Date: November 8th, 2024

# Brain Tumor Classification using CNN and Transformers


*Presentation by:*

**Neerajdattu Dudam**

**Nagamani Motupalli**

**Madhu Sree Sane**

**Mounika Bollina**


*Supervisor:* ***Dr. Lina Chato***

*Class : Computer Vision CSC 752 –U18 (Fall 2024)*

# Inspiration and Objective

**Inspiration:**
Accurate brain pathology classification is crucial for diagnosing neurological diseases. Inspired by retinal disease classification, we aim to use deep learning for brain tumor classification.

**Objective:**
Develop a reliable and accurate system for brain tumor identification using advanced computer vision and machine learning techniques.

**Key Goals:**

- Classify brain tumors with CNNs using a supplied dataset.
- Optimize network topologies and layer configurations.
- Compare our CNN models with advanced models like ResNet50.
- Explore Vision and Swin transformers for performance improvement.
- Apply transfer learning to enhance accuracy.
- Analyze each strategy to determine the best approach.

# Presentation Outline

- Introduction
- CNN (Convolutional Neural Networks)
- ViT (Vision Transformer)
- Workflow
- Data
- Experiments
- Results and Conclusions
- Future Directions
- Team Contributions
- References

# Introduction

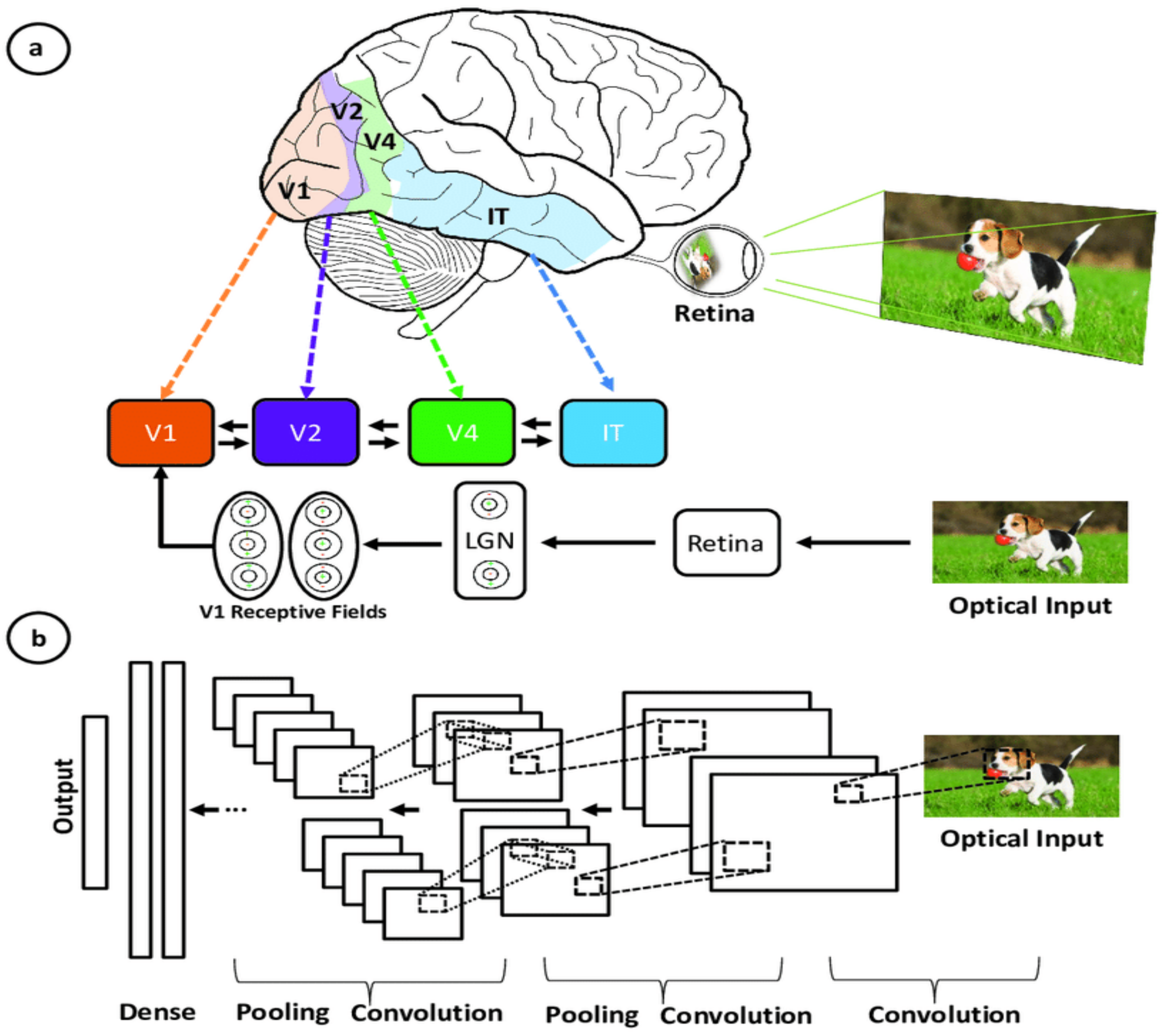
- **Brain Tumor Classification**: Brain tumors are complex and vary significantly in type, making accurate classification crucial for treatment planning and prognosis.

