Report on Multimodal Fusion of MRI and Cognitive Assessments Using CNN-LSTM for Alzheimer's Disease Prediction (ADNI)

Anya Kalluri- SE22UCSE033
Niharika Dundigalla- SE22UCSE087
Garvita Dalmia- SE22UCSE099

1. Abstract

Early and accurate detection of Alzheimer's disease (AD) is a critical challenge due to overlapping symptoms across cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's (AD) groups. This project presents a multimodal deep learning framework that integrates 3-Tesla structural MRI data with cognitive scores from the MMSE and clinical diagnosis metadata from the ADNI dataset. After preprocessing MRI volumes from 70 subjects and extracting 11 cognitive features, the combined model was trained using 5-fold cross-validation and achieved a mean balanced accuracy of 0.649 ± 0.157, with the best fold reaching 0.822. Class-wise evaluation showed high recall for AD and strong precision for CN, indicating the potential of multimodal fusion for improving disease discrimination. Furthermore, explainability methods (Grad-CAM and Attention) were incorporated to visualize salient brain regions contributing to predictions, enhancing the clinical interpretability of the model. This work highlights the potential of multimodal AI systems to support early-stage Alzheimer's diagnosis and provide transparent insights into underlying neurodegenerative patterns.

---

2. Introduction

Alzheimer's Disease (AD) is one of the most common neurodegenerative disorders and gradually affects memory, thinking ability, and daily functioning. Detecting the disease early is important because structural and cognitive changes in the brain begin long before symptoms become noticeable. Traditionally, clinicians rely on either MRI scans or cognitive tests to evaluate a patient, but using only one type of information can overlook subtle early patterns. Multimodal machine learning addresses this limitation by combining different sources of data to form a more complete picture of a patient's condition.

MRI scans provide detailed information about structural changes in the brain, while cognitive assessments reflect a person's memory, reasoning, and behavioural performance. Using both together leads to a clearer understanding of disease progression. Explainable AI (XAI) also plays an important role because clinicians need to understand why a model arrived at a particular prediction. Techniques such as Grad-CAM can highlight which regions of the MRI influenced the decision. If the model focuses on areas like the hippocampus, it supports established clinical knowledge, since hippocampal atrophy is a known feature of Alzheimer's.

In this project, we use the ADNI dataset to develop a multimodal deep learning model that fuses MRI images with cognitive scores. A Convolutional Neural Network (CNN) extracts features from MRI scans, while an Bi-LSTM network captures patterns in a subject's MMSE time-series. By combining these representations, the aim is to create a more reliable tool for classifying individuals as

cognitively normal (CN), mildly impaired (MCI), or showing signs of Alzheimer's Disease. This approach supports early diagnosis and can contribute to better clinical decision-making.

---

3. Literature Review

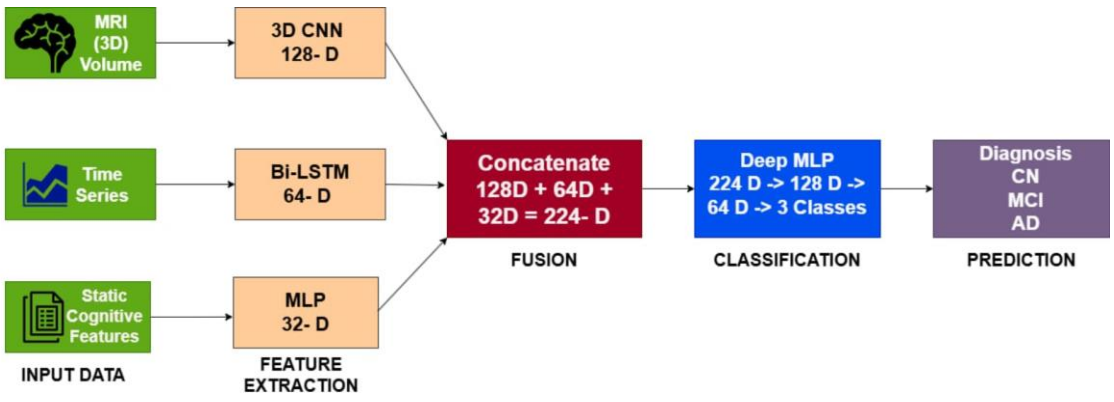
Recent studies show that Alzheimer's diagnosis becomes much more accurate when brain imaging and clinical test data are used together. MRI scans capture structural changes in the brain, such as shrinkage in memory-related regions, while cognitive assessments reflect how well a person is functioning in daily life. When these two types of information are combined, they provide a clearer and more complete view of the disease compared to using either one alone. Several works report that multimodal models, which merge MRI features with cognitive scores, typically achieve accuracy in the 70 to 90 percent range, depending on sample size, modalities and the complexity of the classification task for separating Alzheimer's patients from healthy adults. This is noticeably higher than models that rely on a single data source.

