Early Detection of Alzheimer's Disease

Habiba Mohsen Ateya
Systems and Biomedical Engineering
Cairo University
habiba.shaarawy01@eng-st.cu.edu.eg

Hazem Raafat
Systems and Biomedical Engineering
Cairo University
hazem.gomaa02@eng-st.cu.edu.eg

Hana Hesham
Systems and Biomedical Engineering
Cairo University
hana.morsi02@eng-st.cu.edu.eg

Malak Nasser
Systems and Biomedical Engineering
Cairo University
malak.mohamed021@eng-st.cu.edu.eg

***Abstract*—This paper explores the application of machine learning techniques to detect Alzheimer's disease using MRI scan data. The focus is on evaluating different algorithms' effectiveness and improving diagnostic accuracy. Our results show promising directions for future research in automated Alzheimer's detection.**

Keywords— *Alzheimer's disease, early detection, predictive modeling, neurodegenerative diseases, automated diagnosis*

I. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that significantly impacts cognitive function, memory, and behavior. It is the leading cause of dementia among older adults, affecting millions of individuals worldwide. The early detection of Alzheimer's disease is crucial for effective management and treatment, as it allows for timely interventions that can slow the progression of the disease, improve quality of life, and enable better planning for patients and their families. However, diagnosing AD in its early stages remains a significant challenge due to the subtlety of initial symptoms and the reliance on subjective clinical assessments.

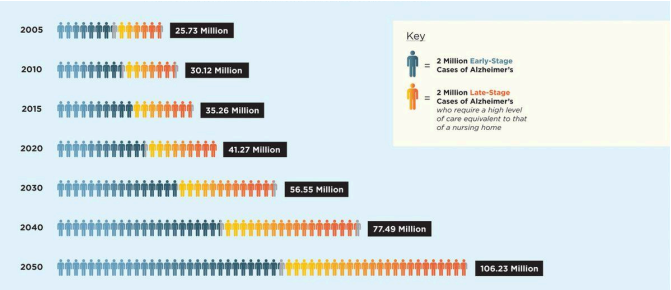

Figure 1. Projections of ALZHEIMER'S prevalence

The motivation behind this study stems from the urgent need to enhance the early detection of Alzheimer's disease through automated methods. By leveraging machine learning techniques, we aim to develop a diagnostic tool that can analyze MRI data more quickly and potentially with greater accuracy than traditional methods. Machine learning algorithms can identify complex patterns and subtle changes in brain structure that may be indicative of AD, providing a more objective and consistent approach to diagnosis.

In this study, the input comprises MRI scan data formatted as CSV files from the OASIS-2 dataset, an open-access series of imaging studies provided by Washington University in St. Louis. The output of this study is the classification of each MRI scan, indicating the presence or absence of Alzheimer's disease. By evaluating the effectiveness of different machine learning algorithms, we seek to improve diagnostic accuracy and contribute to the development of reliable, automated tools for early Alzheimer's detection.

II. Related Works

The detection and diagnosis of Alzheimer's disease have been explored through various computational approaches. This literature review categorizes existing studies based on their methodologies and discusses their comparative strengths, weaknesses, and their relation to our project.

1. Traditional Machine Learning Techniques:

- **Logistic Regression:** this method is valued for its interpretability. Gray et al. (2013) showed its effectiveness in feature importance evaluation, aiding in understanding disease indicators. The model's coefficients highlight significant predictors of conditions like dementia.
- **Support Vector Machines (SVM)**: Cuingnet et al. (2011) utilized SVM to classify Alzheimer's from MRI data, showcasing decent performance with properly selected features.
- **Decision Trees and Random Forests:** These methods are favored for their interpretability and ease of handling feature-rich data. Gray et al. (2013) demonstrated their effectiveness in feature importance evaluation, which aids in understanding disease indicators.

2. Deep Learning Approaches:

Convolutional Neural Networks (CNNs): Leveraged extensively due to their capability in feature extraction directly from images, CNNs reduce the need for manual feature engineering. Sarraf and Tofighi (2016) highlighted their efficiency in classifying Alzheimer's with higher accuracy rates. The challenge with CNNs lies in the need for extensive data to prevent overfitting and computational resources.