- **Role of AI in Medical Imaging**: Recent advances in AI, especially in deep learning, have revolutionized medical imaging by providing faster, more accurate diagnostic support.

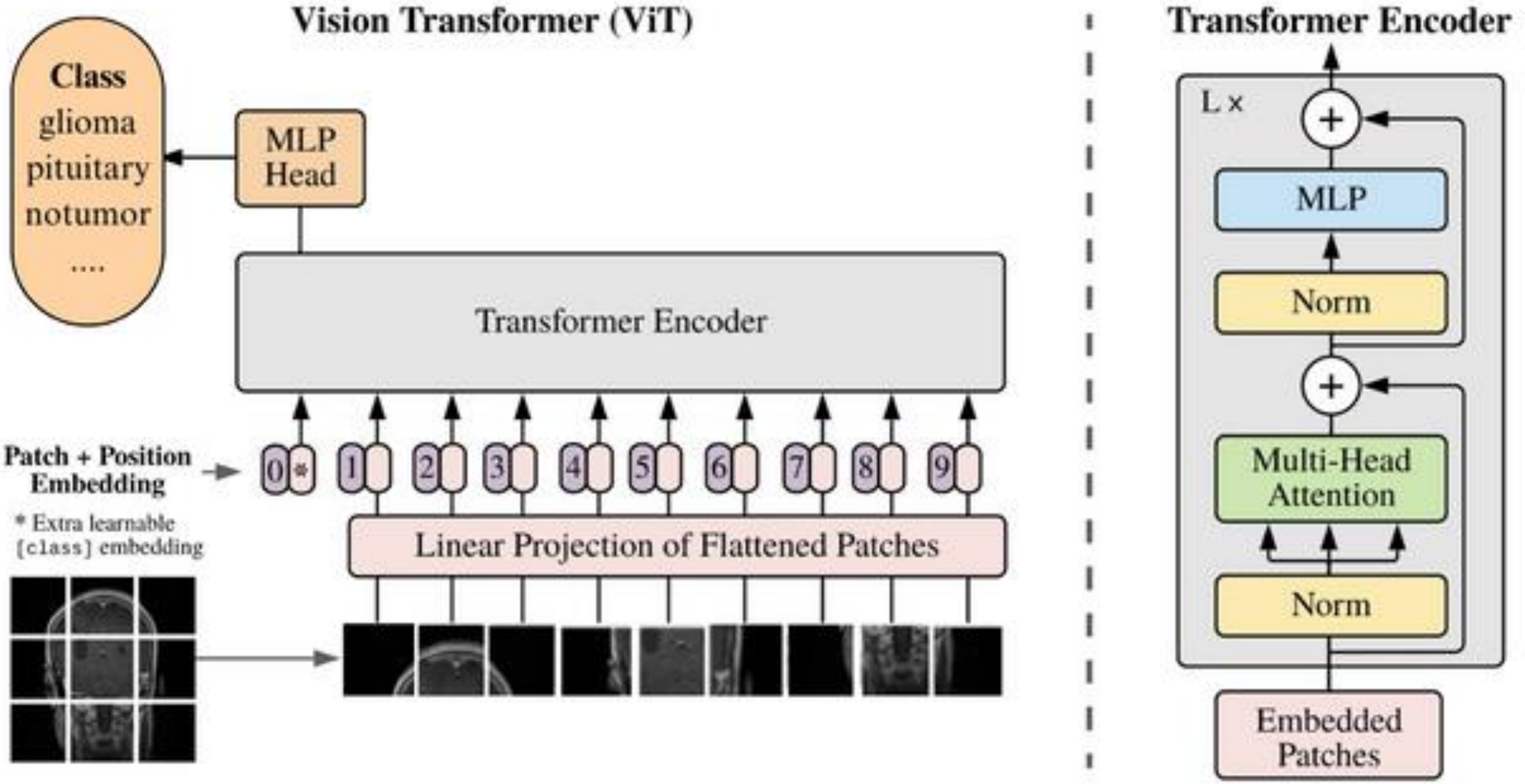
- **Deep Learning Techniques**: CNNs have become the standard for image classification due to their ability to capture intricate spatial hierarchies in visual data.

- **Emergence of Transformers**: Recently, transformers, originally used in NLP, have shown promise in computer vision, enabling models to identify complex patterns over large datasets.

