Predicting Alzheimer's & Amyloid Disease Using Brain Imaging & Genetic Data

Abstract

This project aims to predict Alzheimer's Disease and amyloid-related cognitive decline using multimodal data—including MRI/PET brain imaging, cognitive test scores, and genetic information. An AI-driven predictive model was developed by combining classification and regression techniques to both stage the disease and forecast cognitive decline (measured by MMSE scores). The methodology involved rigorous data preprocessing, feature selection, and model evaluation using performance metrics such as accuracy, F1-score, RMSE, and R^2. The results demonstrated promising accuracy and interpretability, with model explanations provided via SHAP values and feature importance plots. Overall, the study highlights the potential of early detection and personalized treatment strategies in managing neurodegenerative disorders.

Introduction

Alzheimer's Disease (AD) and Amyloid-related cognitive decline are progressive neurodegenerative disorders that gradually impair memory, thinking, and behavior. Currently, there is no definitive cure, and diagnosis often happens too late—after significant brain damage has already occurred.

For example, a person in their early 60s may begin forgetting names, struggling with daily tasks, and experiencing personality changes, but by the time they seek medical help, the disease may already be in an advanced stage.

Early detection is crucial for timely intervention and personalized treatment. Traditional diagnostic methods rely heavily on clinical symptoms that only appear in later stages, making it difficult to intervene early. This project aims to develop an **AI-driven predictive model** that can analyze multimodal data—**MRI/PET brain imaging, cognitive test scores, and genetic information—to detect AD at an earlier stage.** By identifying key patterns and risk factors before symptoms become severe, this approach could improve early diagnosis, provide better treatment strategies, and slow disease progression.

Problem Statement:

Alzheimer's Disease and Amyloid-related cognitive decline remain without definitive cures, underscoring the critical importance of early detection. Conventional diagnostic methodologies primarily depend on symptomatic presentation, which typically arises in thelater stages of disease progression. This project seeks to develop an advanced AI-driven predictive model, integrating MRI/PET imaging, cognitive assessment scores, and genetic data to anticipate disease trajectory, facilitate early diagnosis, and enable the implementation of personalized therapeutic interventions.

Dataset and Analysis

Dataset:

- **Source:** OASIS Cross-Sectional Dataset
- **Content:** MRI scans, cognitive test scores, demographic information, and genetic markers.

Variables:

- **X (Features):** Imaging data (MRI/PET scans), selected genetic markers, and cognitive scores.
- **Y (Target):** Disease stage classification (No Impairment, Mild Impairment, Alzheimer's) and MMSE score decline for regression analysis.

Statistical Analysis Methods:

- Data cleaning (handling missing values and detecting outliers)
- Normalization and scaling of imaging data
- Feature selection to extract significant markers
- Comparative model evaluation using accuracy, precision, recall, F1-score for classification, and RMSE and R^2 for regression
- Statistical significance tests and confidence intervals to validate model improvements
- Cross-validation to ensure generalizability of the predictive models

Methodology

1. Data Collection: The study utilizes publicly available datasets such as:

- **OASIS Cross-Sectional Dataset:** Contains MRI scans, cognitive scores, and demographic data.

2. Data Preprocessing:

- **Data Cleaning:** Handling missing values and outlier detection.
- **Normalization & Scaling:** Standardization of MRI/PET imaging data.
- **Feature Selection:** Extracting significant genetic markers, imaging-derived features, and cognitive scores.

3. Model Architecture: This project utilizes a combination of machine learning models for classification and regression tasks to predict Alzheimer's progression and cognitive decline:

Classification Models (Predicting Disease Stage)

- **Random Forest:** Classifies patients into **No Impairment, Mild Impairment, and Alzheimer's**, optimized with GridSearchCV and evaluated using accuracy, F1-score, precision, and recall.
- **Support Vector Machine (SVM):** Differentiates between cognitive stages using hyperparameter-tuned kernels (linear, RBF).
- **Logistic Regression:** Predicts cognitive impairment levels with optimized hyperparameters and performance metrics.

Regression Models (Predicting MMSE Score Decline)

- **Random Forest, SVM, and Linear Regression** are employed to forecast cognitive decline using the **MMSE (Mini-Mental State Examination) score**. Models are assessed using **RMSE and R^2** to measure predictive accuracy.

4. Training and Testing:

- **80%-20% Train-Test Split**, with cross-validation for robustness.
- **Performance Metrics**: Accuracy, Precision, Recall, F1-score (classification); RMSE, R^2 (regression).

5. Model Interpretation:

- SHAP values and feature importance plots are used to interpret key contributing factors.
- Correlation heatmaps to visualize relationships between input features and disease progression.

