AI Model(s) Used in Research Project

1.1 Models Selected

For this study, we utilized multiple machine learning and deep learning models to predict and classify early-stage Alzheimer's Disease (AD). The models include:

- Random Forest (RF)
- XGBoost
- 1D Convolutional Neural Network (1D CNN)

1.2 Architecture and Key Components

1. Random Forest (RF)
    - Architecture/Key Components:
        - An ensemble of decision trees, each trained on a bootstrapped subset of the training data.
        - Uses a random subset of features to split each node.
        - Final prediction is typically done via majority vote across all trees.
    - Why This Model?
        - Handles high-dimensional data well.
        - Robust to outliers and missing data.
        - Often a strong baseline for classification tasks in medical contexts.
2. XGBoost
    - Architecture/Key Components:
        - Gradient-boosted trees, which iteratively add new trees that correct errors made by existing trees.

        - Implements second-order gradient-based optimization for improved speed and accuracy.
        - Features built-in regularization to reduce overfitting.
    - Why This Model?
        - State-of-the-art performance in many structured data problems.
        - Known for fast training and good scalability.
        - Regularization helps handle smaller, imbalanced medical datasets effectively.

3. 1D Convolutional Neural Network (1D CNN)
    - Architecture/Key Components:
        - Input layer shaped as (number_of_features, 1), treating each feature as a "time step."
        - Convolutional layer(s) with 1D filters to learn local feature patterns.
        - Pooling layers to reduce dimensionality and capture dominant signals.
        - Dense (fully connected) layers for classification output.
    - Why This Model?
        - CNNs can automatically learn important feature representations.
        - Potential to uncover subtle relationships between variables when the data is carefully structured.
        - Convolution and pooling operations can capture local patterns relevant to biomarkers or brain region measurements.

1.3 Rationale for Model Choices

- Clinical Relevance: Tree-based models (RF and XGBoost) provide feature importance, helping clinicians understand which brain regions or demographic measures are most critical.

- Performance and Interpretability: Random Forest and XGBoost offer strong classification performance plus interpretable feature importance. The CNN can potentially capture complex relationships not easily modeled by traditional methods.

- Versatility: Comparing a classic ensemble approach with a deep learning approach allows us to identify the best performer in terms of accuracy, recall, and other relevant metrics for early Alzheimer's detection.

---

2. Performance Metrics Analysis

2.1 Metrics Being Tracked

1. Accuracy
    - Measures the overall fraction of correct predictions.
    - Significance: In a diagnostic setting, a quick snapshot of how often the model correctly classifies patients as having Alzheimer's or not.
2. Precision
    - For each class, precision = (True Positives) / (True Positives + False Positives).
    - Significance: Indicates how many patients predicted positive for dementia actually have it (important to reduce false positives).
3. Recall (Sensitivity)
    - For each class, recall = (True Positives) / (True Positives + False Negatives).
    - Significance: Measures the model's ability to detect cases of AD (important to reduce missed diagnoses).
4. F1-Score
    - Harmonic mean of precision and recall.
    - Significance: Balances the trade-off between precision and recall, especially useful for imbalanced classes (mild cognitive impairment or early AD often have fewer samples).

5. Confusion Matrix

    - Displays counts of true positives, false positives, false negatives, and true negatives for each class.

    - Significance: Helps understand the distribution of errors (e.g., which classes are commonly confused).

2.2 Current Values of Each Metric

	Class	Model	precision	recall	f1-score	support
0	AD	1D CNN	0.83	0.71	0.77	7.0
1	AD	1D CNN + SMOTE	0.75	0.43	0.55	7.0
2	AD	RF + SMOTE	0.75	0.50	0.60	6.0
3	AD	RandomForest	0.60	0.50	0.55	6.0
4	AD	XGBoost	0.62	0.83	0.71	6.0
5	AD	XGBoost + SMOTE	0.71	0.83	0.77	6.0
6	CN	1D CNN	0.86	0.94	0.90	32.0
7	CN	1D CNN + SMOTE	0.94	0.94	0.94	32.0
8	CN	RF + SMOTE	0.97	0.90	0.93	31.0
9	CN	RandomForest	0.88	0.97	0.92	31.0
10	CN	XGBoost	0.94	0.94	0.94	31.0
11	CN	XGBoost + SMOTE	0.97	0.97	0.97	31.0
12	MCI	1D CNN	0.71	0.62	0.67	16.0
13	MCI	1D CNN + SMOTE	0.68	0.81	0.74	16.0
14	MCI	RF + SMOTE	0.65	0.85	0.73	13.0
15	MCI	RandomForest	0.64	0.54	0.58	13.0
16	MCI	XGBoost	0.82	0.69	0.75	13.0
17	MCI	XGBoost + SMOTE	0.83	0.77	0.80	13.0

2.3 Significance in the Context of This AI Medical/Health Project

- AD (Alzheimer's Disease): Highest F1-score = 0.77 (tied between *1D CNN* and *XGBoost + SMOTE*).
- CN (Cognitively Normal): Highest F1-score = 0.97 (*XGBoost + SMOTE*).
- MCI (Mild Cognitive Impairment): Highest F1-score = 0.80 (*XGBoost + SMOTE*).

