Deep Multi-Modal Detection of Early Alzheimer's

Team Members:

Ali Alsaihati, Jennifer Li, Maria Faustina Inoma, Rachel Yu, Ryan Banko

Supervised By:

Dr. Tan Songdechakraiwut

Dr. Michael Lutz

10 August 2025

Table of Contents

1. Project Objectives and Goals 3
2. Current Progress and Milestones Achieved 3
3. Methodology and Approach 4
4. Preliminary Results 4
5. Timeline and Schedule 5
6. Challenges and Obstacles Encountered 5
7. Next Steps and Future Work 5
8. Resources and Tools Used 6

1. Project Objectives and Goals

The primary objective of this project is to predict future cognitive impairment in individuals at risk of Alzheimer's Disease (AD) using machine learning. Our goals include:

- Integrating structural MRI (sMRI), functional MRI (fMRI), and clinical data.
- Constructing a unified functional connectome from fMRI to extract topological features.
- Capturing both current state and change over time.
- Developing a machine learning model capable of forecasting cognitive scores 12 months into the future.
- Identifying individuals at risk of cognitive decline with high accuracy and interpretability.

2. Current Progress and Milestones Achieved

Data Acquisition:

- Successfully accessed and downloaded data from the ADNI database, including sMRI, fMRI, and clinical data for 1302 subjects.

Preprocessing Completed:

- Converted DICOM to BIDS-compliant format.
- Executed full fMRIPrep pipeline across 1027 valid subjects.

Feature Extraction:

- Structural features from sMRI.
- Topological features from fMRI-based functional connectomes.
- Demographic and clinical features (e.g., MMSE, genotype, pulse, blood pressure, gender).

Model Development:

- Built and evaluated regression models for predicting future cognitive scores.
- Conducted binary classification to detect decline using prediction deltas.

Poster Presentation:

- Project results were presented in a structured scientific poster.

3. Methodology and Approach

Data Sources:

- fMRI and sMRI scans from ADNI.
- Clinical data including cognitive scores, vitals, genetics, and demographics.

Preprocessing:

- Utilized Duke's CS cluster and fMRIPrep for fMRI preprocessing.
- Converted raw DICOM files to NIfTI format and reorganized data into BIDS format via custom shell and Python scripts.
- Cleaned and validated subject folders; 275 subjects were discarded due to incomplete data.

Feature Engineering:

- Constructed unified functional connectomes and extracted topological features (e.g., persistence statistics).
- Extracted anatomical features from sMRI (e.g., brain region volumes).
- Computed temporal change between past and current scans for both fMRI and sMRI.
- Incorporated clinical features into the feature set.

Modeling:

- Trained a regression model (best: XGBoost) with log-transformed cognitive scores.
- Evaluated using subject-level cross-validation and randomized hyperparameter tuning.
- Derived binary classification labels from score deltas to assess cognitive decline.

4. Preliminary Results

- The regression model demonstrated strong predictive performance for forecasting cognitive scores 12 months in advance.

- Log-transformation of target values helped mitigate skewness and improved model fit.
- Top-performing model (XGBoost) outperformed other tested models.
- In classification settings, the model accurately identified subjects experiencing cognitive decline, achieving meaningful metrics on precision, recall, and F1-score.
- Combining topological features from fMRI with structural and clinical features significantly improved predictive performance over single-modality baselines.

5. Timeline and Schedule

Week	Milestone
1–2	Planning, learning more about the domain, getting familiar with the dataset
3–4	Data Acquisition & Preprocessing
5–7	Model Development
8	Poster day, submitting deliverables

6. Challenges and Obstacles Encountered

High Data Volume: Downloading and preprocessing over 1300 subjects required significant storage and compute resources.

IP Binding on ADNI: Required use of the same cluster node to generate and download scan URLs.

Resource Limitations: Cluster enforced per-user CPU limits.

Data Loss: 275 subjects were removed due to missing or corrupted data during the BIDS validation step.

BIDS Conversion Complexity: Required writing and adapting multiple scripts to ensure correct metadata formatting and naming conventions.

7. Next Steps and Future Work

- **Longitudinal Modeling**: Develop time-series or RNN-based models to capture sequential scan data.

- **Expanded Evaluation**: Use more comprehensive clinical metrics and different time horizons (e.g., 24-month prediction).
- **Collaboration**: Consider working with clinical researchers to validate model predictions in real-world cohorts.

8. Resources and Tools Used

- Python (pandas, numpy, scikit-learn, xgboost)
- Duke CS Cluster (storage, bash script, slurm jobs, xquartz)
- ADNI Dataset (clinical, sMRI, fMRI)
- fMRIPrep
- dcm2niix
- Google colab
- Google Slides
- Google Docs