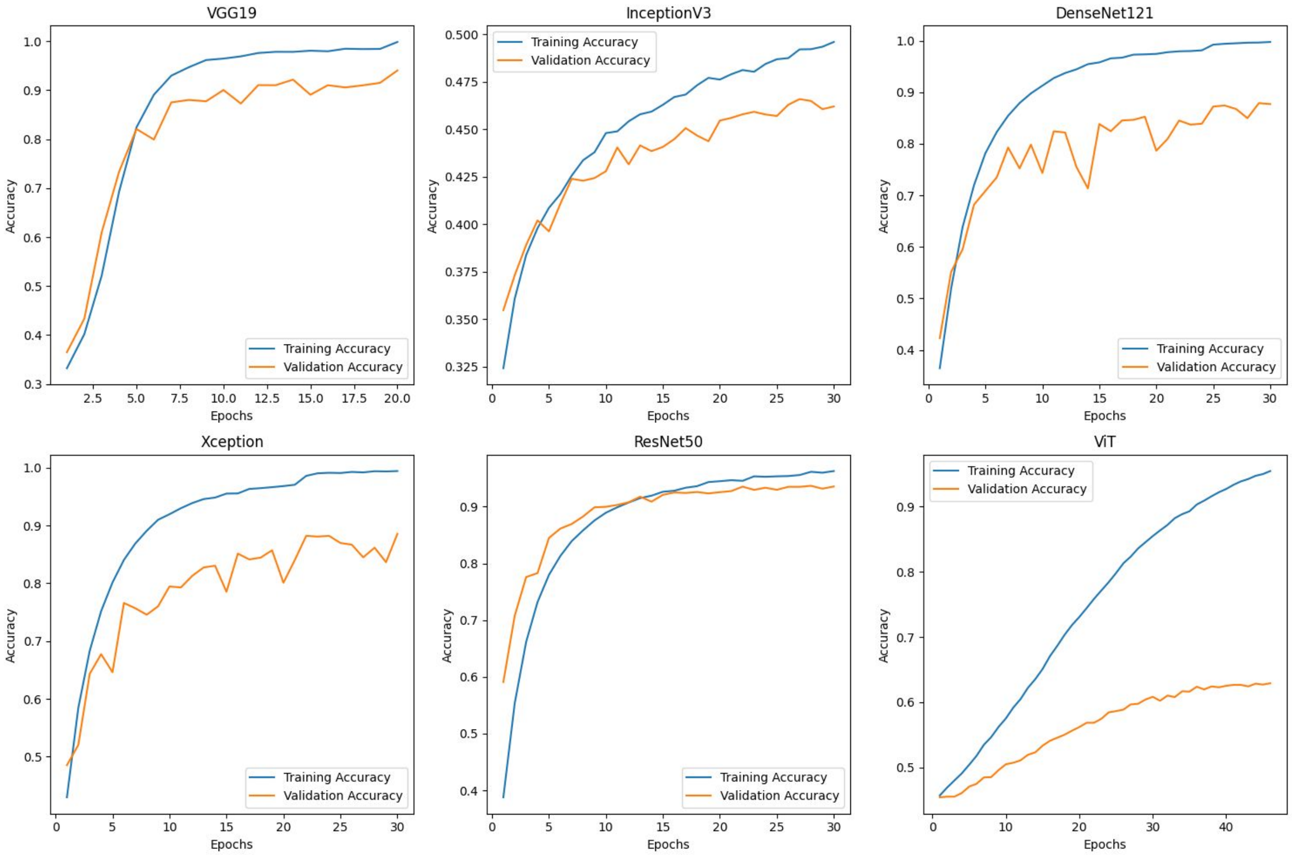
- **Training Time** is the time to train the model on training data.
- **Convergence (Training vs Validation Loss)** measures the stability of training and validation loss curves in a graph over epochs.
- **ROC-AUC curve** is the area under the receiver operating characteristic (ROC) curve and AUC score.

