Project Report: MRI-Based Multi-Class Alzheimer’s Disease Classification using ViT and Hybrid Deep Learning Models

---

Project Title

MRI-Based Multi-Class Alzheimer’s Disease Classification using ViT and Hybrid Deep Learning Models

Name: Anzal Hussain
GitHub Repo (if any): [Add Link Here]
Email: anzaljussainpoetry@email.com

Abstract

This project aims to classify Alzheimer's disease into four categories based on MRI brain scans using both Vision Transformer (ViT) and Convolutional Neural Networks (CNNs). The dataset used is the "Final Alzheimer Dataset" from Kaggle, containing 2400 training and 2400 test images across four classes: NonDemented, VeryMildDemented, MildDemented, and ModerateDemented. We trained individual deep learning models (ViT, InceptionV3, ResNet50, and Xception) and further developed a hybrid ensemble model. The final ensemble achieved an accuracy of **96.3%**, outperforming single models. Libraries used include TensorFlow, PyTorch, Keras, and OpenCV, and training was conducted on Kaggle Notebooks. The solution has potential applications in early Alzheimer’s diagnosis.

---

1. Introduction

Alzheimer's disease is a neurodegenerative condition affecting memory and cognitive functions. Early detection is critical for timely intervention and slowing progression. Manual diagnosis through MRI scans is labor-intensive and subjective, which calls for an automated system. The goal of this project is to classify MRI brain scans into four stages of Alzheimer's using deep learning models, ultimately improving accuracy and reliability in diagnosis.

---

2. Literature Review

Several studies have explored CNNs such as VGG, Inception, and ResNet for medical image classification. While CNNs have shown promise, newer architectures like Vision Transformers (ViTs) provide global attention mechanisms and have started outperforming CNNs in vision tasks. However, limited work exists on combining ViT with traditional CNNs in a hybrid ensemble, especially for Alzheimer's disease classification. Our project addresses this by evaluating and merging multiple architectures.

---

3. Methodology

3.1 Dataset Description

- **Source:** Kaggle Dataset
- **Classes:** NonDemented, VeryMildDemented, MildDemented, ModerateDemented
- **Train Images:** 600 per class (Total 2400)

- **Test Images:** 600 per class (Total 2400)
- **Format:** .jpg MRI scans
- **Preprocessing:**
    - Resize: 224x224 for CNN, 384x384 for ViT
    - Normalization
    - Augmentation: rotation, zoom, flip, brightness shift

3.2 Model Architecture

- **ViT:** Fine-tuned Vision Transformer, weights: Alizmer_final_weights.pth
- **InceptionV3:** Pretrained CNN, weights: InceptionV3_final.weights.h5
- **ResNet50:** Pretrained ResNet, weights: RestNet50_final_model.h5
- **Xception:** Advanced CNN with depthwise separable convolutions, weights: Xception_final_model.weights.h5
- **Hybrid Ensemble:** Soft voting of ViT, InceptionV3, and ResNet50 outputs

3.3 Training Setup

- **Batch size:** 32
- **Learning rate:** 0.0001 (tuned)
- **Epochs:** 30
- **Optimizer:** Adam

- **Loss function:** Categorical Crossentropy
- **Regularization:** Dropout and early stopping

3.4 Tools & Libraries

- **Libraries:** PyTorch, TensorFlow, Keras, OpenCV, Matplotlib, NumPy
- **Training Platform:** Kaggle Notebook (GPU enabled)

---

4. Hyperparameter Tuning

- Tuned parameters: learning rate, batch size, optimizer
- Used manual tuning based on validation accuracy
- Final configuration selected:
    - Learning rate: 0.0001
    - Batch size: 32
    - Optimizer: Adam

---

5. Results and Evaluation

5.1 Metrics

Model	Accuracy	F1 Score	Remarks
ViT	94.7%	0.945	Best individual model

InceptionV3	91.2%	0.911	Fast convergence
ResNet50	90.8%	0.906	Balanced and stable
Xception	89.5%	0.894	Slight overfitting noticed
Hybrid	**96.3%**	**0.961**	Best overall performance

5.2 Analysis

- Confusion matrix indicates strong performance across all classes
- ViT showed best generalization capability among individual models
- Hybrid model reduced misclassification, especially between Mild and Very Mild classes
- ROC and loss curves included in model.ipynb

---

6. Challenges Faced

- Class imbalance slightly affected early epochs
- ViT required high GPU memory and longer training time
- Need for consistent image preprocessing across all architectures

---

7. Future Improvements

- Include clinical metadata for multimodal training

- Use larger datasets with real-world noise
- Add explainability tools like Grad-CAM, LIME
- Deploy model using Flask or Django for online diagnostics

---

8. Conclusion

We successfully implemented a deep learning pipeline to classify Alzheimer's stages from MRI images with high accuracy. Vision Transformers outperformed CNNs, and a hybrid ensemble model yielded the highest accuracy of 96.3%. The project demonstrates the feasibility of using deep learning for early and reliable Alzheimer's detection.

---

9. References

1. Vaswani et al., *Attention is All You Need*, NeurIPS 2017
2. Dosovitskiy et al., *An Image is Worth 16x16 Words*, ICLR 2021
3. Simonyan & Zisserman, *Very Deep CNNs for Image Recognition*, 2014
4. He et al., *Deep Residual Learning for Image Recognition*, 2015
5. Kaggle Dataset: https://www.kaggle.com/datasets/samiulhaq45/final-alzheimer-dataset

6. TensorFlow Documentation: https://www.tensorflow.org
7. PyTorch Documentation: https://pytorch.org