# CNN (Convolution Neural Networks)



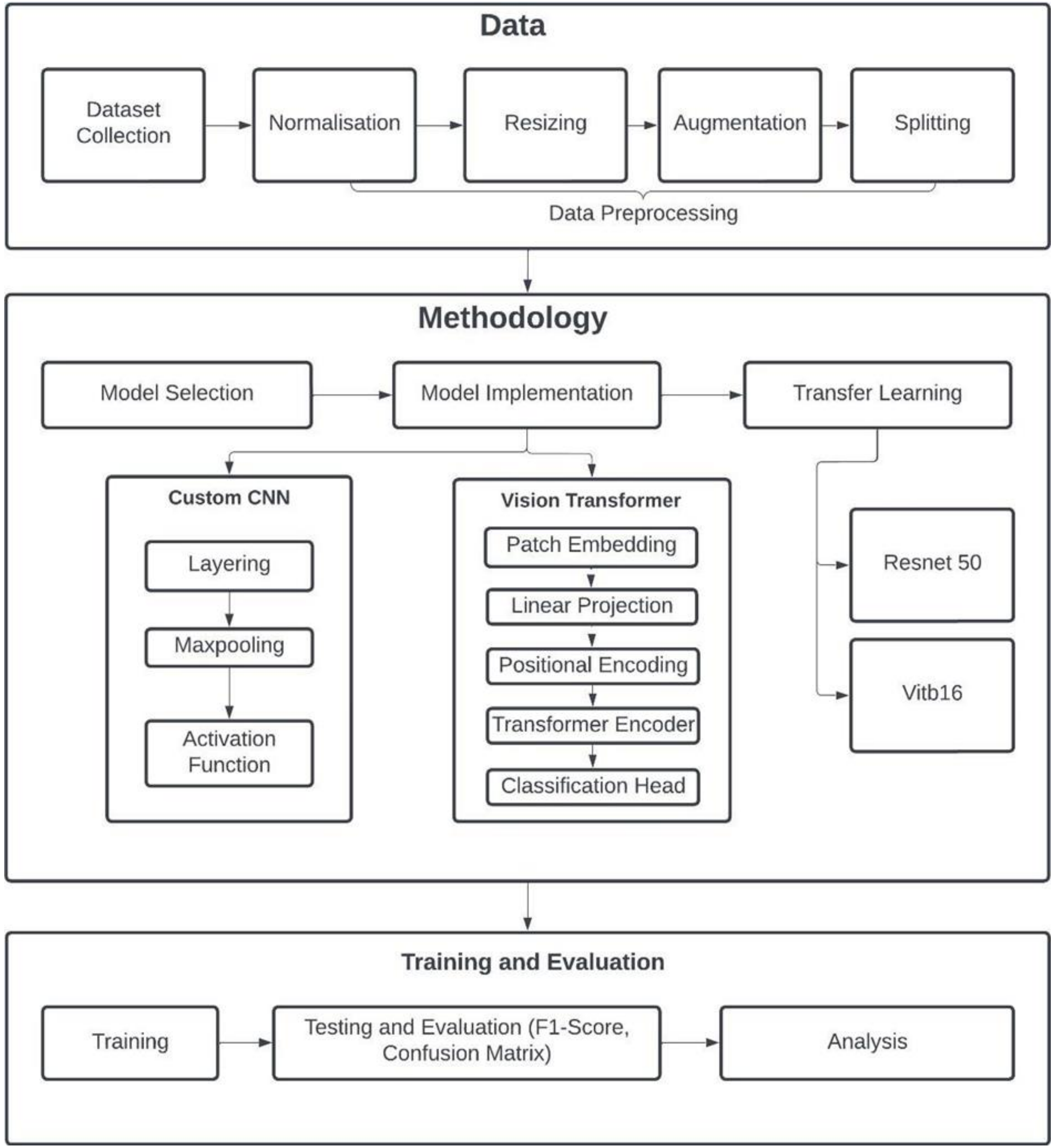
# Vision Transformer (ViT)

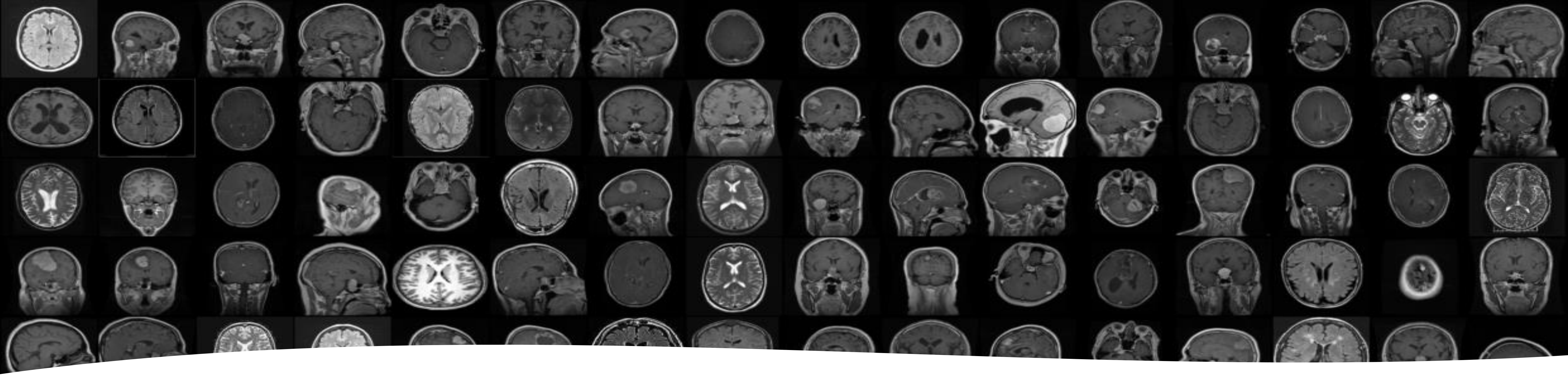


# Why ViT?

- Global Context and Long-Range Dependencies
- Flexibility with Image Size and Resolution
- Better Performance on Large Datasets
- Parallelization and Efficiency with Larger Models
- Transfer Learning Potential
- Handling Complex and Non-Local Patterns