Results

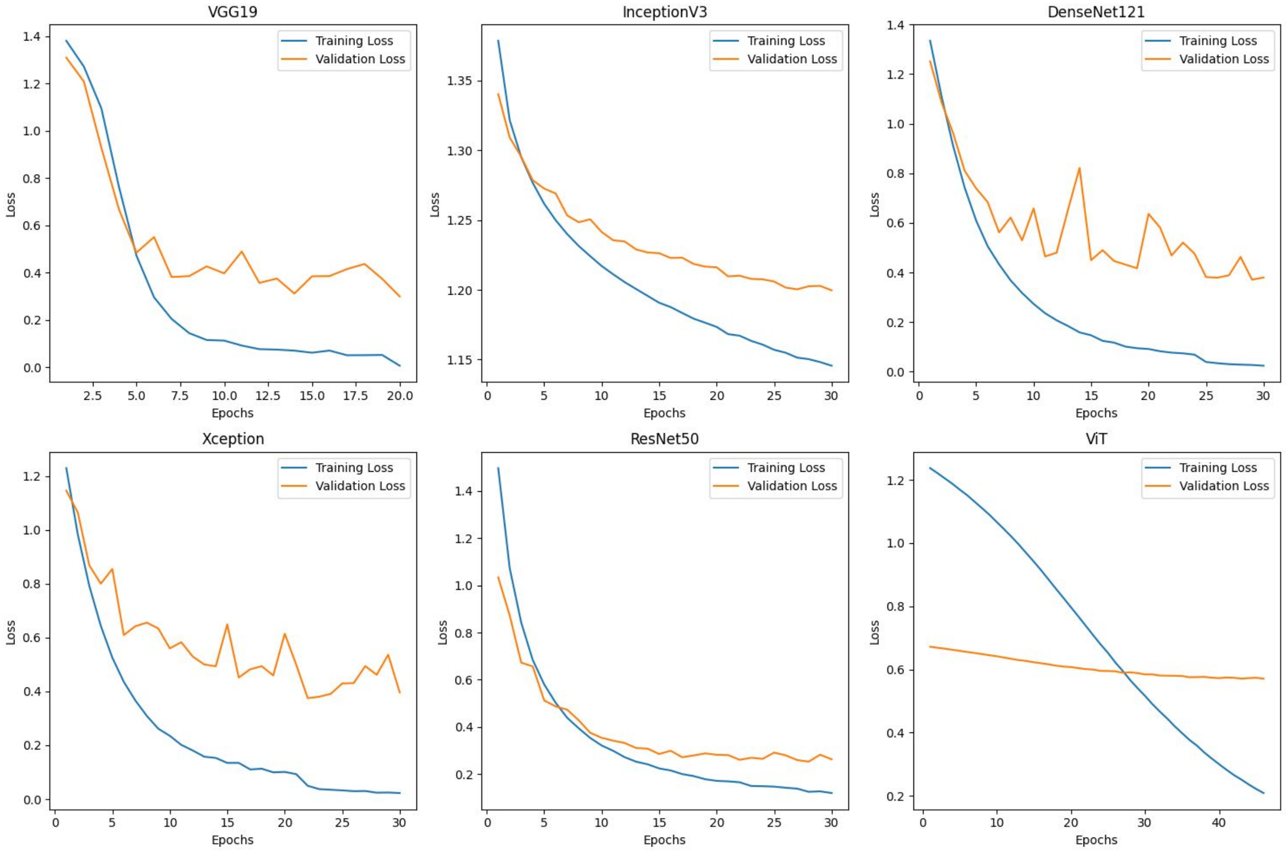
Model	Type	Accuracy			Loss	
		Training	Validation	Testing	Training	Validation
VGG19	Pre-trained CNN	0.984	**0.940**	0.931	0.052	0.299
InceptionV3	Pre-trained CNN	0.516	0.475	0.477	1.111	1.177
DenseNet121	Pre-trained CNN	**0.997**	0.879	0.871	0.027	0.371
Xception	Pre-trained CNN	0.993	0.886	0.884	**0.025**	0.397
ResNet50	Pre-trained CNN	0.960	0.937	**0.934**	0.127	**0.253**
ViT	Pre-trained ViT	0.962	0.630	0.584	0.166	0.574

Model	Type	Metrics			Training
		Precision	Recall	F2 Score	Training Time (sec)
VGG19	Pre-trained CNN	0.931	0.931	0.931	**1004**
InceptionV3	Pre-trained CNN	0.477	0.477	0.472	1740
DenseNet121	Pre-trained CNN	0.871	0.871	0.871	1209
Xception	Pre-trained CNN	0.885	0.884	0.884	1113
ResNet50	Pre-trained CNN	**0.934**	**0.934**	**0.934**	1936
ViT	Pre-trained ViT	0.260	0.229	0.215	3702