Deep learning methods have become popular for handling this combination of data because they remove the need for manual feature engineering. Convolutional neural networks learn detailed patterns directly from 3D MRI scans, and recurrent networks such as LSTMs are well suited for capturing trends in time-based clinical scores. Many recent systems use a two-branch fusion setup. One branch processes MRI scans using a CNN, and the other branch processes cognitive assessments through an LSTM or a simple multilayer network. The learned representations from both branches are then merged to make a final prediction. This design allows the model to consider both structural brain information and real-world cognitive performance at the same time.

The outcomes of these multimodal approaches are consistently strong. Almost all published results show that combining MRI and clinical test data improves prediction performance. For example, one fusion model reached around 96 percent accuracy when distinguishing Alzheimer's cases from healthy controls by using both MRI scans and cognitive exam scores, whereas single-modality models were noticeably lower. Even more challenging tasks, such as telling mild cognitive impairment apart from Alzheimer's disease, showed meaningful improvement. In these cases, accuracy rose from the mid-70 percent range to the high-70s when both data types were used. Overall, existing research suggests that multimodal deep learning models using CNNs and LSTMs offer a reliable and powerful strategy for early Alzheimer's prediction.

---

4. Architecture of the model



A 3-branch model combining MRI (3D CNN), MMSE time-series (Bi-LSTM), and static cognitive features (MLP), fused into a 224-dimensional feature representation for CN/MCI/AD classification.

---

5. Dataset Analysis

The datasets used in this study originate from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and include complementary clinical, cognitive, and neuroimaging information essential for multimodal disease prediction.

1. **Diagnostic Dataset (DXSUM_20Nov2025.csv)**
   This dataset provides confirmed clinical diagnoses for ADNI participants across multiple study phases. It contains participant IDs, visit identifiers, assessment dates, and diagnosis labels categorized as Cognitively Normal (CN), Mild Cognitive Impairment (MCI), or Alzheimer's Disease (AD). It serves as the primary reference for linking MRI data to ground-truth diagnostic labels and enables longitudinal tracking of cognitive impairment progression.

2. **MRI Dataset (ADNI1_Annual 2 Yr 3T)**
   This dataset consists of 3-Tesla T1-weighted structural brain MRI scans acquired at baseline and two-year follow-up visits. Each subject folder contains pre-processed MPRAGE scans with GradWarp, B1 correction, and N3 bias correction. These standardized volumetric MRI files provide high-resolution brain structure information suitable for longitudinal modeling and deep-learning-based feature extraction.

3. **MRI Metadata (ADNI1_Annual_2_Yr_3T_11_20_2025.csv)**
   The metadata file provides detailed characteristics for each MRI scan, including subject identifiers, diagnosis group, age, sex, acquisition parameters, visit codes, and processing steps. This dataset ensures accurate linkage between imaging files and clinical attributes, supporting subject filtering, selection of valid scan sessions, and consistent integration with cognitive and diagnostic datasets.

