Alzheimer MRI Disease Classification

1st Sravani Reddygari
Computer Science
The University Of Memphis
Memphis, TN USA
srddygri@memphis.edu

***Abstract*—Alzheimer's Disease (AD) is a significant and growing global health issue, affecting millions of people around the world. Despite ongoing research, there is currently no cure or highly effective treatment for AD, making early detection essential to slowing its progression and alleviating its impact on patients and healthcare systems. This project explores the use of the "Augmented Alzheimer MRI Dataset V2," a specialized dataset developed to improve the precision and reliability of machine learning models for diagnosing and classifying different stages of AD from MRI scans. The dataset includes MRI images categorized into four classes representing different stages of cognitive impairment: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. Traditional datasets often face challenges like limited sample sizes and low variation, which can restrict model effectiveness. To address these issues, this dataset includes both original and augmented images. By applying data augmentation techniques such as rotation, scaling, and flipping the dataset generates diverse image variations while retaining clinically relevant features. This augmented data enhances model training, helping models generalize better and avoid overfitting, which in turn increases resilience to variations in real-world data. The dataset is organized with clear folder structures for both original and augmented images, simplifying the separation of training, validation, and test sets. This structure aids in systematic model development and enables models to learn from a broad range of data types, thereby enhancing their ability to accurately classify stages of AD and advancing the effectiveness of AI in AD diagnosis. In developing this dataset, this project highlights the crucial role of data augmentation in medical imaging to bolster model performance. The "Augmented Alzheimer MRI Dataset V2" marks progress toward creating reliable, AI-based tools for early and accurate detection of Alzheimer's, aiming to reduce the disease's burden on patients, their families, and healthcare providers worldwide.**

***Index Terms*—low variation,augmented images,overfitting,systematic model**

I. Introduction

Alzheimer's Disease (AD) is a progressive neurological disorder and the most common cause of dementia, affecting millions worldwide. It accounts for around 70% of dementia cases and has become a major global health challenge. AD gradually impairs cognitive functions, such as memory, reasoning, and language, making it increasingly difficult for individuals to carry out daily activities. The condition is irreversible, and despite decades of research, there is no cure or highly effective treatment that can halt or significantly slow its progression. This situation contrasts with other major diseases, like certain types of cancer, where advancements in early detection and treatment have led to improved patient outcomes. In the absence of a cure, early diagnosis of AD becomes critically important, as it provides opportunities to implement interventions that may slow cognitive decline, improve quality of life, and allow patients and their families more time to plan for the future. AD is thought to begin at least two decades before the onset of noticeable symptoms, with gradual and subtle changes occurring in the brain. Initially, these changes are undetectable to the individual, as they often affect brain cells in ways that do not immediately impact daily life. However, as the disease progresses, neurons in certain brain regions become damaged or die, leading to symptoms that include memory loss, language difficulties, and confusion. Over time, these impairments worsen, eventually impacting a person's ability to function independently. Individuals with AD often experience a long course of symptoms that increase in severity, which places a growing burden not only on patients but also on caregivers and healthcare systems. Given that AD predominantly affects people over the age of 65, the aging population in many countries means the prevalence of AD is expected to rise significantly in the coming years. As more people live longer, the need for reliable, accessible methods of early AD detection becomes ever more urgent. Early diagnosis not only benefits patients by providing time for therapeutic interventions but also enables healthcare providers to manage the disease's progression more effectively. Recent advances in artificial intelligence (AI) and medical imaging have opened new pathways for early AD diagnosis. Machine learning models, particularly deep learning (DL) techniques, have shown promise in analyzing complex medical data, including MRI scans, to detect early signs of neurological diseases. MRI, a non-invasive imaging technique, provides detailed insights into brain structure and function, making it an invaluable tool in AD research. By training DL models on large datasets of brain MRI images, researchers can develop algorithms that recognize patterns associated with different stages of AD, even in its earliest phases. This study utilizes the "Augmented Alzheimer MRI Dataset V2," a curated dataset that addresses common challenges in AD research, such as limited sample sizes and variability in image data. This dataset is categorized into four classes representing different stages of cognitive impairment: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. To enhance the dataset, a