Autoencoders: Used primarily for unsupervised learning, autoencoders like those explored by Liu et al. (2018) help in reducing dimensionality and extracting useful features from MRI images. Their application is limited by the necessity for large datasets to train effectively.

3. Hybrid Models:

Combining CNNs and SVMs: This approach utilizes CNNs for robust feature extraction followed by SVMs for classification, providing a balance of deep learning and traditional machine learning benefits. Suk et al. (2014) showed that this combination could yield better diagnostic accuracy than using either method alone.

III. Dataset and features

1. Dataset Description

We will be using the longitudinal MRI data from the OASIS-2 dataset. This dataset consists of MRI and associated clinical data for 150 subjects aged 60 to 96. Each subject was scanned at least once, with follow-up scans occurring at intervals of one year or more. All subjects are right-handed. Within the dataset, 72 subjects were classified as 'Nondemented' throughout the study, while 64 subjects were classified as 'Demented' at their initial visits and remained so throughout. Additionally, 14 subjects were initially classified as 'Nondemented' but later transitioned to 'Demented', these subjects fall under the 'Converted' category.

Column Descriptions:

- EDUC: Years of Education
- SES: Socioeconomic Status
- MMSE: Mini Mental State Examination
- CDR: Clinical Dementia Rating
- eTIV: Estimated Total Intracranial Volume
- nWBV: Normalize Whole Brain Volume
- ASF: Atlas Scaling Factor
- Mini–Mental State Examination (MMSE)

Mini–Mental State Examination (MMSE): A 30-point questionnaire used to measure cognitive impairment and screen for dementia.

Cognitive Impairment: Difficulties in memory, learning, concentration, or decision-making that affect daily life.

Clinical Dementia Rating (CDR): A 5-point scale assessing cognitive and functional performance in Alzheimer's disease and related dementias.

Estimated Total Intracranial Volume (eTIV): A crucial covariate in brain volumetric analyses, representing maximum pre-morbid brain volume.

Atlas Scaling Factor (ASF): A normalization technique for head size in brain volume measurements, enhancing reliability across different age groups and cognitive statuses.

Visualizations:

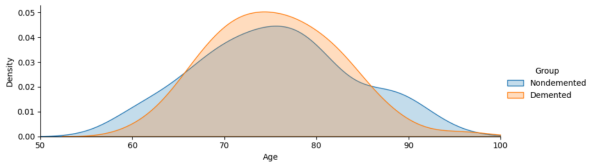

Figure 2. Age Distribution within each class

The Demented patient group exhibits a greater prevalence of individuals aged 70-80 compared to the nondemented patient group. We hypothesize that individuals afflicted with this condition may have a reduced life expectancy, resulting in fewer individuals aged 90 years and older.

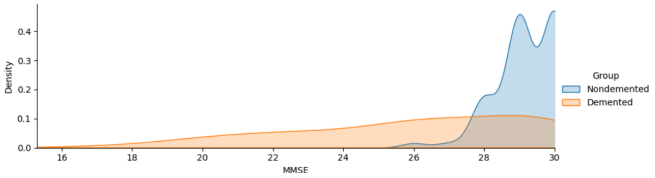

Figure 3. MMSE Distribution within each class

We observed a notable distinction in Mini–Mental State Examination (MMSE) scores between the Demented and Nondemented groups. The chart illustrates that individuals in the Nondemented group consistently obtained significantly higher MMSE scores compared to those in the Demented group. This discrepancy suggests a potential correlation between cognitive impairment, as assessed by the MMSE, and the presence of dementia.

2. Preprocessing Steps

To prepare the dataset for analysis, several preprocessing steps were performed to ensure data integrity and relevance:

Removing Duplicates: Duplicate entries were removed to prevent redundant data from skewing the results.

Handling Null Values: The "SES" (Socio-Economic Status) column's missing values were imputed using the median value of the column to mitigate the impact of outliers. The "MMSE" (Mini-Mental State Examination) column's missing values were imputed using the mean value to maintain the overall distribution of scores.

Group Label Adjustment: The "Converted" label in the "Group" column was replaced with "Demented" to unify all dementia-related labels under a single category.

Visit Selection: Data were filtered to include only the first visit for each patient, as initial diagnoses were made at this time and patients' statuses did not change in subsequent visits.