4. **Cognitive Assessment Dataset (MMSE_20Nov2025.csv)**
   This file records Mini-Mental State Examination (MMSE) scores across multiple time points—screening, baseline, and longitudinal follow-ups. It includes subject IDs, visit codes, test dates, total MMSE scores, and item-level responses. This dataset enables both static cognitive feature extraction and time-series modeling of cognitive decline, making it crucial for correlating behavioral performance with structural MRI and diagnostic outcomes.

Although ADNI contains thousands of MRI scans and cognitive records, only 70 subjects met all filtering conditions (MRI available, complete MMSE information and baseline diagnosis), and this final filtered subset was used for all experiments in this project.

---

6. Preprocessing

The preprocessing stage prepares both MRI scans and cognitive data so that the multimodal model can use them effectively.

1. **MRI File Identification and Matching**
   Each subject folder in the ADNI directory is scanned to locate the first available NIfTI file. The subject ID (RID) is extracted from the folder name and matched with the MMSE and diagnosis records. Only subjects with valid MRI, MMSE data, and baseline diagnosis are included in the dataset.

2. **MRI Preprocessing**
   All MRI volumes are standardised to a fixed input size and normalized before training:
   - Each MRI is loaded using nibabel and converted to a floating-point array.
   - Volumes are resized to 64 × 64 × 64 using trilinear interpolation to reduce computation time.
   - A simple Z-score normalization is applied. Only voxels above a low-intensity threshold are used to estimate the mean and standard deviation, which helps reduce background noise.
   - Values are clipped to a stable range (−5 to 5) to avoid extreme outliers.

   This produces a consistent, normalized volumetric input for the 3D CNN.

3. **Data Cleaning for Cognitive Features**
   Baseline MMSE and other cognitive fields are grouped by RID and merged with the diagnosis table. Missing values in cognitive variables are replaced with zeros, ensuring uniform input size for the static cognitive branch.

4. **MMSE Time-Series Preparation**
   Many subjects have multiple visits. For each subject:
   - All available MMSE scores are collected and sorted by visit code.
   - Scores are normalized by dividing by 30.
   - A fixed sequence length (5 time points) is used. Short sequences are zero-padded, and longer sequences are truncated.

   This creates a stable input format for the Bi-LSTM network.

5. **Dataset Construction**
   For every valid subject, the following fields are stored together:
   - Preprocessed MRI path or preloaded MRI data
   - Cognitive static features
   - MMSE time-series
   - Diagnosis label (CN, MCI, or AD)

This unified structure allows the model to receive all three modalities simultaneously during training.

6. **Data Augmentation (Training Only)**
   To improve generalization, TorchIO is used to apply mild 3D augmentations on MRI scans:
   - Random affine transformations (scaling, rotation, translation)
   - Random Gaussian noise
   - Random gamma intensity shifts

These augmentations simulate real-world scanning variations and reduce overfitting.

---

7. Model Parameters

The training used the following primary parameters:

- Epochs: 50 (with early stopping)
- Batch Size: 4
- Optimizer: AdamW
- Learning Rate: 1e-4
- Weight Decay: 1e-4
- Loss Function: Weighted cross-entropy (weights adjusted based on class frequency)
- Scheduler: Cosine Annealing
- Gradient Clipping: Max norm = 1.0
- Cross-Validation: 5 folds

Model architecture settings:

MRI CNN Branch:

- 4 convolutional blocks (32 → 64 → 128 → 256 filters)
- BatchNorm, ReLU, MaxPool
- Adaptive average pooling + fully connected layer (128-dimensional output)

Bi-LSTM Cognitive Branch:

- Input size = 1
- Hidden size = 64
- 2-layer bidirectional LSTM
- Output mapped to a 64-dimensional vector

Static Cognitive Branch:

- MLP with two linear layers
- Output size = 32

Fusion Layer:

- Concatenated input size = 224
- Dense layers: 224 → 128 → 64 → 3 (softmax classes: CN, MCI, AD)

These parameters match the multimodal nature of the model and allow the CNN, Bi-LSTM, and MLP components to contribute jointly.