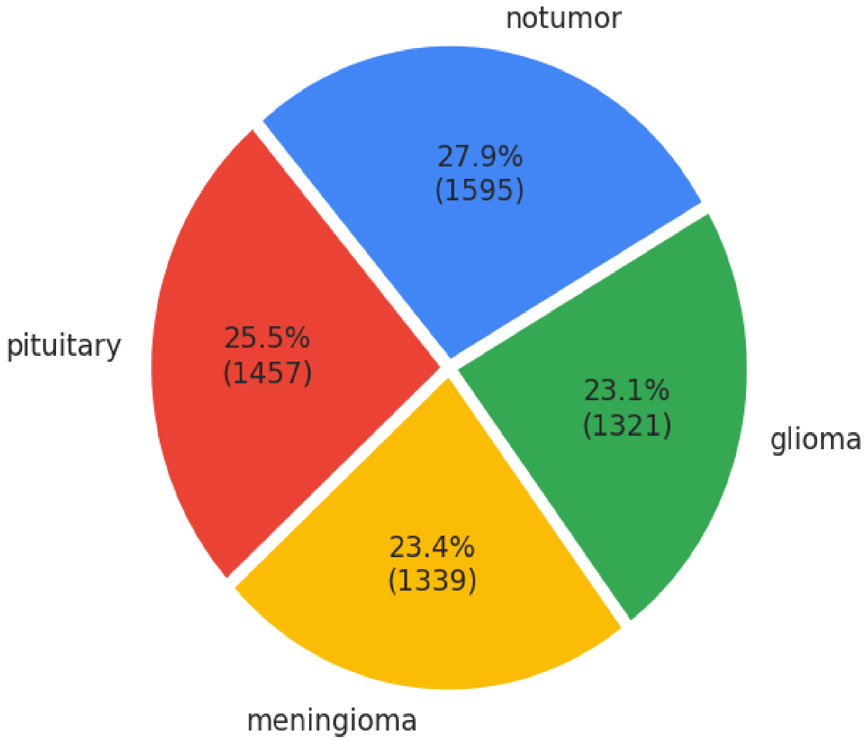
# Workflow


Data
Dataset Collection
Normalisation
Resizing
Augmentation
Splitting
Data Preprocessing
Methodology
Model Selection
Model Implementation
Transfer Learning
Custom CNN
Layering
Maxpooling
Activation Function
Vision Transformer
Patch Embedding
Linear Projection
Positional Encoding
Transformer Encoder
Classification Head
Resnet 50
Vitb16
Training and Evaluation
Training
Testing and Evaluation (F1-Score, Confusion Matrix)
Analysis



# Data

- **Dataset Overview**: The dataset contains MRI images used for classifying brain tumors.
- **Image Types**: Includes various types of brain tumors, such as Meningioma, Pituitary, and Glioma.
- **Image Quality**: High-resolution MRI scans that provide detailed structural views of the brain.
- **Labels**: Images are labeled with tumor types and grades, aiding in supervised learning tasks.
- A total of ~7000 images



# Experiments

| | Experiment 1 (CNN from scratch) | Experiment 2 (CNN with K fold) | Experiment 3 (Resnet 50) |
|---|---|---|---|
| Architecture | ➢ Multiple convolutional layers with increasing filter sizes (32, 64, 128, and 256)<br>➢ ReLU after each layer<br>➢ MaxPooling (2x2) after every 2 layers<br>➢ Dense (1024, 512 and 4) | ➢ Multiple convolutional layers with increasing filter sizes (32, 64, 128, and 256)<br>➢ ReLU after each layer<br>➢ MaxPooling (2x2) after every 2 layers<br>➢ Dense (1024, 512 and 4) | ➢ Pre-trained resnet50<br>➢ GlobalAveragePooling2D<br>➢ Dense layer with 256 nodes and ReLU activation<br>➢ Final Dense layer with softmax activation and 4 nodes |
| Batch Size | Training (128), Testing(256) | Training (128), Testing(256) | 32 for both |
| Epochs | 30 | 30 | 30 |
| Layer Freeze | Unfreeze all | Unfreeze all | Freeze All (except dense) |
| Folds | 0 | 5 | 0 |
| Learning Rate | 0.001 | 0.001 | 0.001 |
| Optimizer | Adam | Adam | Adam |
| Loss Function | Categorical cross-entropy | Categorical cross-entropy | Categorical cross-entropy |