Label Encoding: The "Group" column was encoded with 'Demented' as 1 and 'Nondemented' as 0 to facilitate binary classification. The "M/F" column was encoded with 'M' (male) as 1 and 'F' (female) as 0 to convert categorical gender data into numerical format.

Dropping Irrelevant Columns: Columns not relevant to the analysis, such as 'Subject ID', 'MRI ID', 'Visit', 'Hand', and 'MR Delay', were removed to streamline the dataset.

Outlier Removal: Outliers were identified and removed based on the interquartile range (IQR) method to enhance data quality and analysis reliability.

Feature and Target Separation: The dataset was divided into features (predictors) and the target variable (outcome). The target variable was the "Group" column, indicating the presence or absence of dementia.

Standardization: Prior to analysis, feature data from the dataset is standardized using the StandardScaler from the sklearn.preprocessing library. This process involves normalizing the dataset features to ensure each contributes equally, thus avoiding bias due to the scale of features.

IV. Methods

A- Logistic Regression

1. Learning Algorithms: Logistic Regression

Logistic Regression is a supervised learning algorithm primarily used for binary classification tasks. The objective is to model the probability that a given input point belongs to a particular class, based on one or more independent

variables. In this study, Logistic Regression was employed to classify subjects as either 'Demented' or 'Nondemented' based on their MRI and collected clinical data.

2. Methodological Steps

Model Training and Evaluation:

Grid Search with Cross-Validation: The Logistic Regression model was optimized by performing a grid search over hyperparameters such as

a. Regularization Strength (C):
   Values: [0.001, 0.01, 0.1, 1, 10, 100, 1000]
   Purpose: Controls the inverse of regularization strength.
b. Penalty: Types: ['l1', 'l2'],
   Purpose: Specifies the norm used in the penalization. L1 (Lasso) penalty encourages sparsity in the coefficients, while L2 (Ridge) penalty adds the squared magnitude of coefficients.

3. Mathematical Notation

The decision function for Logistic Regression is defined as:

$$\widehat{y} = \sigma(W^T X + b)$$

Where σ is the sigmoid function: $\sigma(z) = \frac{1}{1+e^{-z}}$, W is the weight vector, X is the input feature vector, and b is the bias term.

Regularization: The objective function to be minimized, incorporating regularization, is:

$$L(w,b) = -\frac{1}{m}\sum_{i=1}^{m}[y_i log(\widehat{y_i}) + (1 - y_i)log(1 - \widehat{y_i})] + \lambda R(w)$$

Where m is the number of training samples, λ is the regularization parameter, inversely related to C and R(w) is the regularization term (either L1 or L2 norm).

B - Random Forest

1. Learning Algorithms: Random Forest Classifier:

Random Forest is used as the primary learning algorithm. This ensemble method constructs multiple decision trees during training and outputs the class that is the mode of the classes predicted by the individual trees. It effectively reduces the overfitting problem common with decision trees.

2. Methodological Steps

Model Training and Evaluation:

Hyperparameters of the Random Forest model are tuned using GridSearchCV.

a. n_estimators: Chosen from [50, 100, 150] based on model complexity and overfitting considerations.
b. max_depth: Tested with values [None, 10, 20] to control the depth of trees and prevent overfitting.
c. min_samples_split: Evaluated at [2, 5, 10] to determine the minimum number of samples required to split an internal node.

These parameters were optimized using GridSearchCV, which iteratively tested combinations of hyperparameters across a 5-fold cross-validation scheme to find the best mix that maximizes the model's accuracy.

3. Mathematical Notation

Loss Function: Internally, the Random Forest algorithm utilizes the Gini impurity as a criterion to guide the splitting of nodes in the decision trees. The Gini impurity for a node is defined mathematically as: $G = 1 - \sum_{i=0}^{c} p_i^2$

C- Support Vector Machine

1. Learning Algorithms: Support Vector Machine (SVM)

SVM is a supervised learning model used for classification and regression analysis. The primary objective of SVM is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. In the context of this study, SVM was employed to classify subjects as either 'Demented' or 'Nondemented' based on their MRI and clinical data.