series of augmentation techniques such as rotation, scaling, and flipping have been applied to create diverse variations of the images while preserving their clinical relevance. By expanding the dataset in this way, the models are exposed to a wider array of data, which reduces overfitting and improves their ability to generalize to new, unseen images. This enriched data set-up ultimately aims to make deep learning models more robust and reliable in detecting and classifying AD at different stages. The dataset's structured organization, with separate folders for original and augmented images, allows for efficient model training, validation, and testing. This clear separation supports systematic evaluation of model performance and enables researchers to gauge the effectiveness of using augmented data in clinical model development. By leveraging this enhanced dataset, this study seeks to demonstrate the value of data augmentation in improving machine learning applications in medical imaging and contributes to the growing field of AI-based tools for early and accurate AD diagnosis. These efforts are essential to reducing the future impact of Alzheimer's Disease on individuals, families, and healthcare systems worldwide.

II. BACKGROUND WORK

In recent years, deep learning (DL) and other advanced machine learning techniques have demonstrated impressive capabilities in distinguishing Alzheimer's Disease (AD) stages and identifying early signs in brain imaging. A variety of studies have employed different DL architectures to enhance classification accuracy and offer a more refined understanding of the disease's progression. In 2016, Ortiz et al. applied a deep learning model to differentiate between AD, Mild Cognitive Impairment (MCI), and non-converting (NC) individuals, using the Automated Anatomical Labeling (AAL) software to segment the brain into three-dimensional patches, which were then fed into deep neural networks (DNNs) for training. To improve prediction accuracy, they employed four voting algorithms during the classification phase. Their innovative methodology achieved a 90

Following that, Sarraf and Tofighi (2016) utilized convolutional neural networks (CNNs) to detect AD from MRI data, leveraging the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, which provided 43 MRI images for validation. Their model achieved a mean classification accuracy of 96.85

In 2018, Islam and Zhang further advanced the field by developing a CNN model for AD detection using MRI data. Their model, which was compared against pre-trained deep learning models like ADNet, InceptionV4, and ResNet, demonstrated superior performance. Using 416 images from the OASIS dataset, the model achieved a classification accuracy of 93

Building on these foundations, Jo et al. (2019) conducted a comprehensive study comparing traditional machine learning (ML) methods with DL approaches for early AD detection and predicting the conversion from MCI to AD. Reviewing 16 studies, they found that a hybrid approach combining ML and DL achieved the highest efficiency in feature selection (96

Also in 2019, Lee et al. explored the use of multimodal Recurrent Neural Networks (RNNs) for tracking AD progression from MCI. By integrating various data sources—such as cross-sectional neuroimaging, cerebrospinal fluid (CSF) analysis, and cognitive performance metrics—their model achieved an accuracy of 81

Ahmed et al. (2019) developed a streamlined CNN framework targeting the left and right hippocampal regions using MRI data. Using both the Gwangju Alzheimer's and Related Dementias (GARD) dataset and the ADNI dataset for validation, they analyzed a total of 678 MRI images and demonstrated the model's effectiveness in accurately detecting AD, with the hippocampal region analysis reinforcing its diagnostic value.

In 2020, Yang and Liu, along with Rolls et al., created a DL algorithm focused on early-stage AD detection using fluorine-labeled fluorodeoxyglucose (FDG) PET scans. This algorithm, developed in the CAFFE deep learning framework, extracted and classified features from PET images with a high degree of accuracy. It also incorporated a hierarchical 2D CNN for intra-slice feature extraction and a Gated Recurrent Unit (GRU) RNN for inter-slice processing, achieving high AUC scores when distinguishing AD from normal cognition (NC) and MCI from NC. Furthermore, they applied a contrastive learning approach that amplified key sections of 3D PET images, enhancing the distinction between AD and non-AD classes through contrastive loss. This innovative combination of CNN and RNN models in PET scan analysis showed promising results for early-stage AD detection and classification.