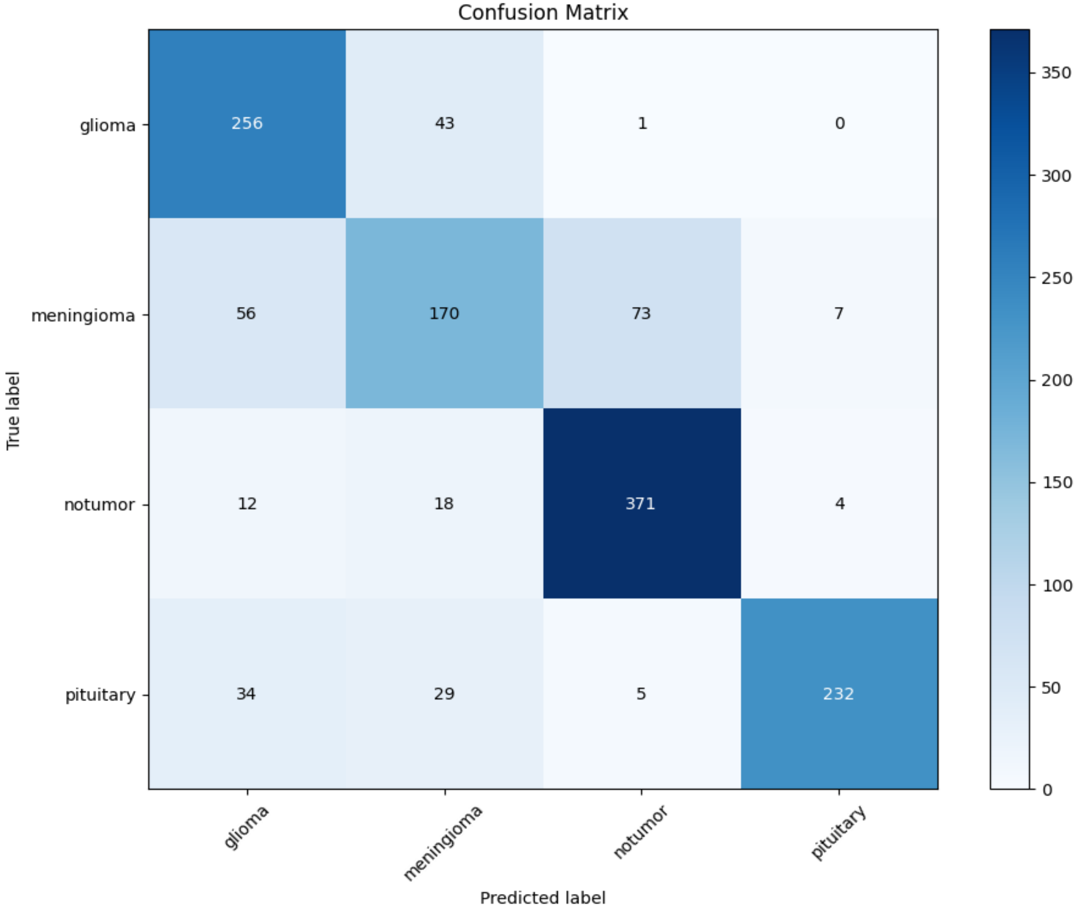
# Experiments

| | Experiment 4<br>(ViT from scratch) | Experiment 5<br>(ViT - B16) |
|---|---|---|
| Architecture | ○ **NewGELUActivation**: An activation function.<br>○ **PatchEmbeddings**: Converts the image into patches and projects them into a vector space.<br>○ **Embeddings**: Combines patch embeddings with class token and position embeddings.<br>○ **AttentionHead and MultiHeadAttention**: Implements the attention mechanism.<br>○ **FasterMultiHeadAttention**: Optimized multi-head attention.<br>○ **MLP**: A multi-layer perceptron module.<br>○ **Block**: A single transformer block.<br>○ **Encoder**: The transformer encoder module.<br>○ **ViTForClassfication**: The Vision Transformer model for classification. | ➢ Pretrained ViT - B16 |
| Batch Size | 32 | 32 |
| Epochs | 30 | 30 |
| Layer Freeze | Unfreeze all | Freeze all |
| Learning Rate | 0.001 | 0.001 |
| Optimizer | Adam | Adam |
| Loss Function | Categorical cross-entropy | Categorical cross-entropy |