Overall, *XGBoost with SMOTE* tends to perform best (especially for CN and MCI), while *1D CNN* performs competitively on the AD class..

2.4 Comparison with Expected Benchmarks

Recent literature on Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI) classification, using structural MRI features or similar clinical markers, often reports classification accuracies ranging from about 75% to 95% across different modeling approaches and study populations. In particular: Random Forest and XGBoost approaches in published research commonly achieve 80%–90% accuracy (sometimes higher), with F1-scores also in the high 0.70–0.90 range, depending on how balanced the dataset is and how many features are used. Deep Learning models (like CNNs) can reach or surpass these metrics but often require larger datasets or advanced data augmentation to avoid overfitting. Our current experiments show that: XGBoost with SMOTE performed best overall, achieving F1-scores up to 0.97 for CN and around 0.77–0.80 for AD and MCI, which aligns well with the upper tier of reported values in the literature—especially for the cognitively normal (CN) class. Random Forest and 1D CNN are also competitive, but each has particular strengths for certain classes (e.g., 1D CNN for AD detection, or Random Forest for robust overall performance before SMOTE). Considering that our dataset may still be limited in size and potentially imbalanced across diagnostic classes, these results are promising. They meet or exceed benchmarks in similar studies that rely on similar clinical data (e.g., brain region volumes, MMSE, and patient demographics). Future improvements—for example, hyperparameter tuning, additional biomarkers, or more robust oversampling—could further elevate our performance into the upper ranges (90%+ in F1-score) as reported by leading AD/MCI classification research.

3. Project Status Summary

Overall Progress

Our project is on track to be and most probably would be completed by April 18th, with most core components in place:

1. Data Preprocessing

    - We have successfully preprocessed our existing data from OASIS-1, OASIS-2, ADNI, and NACC. This includes handling missing values, normalizing features, and performing SMOTE-based oversampling for imbalanced classes.

2. Model Development and Evaluation

    - We have trained and compared six different model configurations (Random Forest, Random Forest + SMOTE, XGBoost, XGBoost + SMOTE, 1D CNN, 1D CNN + SMOTE).

    - Early results show promising performance metrics (F1-scores, precision, recall) for diagnosing AD, CN, and MCI.

3. Initial Findings

    - We have identified XGBoost + SMOTE as a top performer for CN and MCI detection, while the 1D CNN performed competitively for AD detection in certain metrics.

    - These findings align well with published benchmarks in Alzheimer's detection research.

Missing Data / Challenge

Despite this overall progress, we are missing a crucial portion of the dataset, referred to as "dementia Bank" This missing data may include:

- Caregiver logs or advanced clinical details crucial for the Caregiver Support module (e.g., daily symptom tracking, medication adherence).

- Speech or neuropsychiatric data necessary for advanced NLP analyses in Speech and Cognitive Decline Monitoring.

Impact:

- The missing data slightly limits the depth of our analysis for the caregiver and speech components. We have robust pipelines ready for integration, but these components remain partially incomplete until the dementia data is received.

Corrective Action Plan

1. Data Acquisition (by April 15)

    - We are in the process of obtaining the missing dementia data from our collaborators. Once secured, we will integrate it into our existing data pipeline, which already supports new feature sets.

2. Re-Training and Validation (by April 16)

    - After incorporating the missing data, we will re-run the preprocessing steps (cleaning, scaling, SMOTE) and retrain our models to evaluate any changes in performance.

    - This will specifically enhance the caregiver and speech analysis components, ensuring a more comprehensive evaluation.

3. Final Results and Documentation (by April 17)

    - With all data integrated, we will finalize model comparisons, generate updated performance metrics, and prepare our final documentation.

Conclusion

Given the strong progress on data preprocessing, model training, and evaluation thus far, we are pretty confident in delivering a complete AI-powered Alzheimer's system by April 18th. The only gap pertains to the missing dementia data, but we have a clear plan and timeline to incorporate it, thereby ensuring that our modules for caregiver support and speech-based cognitive monitoring are fully operational and integrated before the final deadline.