Pan et al. (2020) expanded upon the use of DL with FDG-PET imaging by introducing the MiSePyNet network, which integrated multiple views of PET scans for robust feature extraction. Their approach employed separable convolution to maintain spatial integrity while reducing the model's computational requirements, contrasting traditional, more resource-intensive 3D convolution methods. Additionally, Pan et al. applied CNN combined with ensemble learning (EL) techniques to analyze MRI data for AD detection. Using 278 images from the ADNI dataset, their combined CNN-EL approach outperformed other models, including 3D-SENet and PCA-SVM, achieving an accuracy rate of 85

These studies collectively underscore the potential of deep learning and advanced data augmentation techniques in enhancing the accuracy and reliability of AD diagnosis from neuroimaging data. The integration of different neural network architectures and multimodal data sources has progressively

Fig. 1. Enter Caption

improved the field, highlighting DL's adaptability and efficiency in early AD detection and in monitoring its progression.

III. METHODOLOGY

A. Dataset Preparation and Organization

1. Dataset Splitting:The Alzheimer's dataset is divided into Train Data and Validation Data sets. This split helps ensure that the model is trained on one portion of the data while another separate portion is used to validate its performance. Within the training data, images are classified into four categories: MildDemented, ModerateDemented, NonDemented, and VeryMildDemented, which correspond to different stages of Alzheimer's disease.

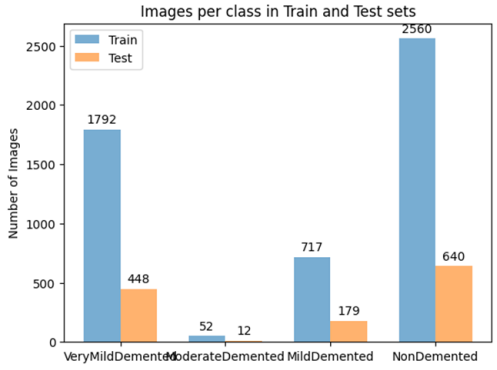

Fig. 2. Images per class in training and testing set

B. Data Preprocessing and Augmentation

1) Preprocessing: Images are resized to a uniform shape (e.g., 208x176 pixels) and normalized by scaling pixel values to the range [0, 1].
2) Data Augmentation: Augmentation techniques such as rotation, shifting, shearing, zooming, and horizontal flipping are applied to the training dataset to increase data variability and reduce overfitting.
3) Class Balancing: To address class imbalance, each class is assigned a weight during model training, calculated based on the class frequencies.

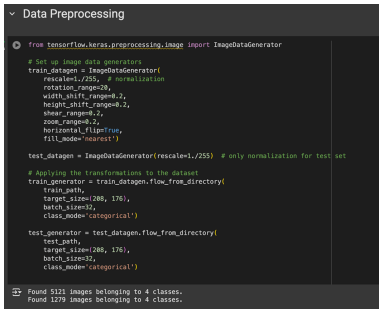

Fig. 3. Data Preprocessing

C. Class Balancing

1) Purpose: In cases where some classes have significantly more samples than others, class imbalance can lead to biased learning, where the model becomes more accurate on the dominant classes and underperforms on underrepresented ones.
2) Method: Class weights are calculated based on the frequency of each class. These weights are then applied during model training to ensure the model gives appropriate attention to all classes, particularly the minority classes.

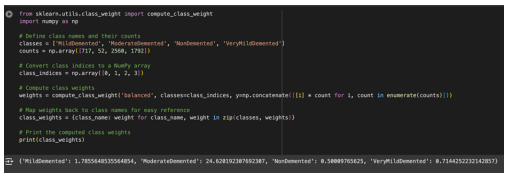

Fig. 4. Data Preprocessing

D. VGG Model

1) Load the Pre-trained VGG16 Model: Initialize the VGG16 model with weights pre-trained on ImageNet, excluding the top layers so we can add custom layers for our specific classification task.
2) Freeze the Base Layers: Prevent the pre-trained VGG16 layers from being modified during training, which allows us to leverage their learned features while focusing the training on the added layers.
3) Add Custom Layers: Attach new layers to the VGG16 base, including a flattening layer, dense layer(s) for feature learning, dropout layer(s) to prevent overfitting, and a softmax output layer for multi-class classification (four classes representing different Alzheimer's stages).
4) Compile the Model: Set the optimizer, loss function, and evaluation metrics. Typically, Adam is used with a learning rate suitable for fine-tuning, categorical crossentropy for multi-class classification, and accuracy as the primary metric.
5) Train the Model: Fit the model using the training data, with validation data to monitor performance. Optionally, apply class weights to address class imbalance.