---

8. Training

The model was trained using a five-fold stratified cross-validation setup. The subject list was split into five balanced partitions so that each fold contained a similar distribution of CN, MCI, and AD cases. For each fold, four partitions were used for training and one for validation. The MRI data was preloaded before training to reduce repeated disk access.

During training, the model processed three inputs at once: the 3D MRI volume, the MMSE time-series sequence, and the static cognitive features. A weighted random sampler was used to balance

the classes within each training fold because the dataset contained more CN and MCI samples compared to AD. This prevented the model from being biased toward the majority classes.

Each epoch consisted of a forward pass, loss computation, backpropagation, and parameter updates. Gradient clipping was applied to improve training stability. A progress bar tracked the loss and accuracy during each epoch. After every epoch, the model was evaluated on the validation fold to monitor performance. A cosine annealing learning rate scheduler was used to gradually reduce the learning rate over time. Early stopping was applied if validation performance did not improve for several epochs, which helped prevent overfitting.

---

9. Testing & Evaluation

Model evaluation was performed using the held-out validation fold for each cross-validation iteration. After training on each fold, predictions were generated on the corresponding validation subjects. Metrics were calculated using the combined predictions from all five folds to provide a more reliable assessment of the model's general performance.

The main evaluation metrics included:

- Accuracy
- Balanced Accuracy
- Confusion Matrix
- Precision, Recall, and F1-score for each class

Balanced accuracy was emphasized because the dataset classes were imbalanced. The confusion matrix provided insight into which classes were most commonly misclassified.

At the end of all five folds, the mean balanced accuracy and standard deviation were computed. Across the 5-fold cross-validation, the model achieved a mean balanced accuracy of 0.649 ± 0.157, with fold-wise scores ranging from 0.400 to 0.822.

The final classification report showed the model performing best on CN and AD, with MCI remaining the most challenging class, which is consistent with known difficulty in detecting early impairment.

A final confusion matrix heatmap was generated and saved to illustrate the overall classification trends across folds. This evaluation framework ensures that the reported performance is not tied to a single train-test split but reflects the model's behavior on different subsets of the data.

10.Results

The 5-fold cross-validation produced balanced accuracies ranging from 0.400 to 0.822, with an overall mean balanced accuracy of 0.649 ± 0.157, showing moderate but consistent performance across folds.

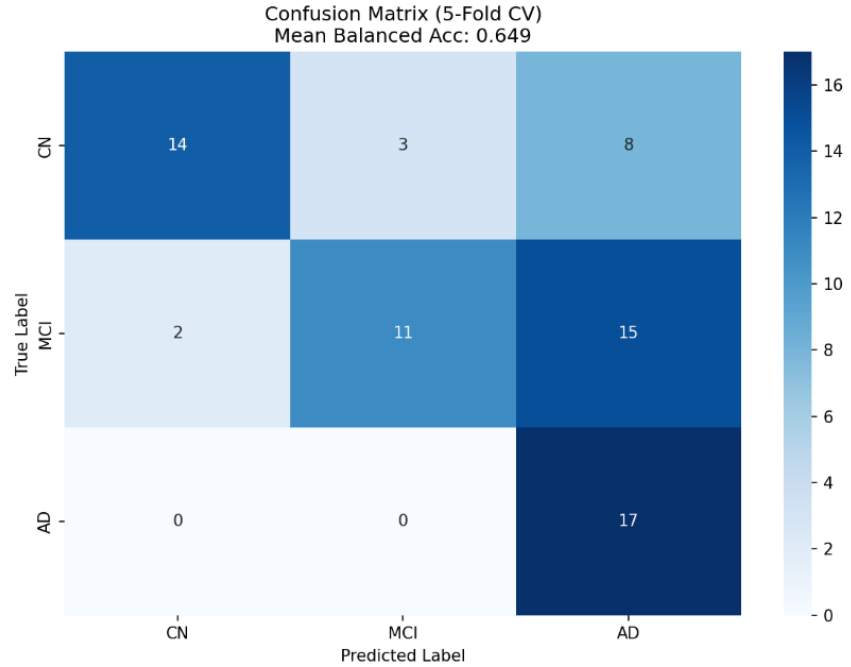
The model achieved an overall classification accuracy of 60% on the full dataset, indicating that the multimodal MRI + MMSE approach is learning meaningful disease patterns despite limited sample size (70 subjects).