2. Methodological Steps:

Model Training and Evaluation:

Grid Search with Cross-Validation: The SVM model was optimized by performing a grid search with 5-fold cross-validation, optimizing for accuracy, over hyperparameters such as kernel type, regularization parameter C, and kernel coefficient γ.

a. Kernel Type: ['linear', 'rbf', 'poly']
b. Regularization Parameter C: [0.1, 1, 10, 100].
c. Kernel Coefficient γ: ['scale', 'auto'].

Where:

Linear Kernel: $K(x_i, x_j) = x_i . X_j$.

Radial Basis Function (RBF) Kernel:

$K(x_i, x_j) = exp(-\gamma\left\|x_i - x_j\right\|^2)$.

Polynomial Kernel: $K(x_i, x_j) = (x_i . x_j + c)^d$.

3. Mathematical Notation

The decision function for SVM is defined as:

$$f(x) = sign(\sum_{n}^{i=1} \alpha_i y_i K(x_i, x) + b$$

Where : $K(x_i, x_j)$ is the kernel function, α_i are the support vectors and b are the bias terms.

D-Decision Tree

1. Learning Algorithms: Decision Tree:

Decision Trees are a supervised learning model used for classification and regression tasks. They work by recursively splitting the data into subsets based on the value of input features, creating a tree-like structure of decision rules.

2. Methodological Steps:

Model Training and Evaluation:

Grid Search with Cross-Validation**:** To optimize the performance of the Decision Tree model, a hyperparameter tuning process was conducted using Grid Search with cross-validation. The grid search was performed over the following ranges:

a. criterion: ['gini', 'entropy']
b. max_depth: [2, 3, 5, 10, None]

The parameters tuned were:

Criterion: The function to measure the quality of a split. Options include 'gini' for Gini impurity and 'entropy' for information gain. And max Depth: The maximum depth of the tree. Limiting the depth helps prevent overfitting.

3. Mathematical Notation

Gini Impurity: $G = 1 - \sum_{i=0}^{c} p_i^2$

Where pi is the probability of an element being classified as class *i*.

Block Diagram

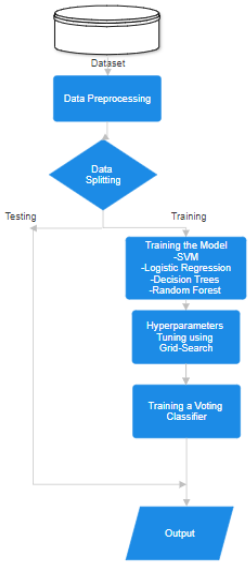

Figure 4. Block Diagram

V. Experiments and Results

1. *Primary Evaluation Metrics*

To assess the performance of the machine learning models in detecting early-stage Alzheimer's disease, we employed the following evaluation metrics:

a. Accuracy: Measured as the ratio of correctly predicted observations to the total observations.

$$Accuracy = \frac{Number\ of\ correct\ predications}{Total\ number\ of\ predications}$$

b. Precision: The ratio of correctly predicted positive observations to the total predicted positives.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positvies}$$

c. Recall (Sensitivity): The ratio of correctly predicted positive observations to all observations in the actual class.

d. F1-Score: The weighted average of Precision and Recall.

e. Confusion Matrix: A table used to describe the performance of a classification model by showing the true positive, true negative, false positive, and false negative counts. This matrix is essential for understanding the distribution of predictions and evaluating the model's effectiveness in distinguishing between demented and non-demented subjects.

2. *Results*

A- Logistic Regression

1. Hyperparameters Selection

For the model, The best hyperparameters identified for the logistic regression model were: {'C': 1.0, 'penalty': 'l2'}

2. *Primary Metrics:* The primary metrics used for model evaluation were:

1. **Accuracy:** Best Cross-Validation Score: 0.92.
   Test Accuracy: 0.88.
2. **Confusion Matrix:** The confusion matrix provides a detailed breakdown of the model's performance across different classes. As shown in the figure below:

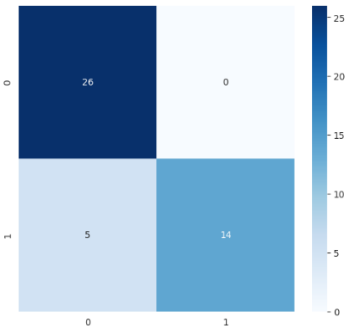

Figure 5. Confusion Matrix of Logistic Regression

B - Random Forest

1. Hyperparameters Selection

the best hyperparameters for the Random Forest model were {'max_depth': 10, 'min_samples_split': 5, 'n_estimators': 50}