6) Evaluate the Model: After training, assess the model's performance on the test or validation set, looking at metrics like accuracy, precision, recall, and F1-score to understand its effectiveness.
7) Save the Model : Save the trained model for future use, enabling you to reload and use it without retraining.

Fig. 5.

E. Results

we have tested the model performance was evaluated based on accuracy for three different architectures: VGG16 Transfer Learning Model, Five-Layer Custom DNN Model, and Simple Custom CNN Model. The bar plot shows a comparison of the accuracy achieved by each model on the test dataset.

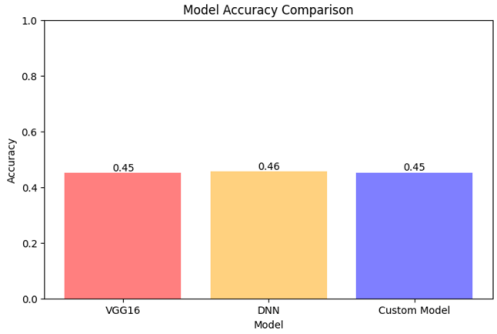

Fig. 6.

1) VGG16 Transfer Learning Model Accuracy: The VGG16 model achieved an accuracy of 0.45 (45Observations: Despite using transfer learning with a pre-trained model, the VGG16 architecture did not outperform the other models significantly. This could be due to differences in the domain of the pre-trained features (ImageNet data vs. Alzheimer's MRI images), which may not align closely enough for optimal feature extraction.
2) Five-Layer Custom DNN Model Accuracy: The custom DNN model with five layers achieved an accuracy of 0.46 (46Observations: This model marginally outperformed the VGG16 model, suggesting that a smaller, custom architecture may be better suited to the dataset. The model's performance indicates that further tuning and optimization might yield better results, given that the architecture is relatively shallow and may not be capturing sufficient depth for complex feature extraction.
3) Simple Custom CNN Model Accuracy: The simple custom CNN model achieved an accuracy of 0.45 (45Observations: This model was designed to be computationally lightweight and easy to train. While its performance is similar to that of the VGG16, it highlights the limitations of shallow architectures for complex classification tasks, such as differentiating between Alzheimer's stages.

F. Summary of Results

1) Data Complexity and Model Depth: The Alzheimer's MRI dataset may contain subtle patterns that require a deeper model with more complex feature extraction capabilities. The current architectures may not be sufficiently deep to capture these nuances, especially when distinguishing between similar stages.
2) Class Imbalance: Class imbalance might have impacted performance, leading the model to underperform on less-represented classes. Although class weights were used, further adjustments or alternative techniques (such as SMOTE for balancing the training set) could be considered.
3) Transfer Learning Limitations: The VGG16 model was pre-trained on ImageNet, a dataset of natural images, which might not share enough relevant features with MRI scans for Alzheimer's. Fine-tuning more layers of VGG16 or using a model pre-trained specifically on medical imaging (e.g., a model pre-trained on brain MRI datasets) might improve performance.
4) Hyperparameter Tuning: Further hyperparameter tuning, including learning rate adjustments, number of layers, and dropout rates, could be explored to optimize each model's performance.

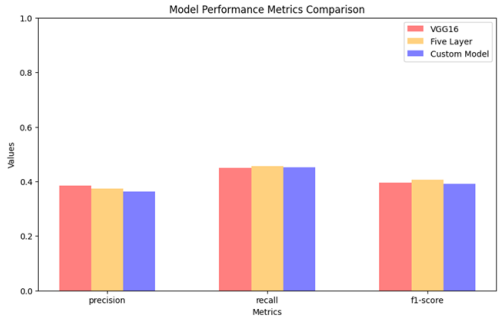

Fig. 7.