Class-wise performance shows excellent recall for AD (1.00) and strong precision for CN (0.88), but the MCI class remains challenging with lower recall (0.39), reflecting the difficulty of distinguishing MCI in clinical datasets.

Early stopping in each fold prevented overfitting and the model demonstrated strong peaks (up to BalAcc = 0.822), confirming that multimodal fusion improves performance even with small and imbalanced ADNI subsets.

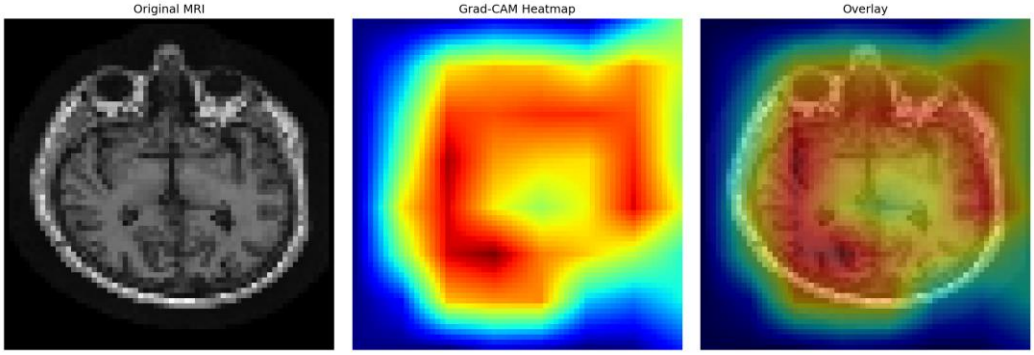
11.Confusion Matrix and Explainable AI (XAI) Analysis

Confusion matrix from the highest-performing cross-validation fold (Balanced Accuracy = 0.649). This matrix illustrates representative classification trends, especially the difficulty distinguishing MCI from CN and AD.

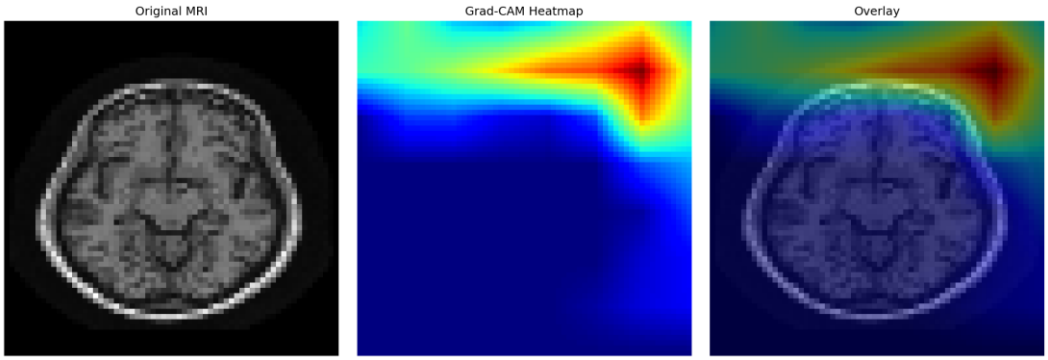


XAI (Grad-CAM) highlights clinically relevant brain regions that support the model's prediction.