2. Primary Metrics: The primary metrics used for model evaluation were:

1. **Accuracy**: The Random Forest model achieved an overall Best Cross-Validation Score: 0.92.
   Test Accuracy: 0.88.
2. **Confusion Matrix** The confusion matrix provides a visual summary of the model's classification accuracy for two classes, 0 and 1. As shown in the figure below:

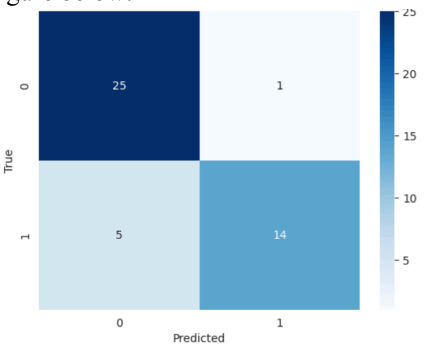

Figure 6. Confusion Matrix of Random Forest

C- SVM

1. Hyperparameters Selection:

In this study, we used an SVM to classify subjects. The SVM model's performance is highly dependent on its hyperparameters, so we employed a grid search with cross-validation to find the optimal set. The best hyperparameters identified for the SVM model were: {'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}

2. Primary Metrics: The primary metrics used for model evaluation were:

1. **Accuracy**: Best Cross-Validation Score: 0.92 (accuracy) and Test Accuracy: 0.88
2. **Confusion Matrix** The confusion matrix provides a visual summary of the model's classification accuracy for two classes, 0 and 1. As shown in the figure below:

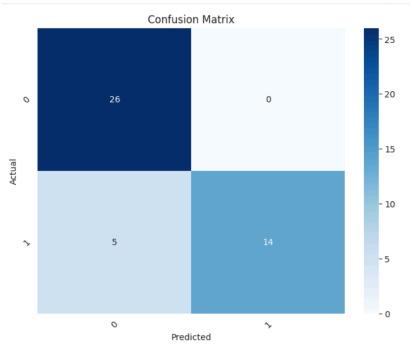

Figure 7. Confusion Matrix of SVM

D-Decision Tree

1. Hyperparameters Selection:

To optimize the performance of the Decision Tree model, a hyperparameter tuning process was conducted using Grid Search with cross-validation. The hyperparameter tuning process identified the best parameters as criterion='gini' and max_depth=2.

2. Primary Metrics

The results of the Decision Tree classifier for this study are

1. **Accuracy**: Best Cross-Validation Score: 0.924 and Test Accuracy: 0.889

2. **Confusion Matrix** The confusion matrix provides a visual summary of the model's classification accuracy for two classes, 0 and 1. As shown in the figure below:

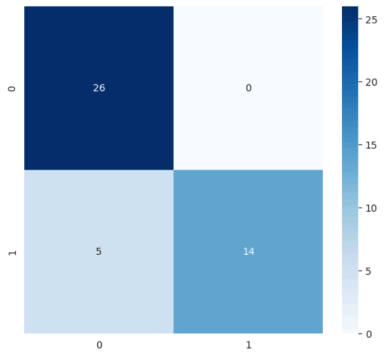

Figure 8. Confusion Matrix of Decision Tree

VI. Conclusion

This study explored machine learning applications for the early detection of Alzheimer's disease using MRI data from the OASIS-2 dataset, focusing on comparing the effectiveness of various algorithms such as logistic regression, random forests, support vector machines (SVM), and decision trees. Among these, the Support Vector Machine (SVM) model, particularly with the radial basis function kernel, was chosen as one of the best performing models not only for its high accuracy but also for its robustness to outliers and ability to handle high-dimensional data effectively. It demonstrated a strong capability to distinguish between demented and nondemented subjects, providing a promising direction for future automated diagnostic tools. Additionally, a Voting Classifier, which combines the predictions of multiple models, was also evaluated. This ensemble method further enhanced the robustness and reliability of the predictions by leveraging the strengths of each individual model. The Voting Classifier achieved comparable high accuracy, highlighting the potential of ensemble learning to improve diagnostic performance. These findings underscore the potential of machine learning to enhance the objectivity, consistency, and speed of Alzheimer's diagnosis, transitioning away from traditional methods that rely heavily on subjective clinical assessments.