In our comparison of the three models—VGG16 Transfer Learning Model, Five-Layer Custom DNN Model, and Custom CNN Model—we evaluated performance using three key metrics: Precision, Recall, and F1-Score, each offering insights into different aspects of model accuracy for classifying Alzheimer's Disease stages. Precision, which indicates how accurately the models identify positive cases, was similar across all models, with scores around 0.45, suggesting that each model has a comparable tendency to correctly predict positive classes, though the low values highlight potential

for improvement. Recall, measuring the models' ability to capture all relevant instances, also yielded close results at approximately 0.46 for each model, indicating a moderate but insufficient capability in recognizing true positives. F1-Score, the harmonic mean of precision and recall and especially relevant in imbalanced datasets, further reflected this trend with values around 0.45 across the board. The consistency in these results implies that all three models face similar challenges in distinguishing Alzheimer's stages from MRI data. This may stem from limitations in model depth or complexity, which might be insufficient to capture subtle patterns in the images. Additionally, the intrinsic complexity of the MRI data and the possible impact of class imbalance, despite using class weights, may contribute to these modest scores, indicating the need for further optimization or exploration of more sophisticated models tailored for medical imaging.

G. Conclusion

In this study, we explored three deep learning models—VGG16 Transfer Learning Model, Five-Layer Custom DNN Model, and Custom CNN Model—to classify stages of Alzheimer's Disease from MRI images. Our results indicated that all three models demonstrated similar performance, with precision, recall, and F1-scores around 0.45, reflecting moderate but limited accuracy across the Alzheimer's stages. This consistency suggests that while the models can capture some features distinguishing the disease stages, they may lack the necessary depth or complexity to fully address the nuanced patterns in MRI data. Additionally, challenges such as class imbalance and the inherent complexity of Alzheimer's-related MRI data likely impacted model effectiveness, despite the use of data augmentation and class weights. Future work could focus on exploring deeper, more specialized architectures, fine-tuning hyperparameters, or leveraging models pre-trained on medical imaging data to better capture subtle differences in Alzheimer's stages. Additional strategies, such as advanced data augmentation techniques or the use of synthetic data, may also help address class imbalance and improve model generalization. By incorporating these improvements, future models may achieve higher precision and recall, making them more valuable for early and accurate Alzheimer's detection, which is critical for timely intervention and better patient outcome

REFERENCES

[1] Classification of Alzheimer's disease using MRI data based on Deep Learning Techniques Author links open overlay panel Shaymaa E. Sorour a b , Amr A. Abd El-Mageed c , Khalied M. Albarrak a , Abdulrahman K. Alnaim a , Abeer A. Wafa d , Engy El-Shafeiy e.

[2] A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease Akhilesh Deep Arya 1, Sourabh Singh Verma 1, Prasun Chakarabarti 2,

[3] Ensemble of ROI-based convolutional neural network classifiers for staging the Alzheimer disease spectrum from magnetic resonance imaging Samsuddin Ahmed 1, Byeong C Kim 2,4, Kun Ho Lee 2,3,5, Ho Yub Jung 1,*; for the Alzheimer's Disease Neuroimaging Initiative.

[4] Multiscale Deep Neural Networks based analysis of FDG-PET images for the Early Diagnosis of Alzheimer's Disease February 2018Medical Image Analysis 46(3) DOI: 10.1016/j.media.2018.02.002 Lab: Functional and Anatomical Image Shape Analysis Lab. Donghuan LuDonghuan LuKarteek PopuriKarteek PopuriWeiguang DingWeiguang DingShow all 5 authorsMirza Faisal BegMirza Faisal Beg.

[5] Journal of King Saud University - Computer and Information Sciences.Shaymaa E. Sorour a b , Amr A. Abd El-Mageed c , Khalied M. Albarrak a , Abdulrahman K. Alnaim a , Abeer A. Wafa d , Engy El-Shafeiy e

[6] deep learning techniques in the effective diagnosis of Alzheimer's disease 1.Tulika Chakrabarti 2, Ahmed A Elngar 3,, Ali-Mohammad Kamali 4, Mohammad Nami 5

[7] Early Diagnosis of Alzheimer's Disease Donghuan LuKarteek PopuriKarteek PopuriWeiguang DingWeiguang DingShow all 5 authorsMirza Faisal BegMirza Faisal Beg.