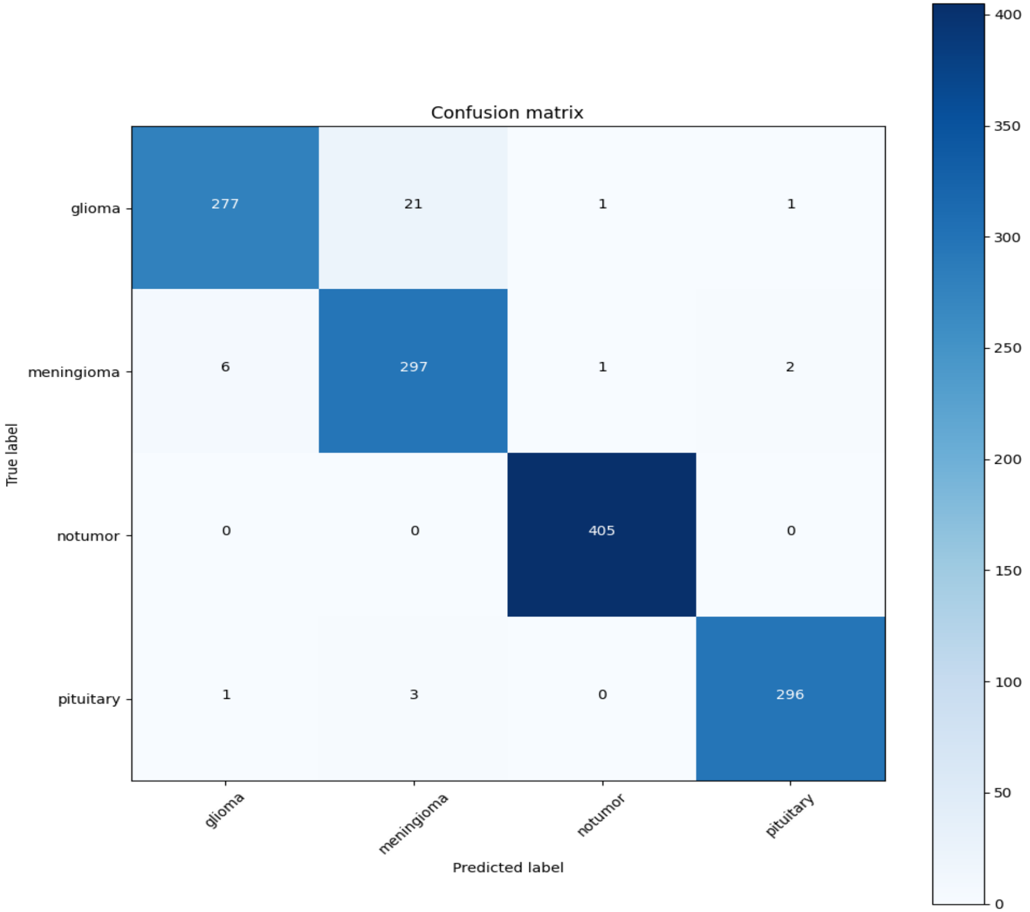
# Results and Conclusions

| | Experiment 1 (CNN from scratch) | Experiment 2 (CNN with K fold) | Experiment 3 (Resnet 50) | Experiment 4 (ViT from scratch) | Experiment 5 (ViT - B16) |
|---|---|---|---|---|---|
| Test Accuracy | 96.16 | Majority Voting: 97.48<br>Average Voting: 97.56 | 77.96 | 78.49 | 97.6 |