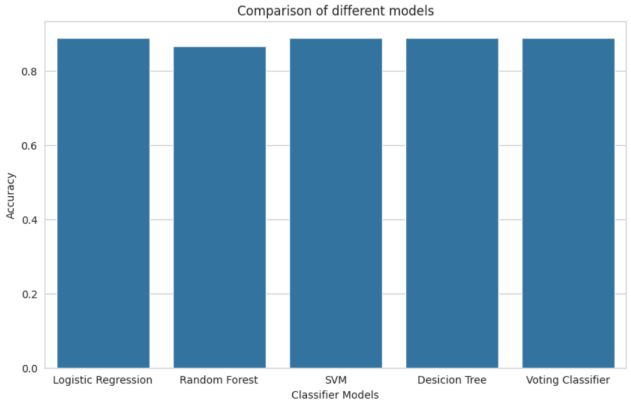

Figure 9. Comparison Between Classifiers

VII. Future Work

Future research in Alzheimer's disease detection using machine learning focus on several pivotal areas:

Dataset Expansion: Enlarge and diversify the dataset through international collaborations to improve model robustness and applicability across different demographics and disease stages.

Advanced Deep Learning: Incorporate advanced architectures like 3D convolutional neural networks and explore transfer learning from extensive medical imaging datasets to enhance diagnostic accuracy.

Multi-Modal Data Integration: Combine MRI data with genetic, cognitive, and biomarker information for a more comprehensive diagnostic approach.

Model Interpretability: Develop interpretable machine learning models that clinicians can trust, providing clear insights into their diagnostic reasoning.

VIII. Limitations

While the study demonstrates promising results in the early detection of Alzheimer's disease using machine learning, several limitations must be acknowledged:

Small Dataset: The dataset consists of only 373 samples initially, reduced to 150 after preprocessing. This small sample size can impact the generalizability and robustness of the model, necessitating further validation on larger, more diverse datasets.

Feature Extraction from MRI Data: Features were extracted from MRI data and stored in CSV format, which may result in the loss of valuable spatial and structural information. Directly working with MRI data using advanced techniques like convolutional neural networks (CNNs) could potentially yield more accurate and robust models by preserving the rich information in the raw images.

VIII. Contributions

In the completion of this project, each team member played an integral and equal role across all stages, including data collection, model implementation, and the development of the research paper. This balanced and collective effort was pivotal in achieving the project's objectives efficiently and effectively.

Team Members	Contribution
Habiba Mohsen	Supported Vector Machine Model.
Hana Hesham	Random Forest Model.
Malak Nasser	Logistic Regression Model.
Hazem Raafat	Decision Tree Model.

a.

IX. References

[1] N. Amoroso et al., "Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge," J. Neurosci. Methods, vol. 302, pp. 3-9, 2018. Available: https://pubmed.ncbi.nlm.nih.gov/29287745

[2] R. Cuingnet et al., "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," NeuroImage, vol. 56, no. 2, pp. 766-781, 2011. Available: https://doi.org/10.1016/j.neuroimage.2010.06.013

[3] Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D; Alzheimer's Disease Neuroimaging Initiative. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. Neuroimage. 2013 Jan 15;65:167-75. Available :http://doi.org/10.1016/j.neuroimage.2012.09.065

[4] Ehsan Hosseini As, Mohammed Ghazal, Ali Mahmoud, Ali Aslantas, Ahmed Shalaby, Manual Casanova, Gregory Barnes, Georgy Gimel'farb, Robert Keynton, Ayman El Baz. Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. Front. Biosci. (Landmark Ed) 2018, 23(3), 584–596. Available: https://doi.org/10.2741/4606

[5] S. Liu et al., "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," IEEE Trans. Biomed. Eng., vol. 62, no. 4, pp. 1132-1140, 2014. Available: https://doi.org/10.1109/TBME.2014.2372011

[6] S. Sarraf and G. Tofighi, "Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks," arXiv preprint arXiv:1603.08631, 2016. Available: https://arxiv.org/abs/1603.08631