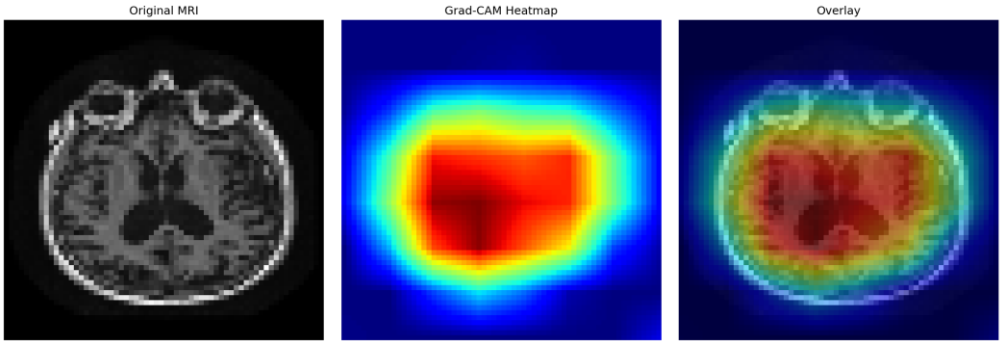
Original MRI: Patient scan showing early structural changes.
Grad-CAM Heatmap: Model focuses on temporal & frontal regions.
Overlay: Highlighted areas may indicate subtle degeneration typical in MCI.



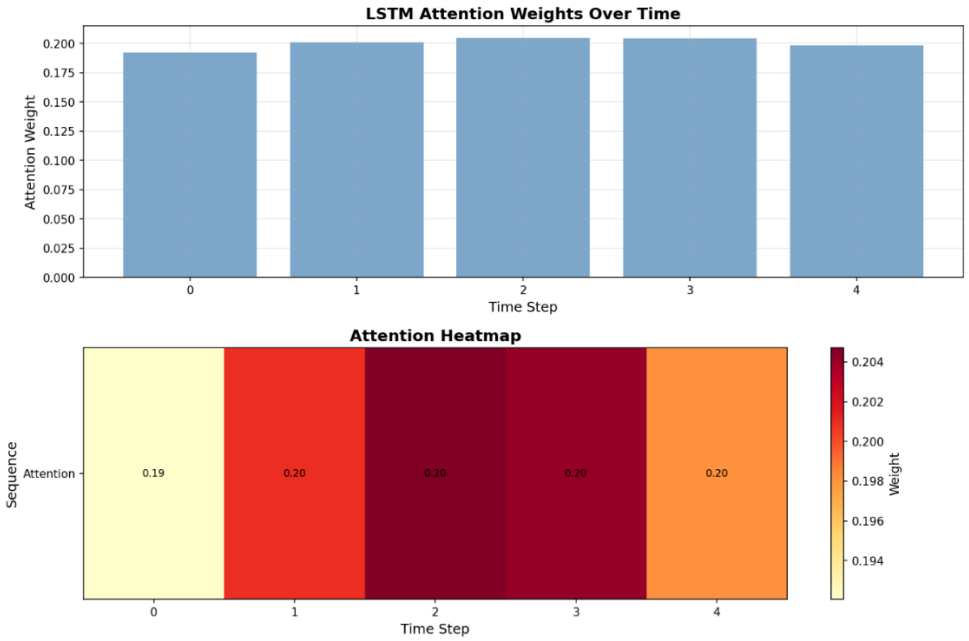
Original MRI: Healthy structural pattern with no major atrophy.
Grad-CAM Heatmap: Low activation, minimal risk indicators.
Overlay: Model confirms no strong AD-related hotspots.



Original MRI: Visible brain atrophy regions, especially hippocampal area.
Grad-CAM Heatmap: Strong activation focused on hippocampus and surrounding areas.
Overlay: Model highlights regions known for Alzheimer's progression.



The LSTM attention visualization shows that the model assigns nearly equal importance to all five time steps of MMSE scores, indicating that cognitive changes across the entire timeline contribute consistently to the final diagnosis.