ViT from Scratch

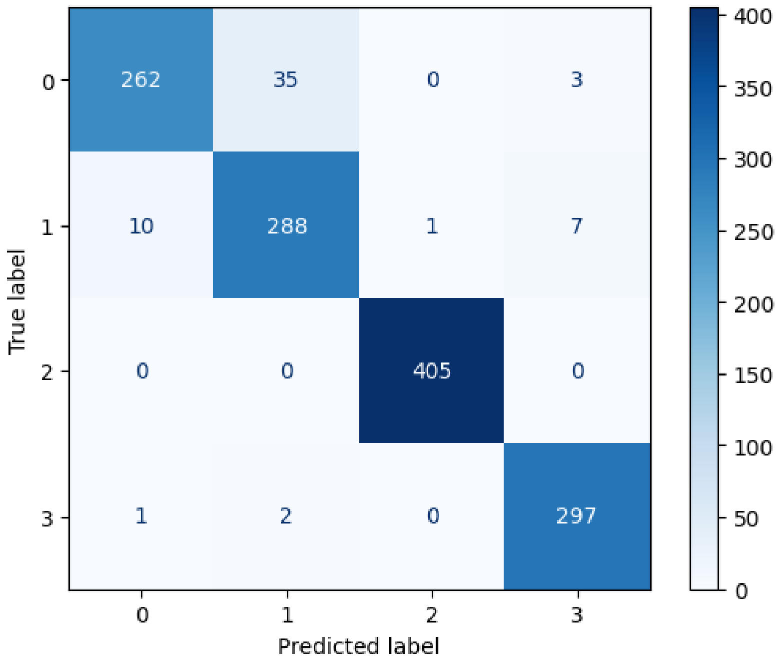



ViT B16

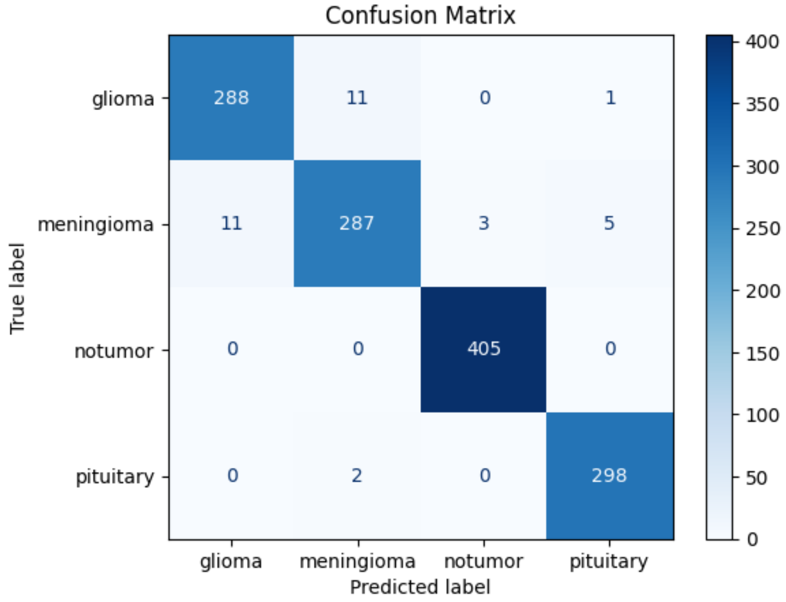


# Results and Conclusions

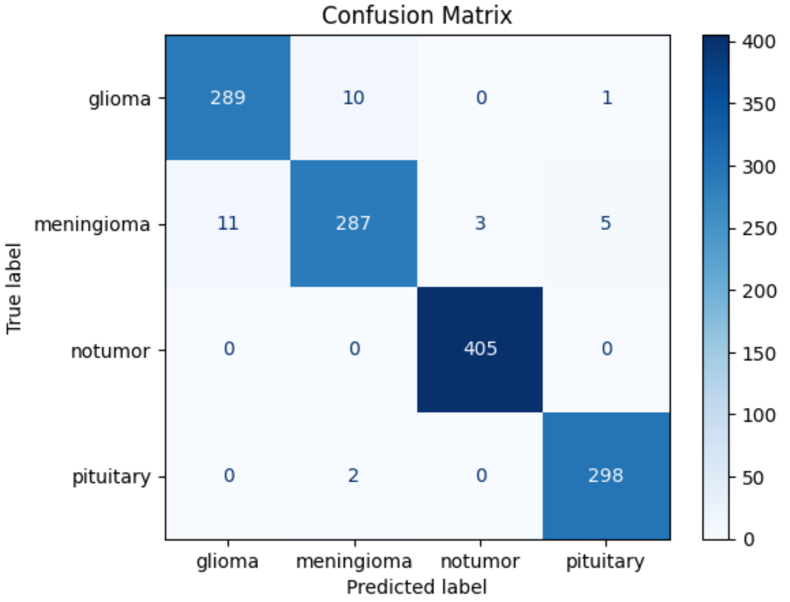
***Confusion matrices***




CNN from scratch



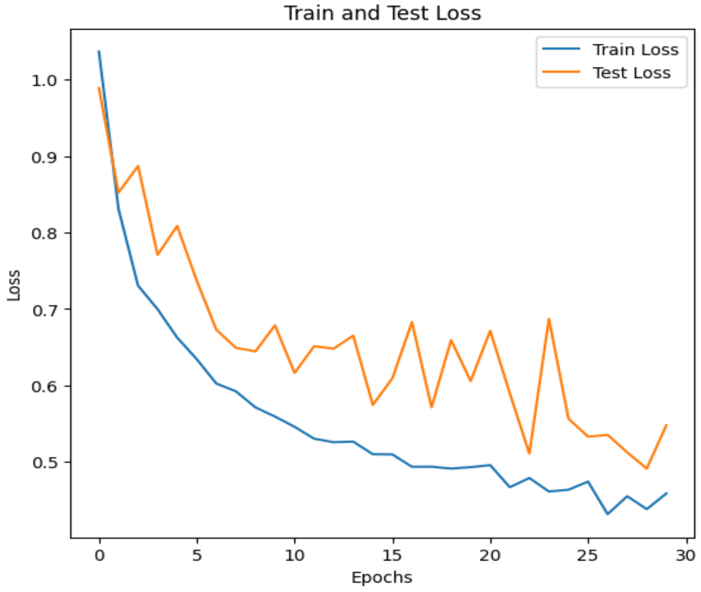

CNN with K fold (Majority Voting)

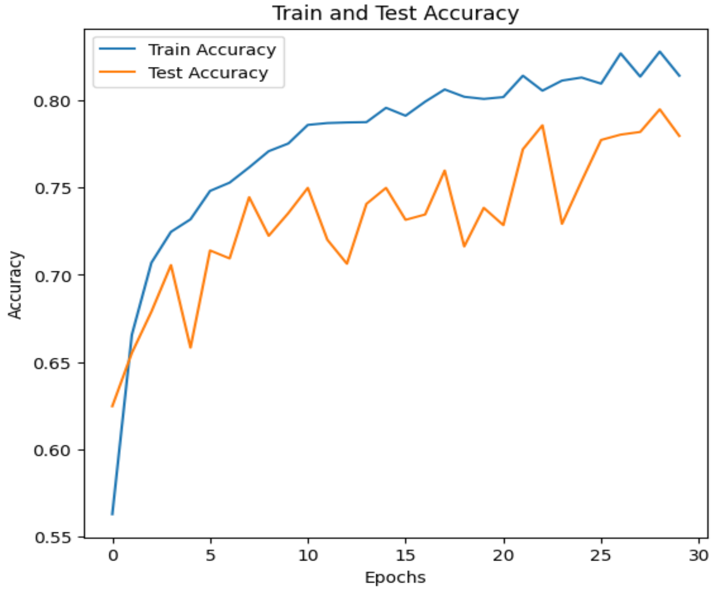



CNN with K fold (Average Voting)

- This indicates our CNN from scratch model can generalize the images very well, as **k-fold cross-validation helps assess its performance across multiple subsets of the data, reducing the risk of overfitting and providing a more reliable estimate of its accuracy and robustness on unseen data.**
- ViT from scratch needs more data to learn better.

