

Explainable Machine Learning for Early Alzheimer's Risk Detection

Team Name: Clever Cubed

Hackathon: AI 4 Alzheimer's – Hack4Health

Team Members: Aadya Patel | Ananya Mishra | Anish Kushwaha

Abstract

Alzheimer's disease is a progressive neurodegenerative disorder that often goes undiagnosed until cognitive decline becomes severe. Early detection remains a critical challenge due to the complexity and heterogeneity of symptoms. This project explores the use of machine learning techniques to support early Alzheimer's risk identification using de-identified clinical and cognitive data.

We present an interpretable machine learning pipeline that classifies individuals into cognitive states such as cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD). Beyond prediction accuracy, we emphasize explainability using feature importance analysis and SHAP values to ensure transparency and trust. The system is designed for research and educational purposes, supporting early risk awareness rather than clinical diagnosis.

1. Introduction

Alzheimer's disease affects millions worldwide and poses significant social and economic challenges. One of the main difficulties in managing Alzheimer's is the lack of early and reliable detection mechanisms. Subtle cognitive and clinical changes may appear years before a formal diagnosis, but these signals are often overlooked or difficult to interpret.

Recent advances in machine learning have shown promise in analyzing complex biomedical data. However, many high-performing models operate as black boxes, limiting their usefulness in healthcare contexts where interpretability is essential. This project aims to bridge that gap by building an explainable machine learning system that balances predictive performance with transparency.

2. Dataset Description

The project uses a de-identified Alzheimer's dataset provided by the hackathon organizers. The dataset contains a combination of:

- Demographic features (e.g., age, gender)
- Cognitive assessment scores
- Clinical and health-related measurements

The target variable represents cognitive status categories such as CN, MCI, and AD. As with most real-world biomedical datasets, the data includes missing values and variability across features, requiring careful preprocessing.

3. Methodology

3.1 Data Preprocessing

- Removed records with missing target labels
- Handled missing feature values using mean imputation
- Normalized numerical features where required
- Split data into training (80%) and testing (20%) sets using stratified sampling

3.2 Model Development

We implemented and compared multiple machine learning models:

- **Logistic Regression** as a baseline linear classifier
- **Random Forest Classifier** to capture non-linear relationships
- **Gradient Boosting / XGBoost** for improved predictive performance

Model evaluation was performed using:

- Accuracy
- Precision and Recall
- Confusion Matrix
- ROC-AUC score (where applicable)

3.3 Explainability

To ensure interpretability:

- Feature importance scores were extracted from tree-based models
- SHAP (SHapley Additive exPlanations) values were used to explain both global and individual predictions

These techniques allowed us to identify which features most strongly influenced model outputs, improving transparency and trust.

4. Results

Among the tested models, tree-based methods such as Random Forest and Gradient Boosting demonstrated superior performance compared to the baseline logistic regression.

Key observations include:

- Cognitive assessment scores were among the most influential features
- Age-related features showed strong contributions to prediction
- Explainability analysis revealed consistent feature importance across models

The confusion matrix and SHAP visualizations indicate that the model learns meaningful patterns rather than relying on noise.

5. Challenges and Limitations

Several challenges were encountered during development:

- Handling missing and noisy biomedical data
- Avoiding overfitting while maintaining reasonable accuracy
- Interpreting complex model behavior in a clinically meaningful way
- Limited dataset size and lack of longitudinal data

Limitations of this work include:

- The model is not validated on external datasets
 - Predictions are not suitable for direct clinical use
 - The dataset represents a simplified snapshot rather than disease progression over time
-

6. Ethical Considerations

This project is designed with ethical responsibility in mind. The model does **not** provide medical diagnoses or treatment recommendations. Instead, it supports research and early risk awareness. All data used is de-identified, and results should be interpreted only by qualified professionals.

7. Conclusion and Future Work

This project demonstrates that explainable machine learning can support early Alzheimer's risk analysis while maintaining transparency and trust. By combining predictive modeling with interpretability techniques, the system highlights meaningful clinical patterns without functioning as a black box.

Future work includes:

- Incorporating additional public datasets
- Exploring longitudinal modeling for disease progression

- Improving uncertainty estimation and calibration
 - Collaborating with healthcare professionals for validation
-

Disclaimer

This project is intended for **research and educational purposes only** and does **not** provide medical diagnosis or clinical advice.