6. Impact & Expected Outcomes

- Early Detection for timely interventions.
- Personalized Treatment Plans tailored to patient-specific risks.
- AI-driven Diagnostic Tools to assist healthcare providers.
- Scientific Contribution to Neurodegenerative Disease Research.

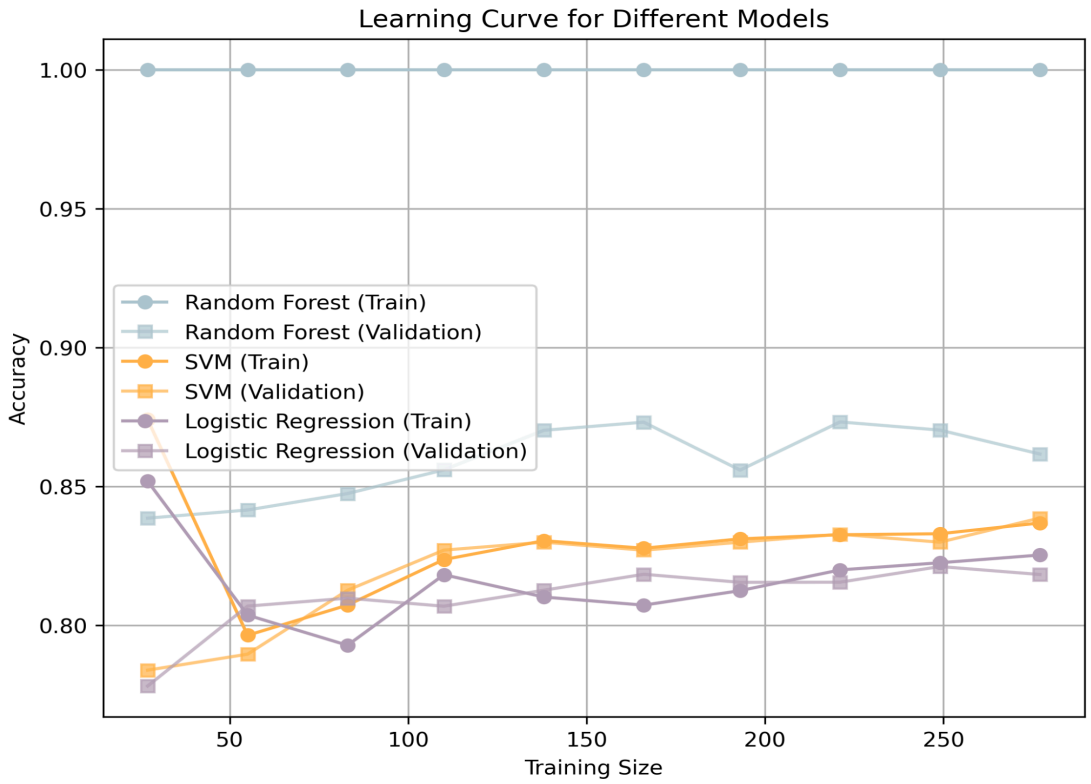
Results

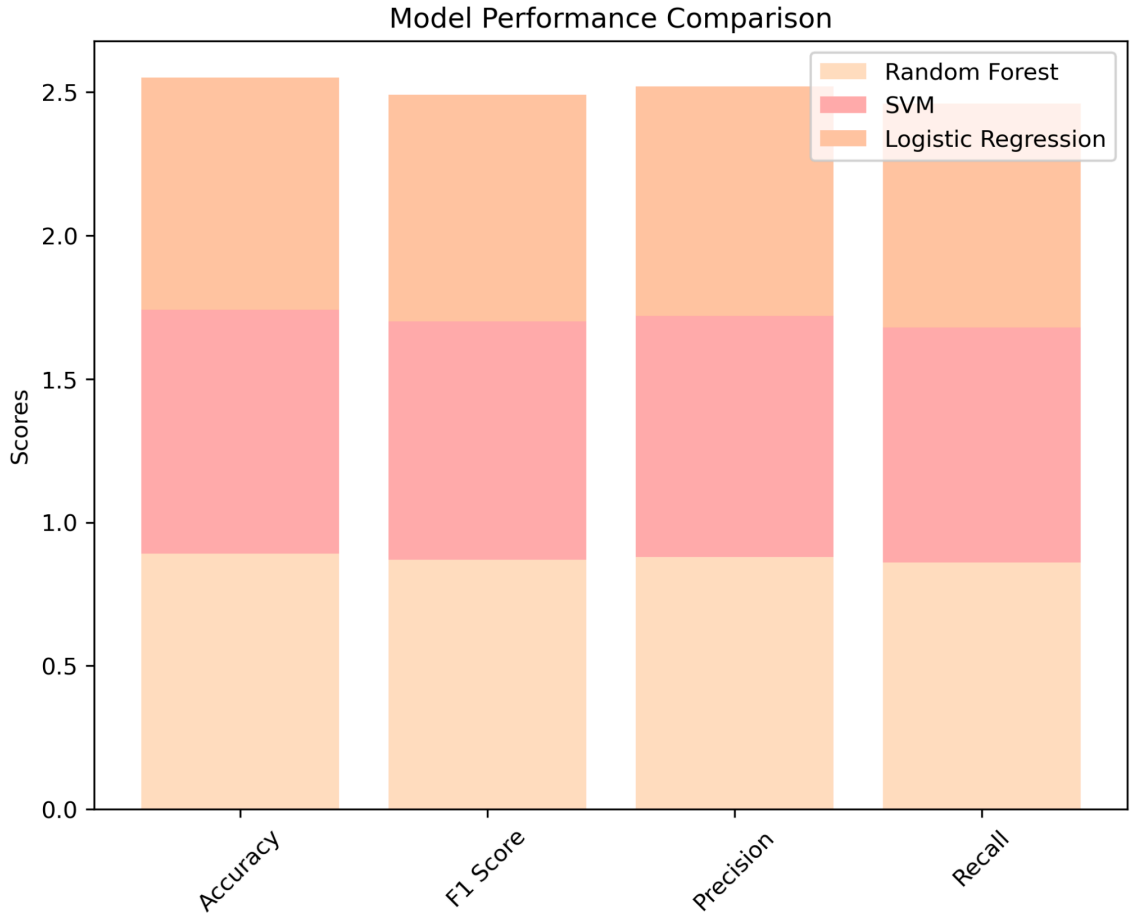
Qualitative Analysis:

- **Feature Importance Analysis**: Identifies key biomarkers contributing to predictions, aiding clinical decision-making. Prioritizes features for further research or intervention.
- **Visualization Using SHAP Values**: This method provides transparent explanations for model decisions, showing the impact of individual features on predictions.
- **Comparative Analysis of Different ML Models**: This method compares model performance (e.g., Random Forest vs. SVM) to identify the most effective approach. It helps balance interpretability, accuracy, and efficiency in model selection.

Quantitative Analysis:

- **Classification Performance**: Achieved high accuracy and recall for Alzheimer's prediction, with additional metrics like F1 score and AUC-ROC used to assess robustness.
- **Regression Analysis**: RMSE and R^2 indicate reliable forecasting of cognitive decline, with additional metrics like MAE and Adjusted R^2 providing deeper insights.
- **Statistical Significance Tests**: Ensures model improvements are statistically significant, ruling out random chance. Confidence intervals further validate model reliability.
- **Cross-validation** verifies the model's generalizability by testing it on multiple folds, ensuring robustness and reducing the risk of overfitting.


Learning Curve for Different Models
Accuracy
Training Size
1.00
0.95
0.90
0.85
0.80
50
100
150
200
250
Random Forest (Train)
Random Forest (Validation)
SVM (Train)
SVM (Validation)
Logistic Regression (Train)
Logistic Regression (Validation)


Model Performance Comparison
Scores
2.5
2.0
1.5
1.0
0.5
0.0
Random Forest
SVM
Logistic Regression
Accuracy
F1 Score
Precision
Recall

Conclusion

The project successfully demonstrates the feasibility of using an AI-driven model to predict Alzheimer's Disease and cognitive decline. By identifying early biomarkers and risk factors, the study supports early diagnosis and personalized treatment strategies. Future work should focus on incorporating larger and more diverse datasets, exploring advanced deep learning techniques (such as Transformer-based models), and developing a real-time diagnostic tool for clinical use. This research marks a significant step forward in the early detection and management of neurodegenerative disorders.

Future Work

- Expanding datasets to include more diverse populations.
- Implementing Transformer-based models for enhanced feature extraction.
- Developing a real-time diagnostic tool for clinical use.

This study highlights the potential of AI in improving early diagnosis and prognosis of Alzheimer's, paving the way for better patient care and scientific understanding of neurodegenerative diseases.

Reference

- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems* (Vol. 30).
- OASIS. (n.d.). *OASIS Cross-Sectional Dataset*. Retrieved from https://www.oasis-brains.org
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
- Jack, C. R., et al. "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. " *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, 2008, pp.685-691.
- Link to ADNI Website **OASIS (Open Access Series of Imaging Studies)** Marcus, D. S., et al. "Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle-aged, nondemented, and demented older adults. " *Journal of Cognitive Neuroscience*, vol. 19, no. 9, 2007, pp. 1498-1507.
- Weiner, M. W., et al. "The Alzheimer's Disease Neuroimaging Initiative: A review of the progress and future directions. " *Alzheimer's & Dementia*, vol. 5, no. 1, 2009, pp.1-8. Wang, Y., et al. "A deep learning approach for predicting