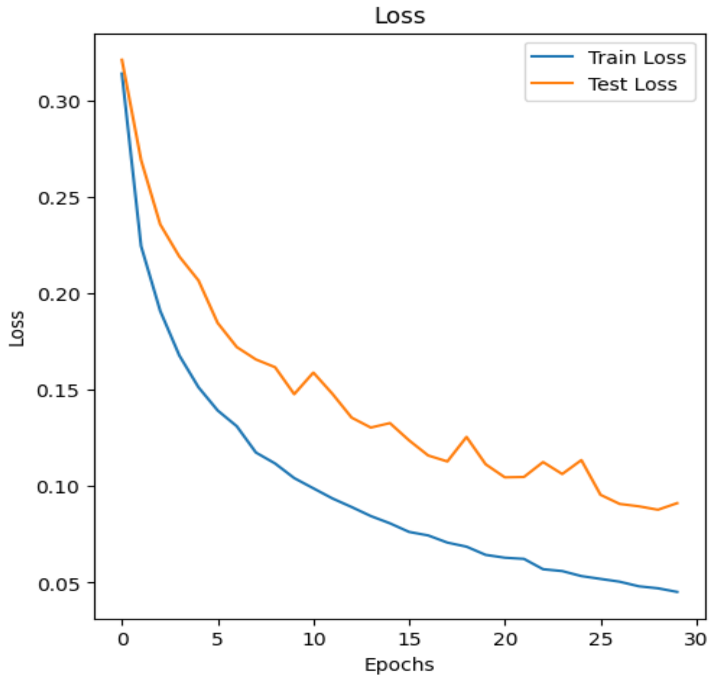
# Results and Conclusions

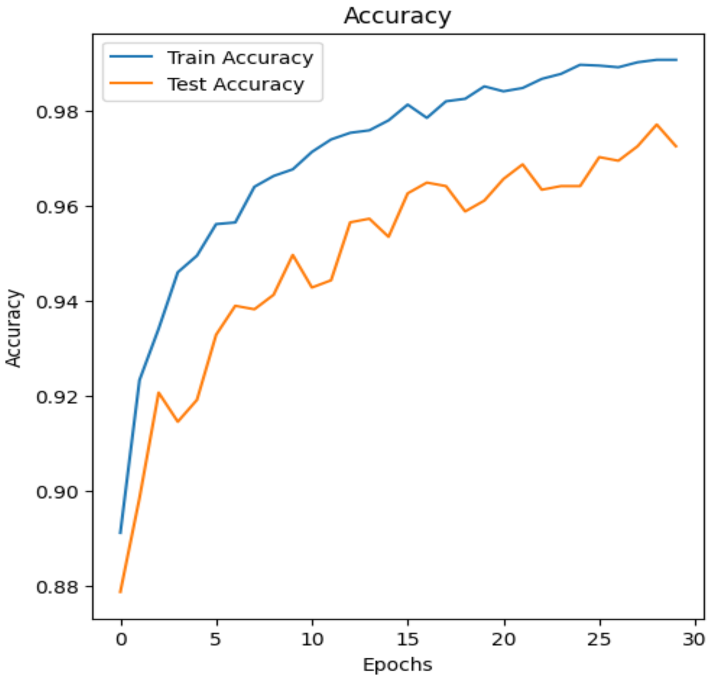



**Resnet 50**

**ViT B16**



# Future Directions

From the results and conclusions drawn from our study, we have identified several key future directions to improve model performance.

- **Improving ResNet-50 and ViT Accuracy:**
  - Decrease learning rate for ResNet-50 to improve convergence.
  - Apply more data augmentation techniques for ViT to enhance performance.
  - Reduce Overfitting

- **Using Pretrained Models:**
  - Fine-tune pretrained models like **VGG16**.
  - Compare their performance with our models to assess improvements.

- **Implementing Multi-label Classification:**
  - Extend the model to handle multi-label classification, where images can belong to multiple categories.

# Team Contributions

- **Neerajdattu Dudam** – Coding and Debugging, PowerPoint, References, Report
- **Nagamani Motupalli** – Literature Review, Report
- **Madhu Sree Sane** – Literature Review, Report
- **Mounika Bollina** – Literature Review, Report

# References


Corominas, Ò., Riera, I., Aditya, S., & Singh, S. (n.d.). *Image Classification with Classic and Deep Learning Techniques*. https://arxiv.org/pdf/2105.04895

Sharma, N., Jain, V., & Mishra, A. (2018). An Analysis Of Convolutional Neural Networks For Image Classification. *Procedia Computer Science*, *132*, 377–384. https://doi.org/10.1016/j.procs.2018.05.198

Wang, D., Lian, J., & Jiao, W. (2024). Multi-label classification of retinal disease via a novel vision transformer model. *Frontiers in Neuroscience*, *17*. https://doi.org/10.3389/fnins.2023.1290803

Keita, Z. (2023, November 14). *An Introduction to Convolutional Neural Networks (CNNs)*. Datacamp.com; DataCamp. https://www.datacamp.com/tutorial/introduction-to-convolutional-neural-networks-cnns

Soni, P. (2021, April 30). *Train CNN Model With PyTorch*. TheCyPhy. https://medium.com/thecyphy/train-cnn-model-with-pytorch-21dafb918f48

Nguyen, T. (2023, March 7). *Implementing Vision Transformer (ViT) from Scratch - Towards Data Science*. Medium; Towards Data Science. https://towardsdatascience.com/implementing-vision-transformer-vit-from-scratch-3e192c6155f0

Tigges, C. (2022, June 26). *Building the Vision Transformer From Scratch - Curt Tigges - Medium*. Medium. https://medium.com/@curttigges/building-the-vision-transformer-from-scratch-d77881edb5ff

Nickparvar, M. (2021). *Brain Tumor MRI Dataset*. Www.kaggle.com. https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

# Q & A