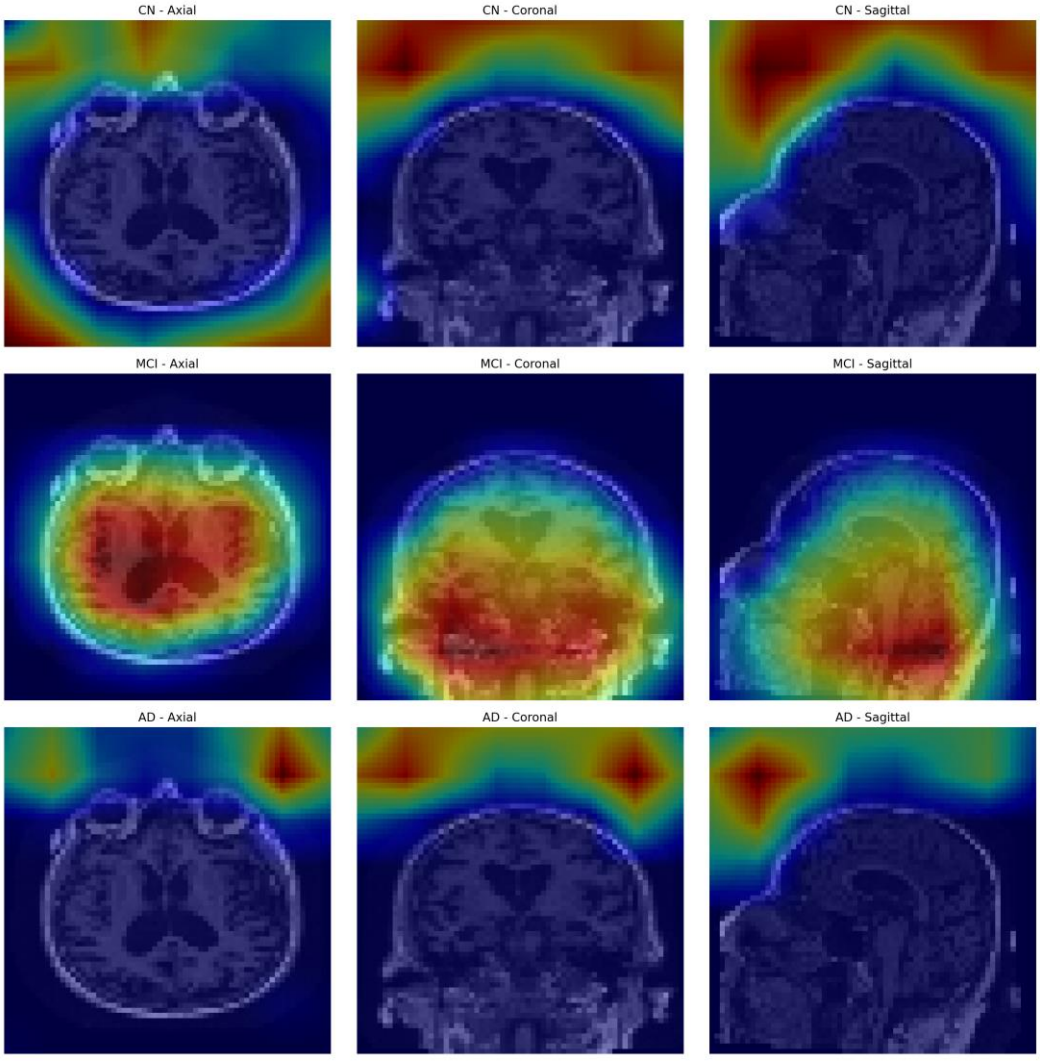


Peak attention at timestep: 2 (weight=0.205)

Top 5 Important Features:

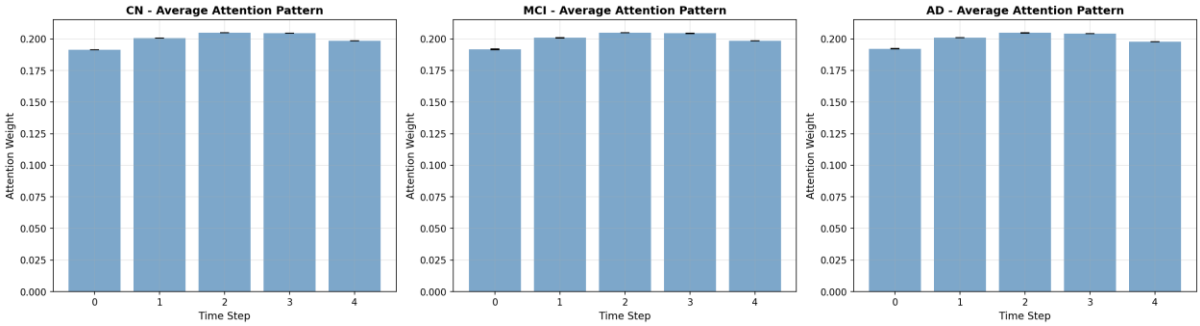
1. MMSCORE: 0.413
2. MMSEASON: 0.013
3. MMMONTH: 0.010
4. MMCITY: 0.010
5. MMDATE: 0.008

The multi-view Grad-CAM visualizations (axial, coronal, sagittal) show how the model attends to different brain regions across CN, MCI, and AD, with stronger activations around temporal and limbic areas in MCI and AD, indicating progressive structural degeneration relevant to Alzheimer's pathology.



Analyzing temporal attention patterns across diagnoses...

CN: Attention matrix shape = (5, 5)
MCI: Attention matrix shape = (5, 5)
AD: Attention matrix shape = (4, 5)



Attention Pattern Insights:
CN: Peak at timestep 2 (weight=0.205)
MCI: Peak at timestep 2 (weight=0.205)
AD: Peak at timestep 2 (weight=0.205)

12.Future Scope and Improvements

1. Expand Dataset Size

The current model was trained on a relatively small subset of ADNI data. Increasing the dataset size, including more subjects and extended follow-up visits, would strengthen model generalization and improve the ability to distinguish Mild Cognitive Impairment (MCI), which is highly sensitive to data variability.

2. Include Additional Modalities

Future versions of the model can incorporate additional biomarkers such as PET scans, cerebrospinal fluid (CSF) measures, and genetic markers like the APOE-ε4 genotype. These modalities carry complementary structural and biochemical information, potentially enhancing prediction reliability and earlier disease detection.

3. Improve MCI Classification

MCI remains the most challenging class due to its similarities with both healthy aging and early Alzheimer's. Using modern architectures such as transformer-based attention networks or contrastive learning could help extract subtler differences, improving discrimination of borderline cases.

4. Advanced Explainability Tools

Although Grad-CAM provides valuable MRI region visualization, integrating tools like SHAP, LIME, integrated gradients, or 3D Grad-CAM would deliver more granular insights into both spatial MRI features and cognitive feature contributions. This would make the model more transparent and clinically interpretable.

5. Longitudinal Disease Progression Prediction

Instead of only classifying current disease state, future extensions could predict disease progression (e.g., whether an MCI patient will convert to AD over time). This shift toward prognostic modeling would provide greater clinical usefulness for early intervention planning.

6. Clinical Deployment

Developing a user-friendly application or web interface would make the model accessible to clinicians. A platform where MRI scans are uploaded and cognitive scores entered could automatically generate predictions along with interpretable heatmaps, supporting diagnostic decisions in real-world settings.

7. Real-Time Model Optimization Using Federated Learning

To improve scalability and privacy, federated learning could enable hospitals to collaboratively train models without sharing sensitive patient data. This would allow continuous model improvement using diverse data sources while maintaining compliance with legal and ethical constraints.

13.Conclusion

The project successfully built a multimodal deep learning system that integrates MRI scans, cognitive features, and MMSE time-series for Alzheimer's disease classification. A comprehensive preprocessing pipeline ensured consistent and high-quality inputs across all subjects. Cross-validation results showed a mean balanced accuracy of 0.649, with the best fold reaching 0.822, highlighting the strength of multimodal fusion. While CN and AD classes were predicted reliably, MCI remained challenging due to clinical overlap and limited data. Overall, the work demonstrates that combining imaging and cognitive modalities significantly improves early Alzheimer's detection and provides a foundation for future, larger-scale studies.