Additional Results - Training and Validation Accuracies

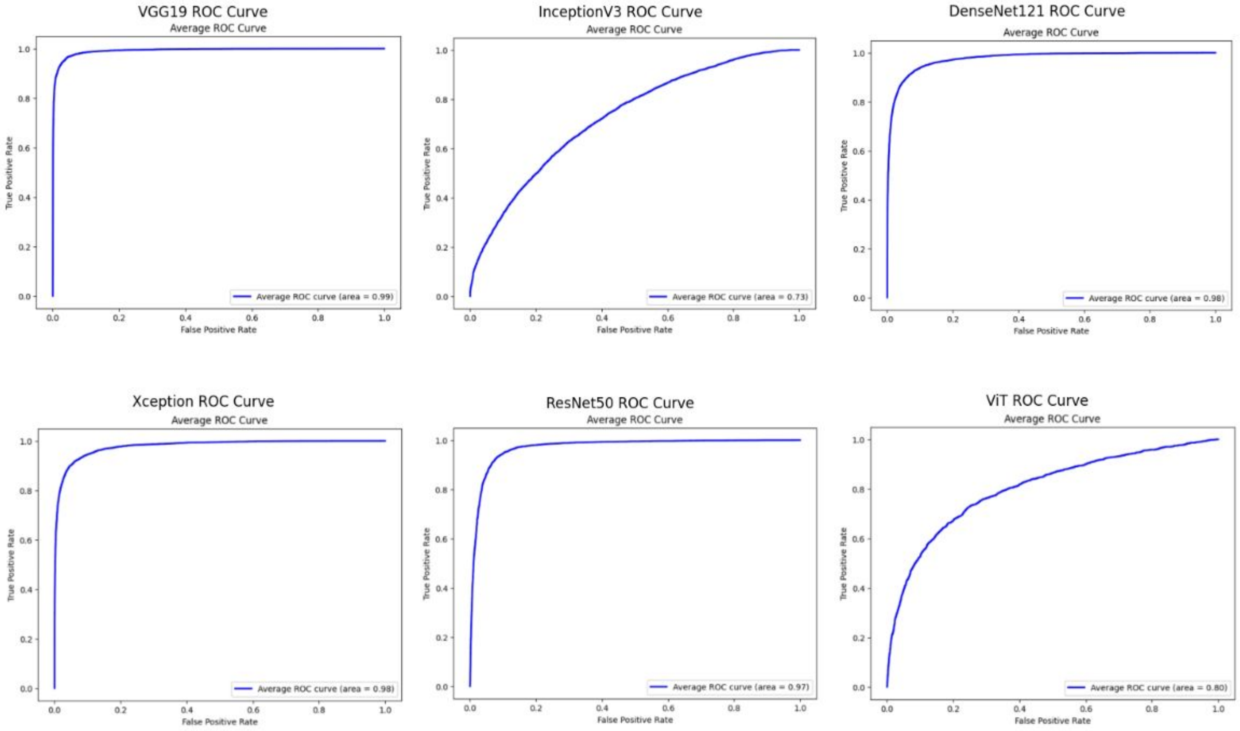


Additional Results - Training and Validation Losses



Additional Results - ROC Curves and AUC Scores

	VGG19	InceptionV3	DenseNet121	Xception	ResNet50	ViT
AUC	**0.99**	0.73	0.98	0.98	0.97	0.80



Discussion

- We use ~43k images to fine-tune 5 CNN models and ~14k for ViT
- VGG19 emerges as the best-recommended model for brain tumor classification via MRI scans
  - High accuracy (train acc 0.984/ test acc 0.931) & stable performance during training
  - Computational efficiency (1004s with A100 GPU)
- Although ResNet50 slightly exceeds VGG19 in performance (test acc 0.934), VGG19's computational efficiency make it the preferred choice (1004s v.s. 1936s)
- ViT, while converged and achieved 0.962 training accuracy, performs the worst in our experiments, suggesting that pre-trained Vision Transformer models may not be easy to handle for adapting to brain tumor multi-label classification task

Limitations & Future Work

- Due to the complex nature of ViT, it's highly sensitive to finetune and adapt to brain tumor classification task:
  - Require larger dataset (since ViT doesn't inherently encode inductive biases like CNNs do)
  - Carefully monitor overfitting through effective data augmentations and different regularizations
  - Optimize the fine-tuning process by grid-search of hyperparameters and layer configurations
- The high-performance CNN models are indeed capable of accurately classifying and hopefully determining the severity of brain tumors from MRI scans

Conclusion

- Pre-trained CNN models, specifically VGG19 and ResNet50, emerge as the optimal choice due to their superior performance (~0.93 test acc) and stability

- We anticipate that our findings will contribute significantly to brain tumor diagnostics, providing rapid, accurate insights that can assist healthcare professionals and inform treatment decisions, ultimately leading to better patient care