

Early Alzheimer's disease diagnosis using an XG-Boost model applied to MRI images

Khoi Nguyen^{1,2,#}, My Nguyen^{3,2,#}, Khiet Dang^{1,2}, Bao Pham^{1,2}, Vy Huynh^{2,3}, Toi Vo^{1,2}, Lua Ngo^{1,2}, Huong Ha^{1,2,*}



Use your smartphone to scan this QR code and download this article

[#]These authors contributed equally to this manuscript

¹School of Biomedical Engineering, International University, Viet Nam

²Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Viet Nam

³Faculty of Biology – Biotechnology, University of Science, Viet Nam

Correspondence

Huong Ha, School of Biomedical Engineering, International University, Viet Nam

Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Viet Nam

Email: htthuong@hcmiu.edu.vn

History

- Received: Jun 20, 2023
- Accepted: Sep 18, 2023
- Published Online: Sep 30, 2023

DOI : 10.15419/bmrat.v10i9.832



Copyright

© Biomedpress. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



ABSTRACT

Introduction: Early Alzheimer's disease (AD) diagnosis is critical to improving the success of new treatments in clinical trials, especially at the early mild cognitive impairment (EMCI) stage. This study aimed to tackle this problem by developing an accurate classification model for early AD detection at the EMCI stage based on magnetic resonance imaging (MRI). **Methods:** This study developed the proposed classification model through a machine-learning pipeline with three main steps. First, features were extracted from MRI images using FreeSurfer. Second, the extracted features were filtered using principal component analysis (PCA), backward elimination (BE), and extreme gradient (XG)-Boost importance (XGBI), the efficiency of which was evaluated. Finally, the selected features were combined with cognitive scores (Mini Mental State Examination [MMSE] and Clinical Dementia Rating [CDR]) to create an XG-Boost three-class classifier: AD vs. EMCI vs. cognitively normal (CN). **Results:** The MMSE and CDR had the highest importance weights, followed by the thickness of the left superior temporal sulcus and banks of the superior temporal lobe. Without feature selection, the model had the lowest accuracy of 69.0%. After feature selection and the addition of cognitive scores, the accuracy of the PCA, BE, and XGBI approaches improved to 74.0%, 90.9%, and 91.5%, respectively. The BE with tuning parameters model was chosen as the final model since it had the highest accuracy of 92.0%. The area under the receiver operating characteristic curve for the CN, AD, and EMCI classes were 0.98, 0.94, and 0.88, respectively. **Conclusion:** Our proposed model shows promise in early AD diagnosis and can be fine-tuned in the future through testing on a multi-dataset.

Key words: Alzheimer's disease, Early mild cognitive impairment, early diagnosis, three-class classification, XG- Boost

INTRODUCTION

Alzheimer's disease (AD) is the most common neurodegenerative disorder that greatly reduces patients' quality of life and makes them utterly dependent on their caregivers^{1,2}. Prolonged medical treatment and care exert a substantial economic strain on patients and their families, potentially costing >1.1 trillion US dollars worldwide¹. Unfortunately, once cognitive symptoms manifest, current medications cannot reverse disease progression due to the continued loss of neurons without replacement by cell division^{3,4}. Therefore, identifying patients at the early mild cognitive impairment (EMCI) stage is critical to improving the success of new treatments or interventions in clinical trials.

Several breakthrough approaches have attempted to predict AD at its preclinical stage, which could allow the application of medications to halt AD development from its onset^{3,5-8}. About 80% of patients diagnosed with mild cognitive impairment (MCI) convert to AD within six years⁹. Recent studies have fo-

cused on this transitional phase to detect the preclinical AD stage, particularly EMCI⁵. One promising approach to detect EMCI is identifying brain morphological changes through neuroimaging data, such as magnetic resonance imaging (MRI).

Early AD detection using brain MRI data remains clinically challenging since the subtle changes during its transitional period cannot be assessed manually³. Automatic computation and artificial intelligence (AI) approaches such as deep learning (DL) or machine learning (ML) are required to identify brain structural features at the EMCI stage. Of numerous AI-assisted methods, DL has been broadly used because of its high performance, especially the convolutional neural network (CNN)^{5,10}. Kang *et al.* combined a 2D CNN with transfer learning to identify EMCI by processing a multi-modal dataset (MRI and diffusion tensor imaging data), achieving the highest accuracy of 94.2% for cognitively normal (CN) vs. EMCI patients⁵. In addition, Kolahkaj *et al.* built a DL architecture based on the BrainNet

Cite this article : Nguyen K, Nguyen M, Dang K, Pham B, Huynh V, Vo T, Ngo L, Ha H. **Early Alzheimer's disease diagnosis using an XG-Boost model applied to MRI images.** *Biomed. Res. Ther.* 2023; 10(9):5896-5911.

CNN model to detect EMCI, achieving high accuracies for binary classification: 0.96, 0.98, and 0.95 for NC/EMCI, NC/MCI, and EMCI/MCI, respectively¹¹. Despite its significant results, DL has several limitations that could hinder clinical applications. Firstly, DL models are prone to encounter overfitting due to the many parameters considered¹². Secondly, analysts cannot provide a plausible explanation for the algorithm's performance, which is called a black box. Therefore, to build an understandable prediction model, making the shift to ML for early AD detection is beneficial for neurologists and doctors.

While most ML studies have focused on binary classification, some have focused on multi-class classification. However, there is a growing need for a multi-class algorithm that can effectively distinguish the prodromal stage (EMCI) from the array of other stages (late MCI [LMCI], AD, and CN), enabling an early AD diagnosis. Moreover, it is important to note that existing multi-class ML models have low accuracies. In 2022, Techa *et al.* showed that a new model based on three CNN architectures (DenseNet196, VGG16, and ResNet50) achieved 89% accuracy in discriminating normal, very mild dementia, mild dementia, moderate dementia, and AD¹³. Alorf *et al.* implemented a Brain Connectivity-Based Convolutional Network in 2022, which provided 84.03% accuracy for six-class classification (AD, LMCI, MCI, EMCI, subjective memory complaints, and CN)¹⁴. Another major difficulty when identifying the initial AD stages is the subtle structural change in subjects with EMCI. EMCI is elusive and cannot be recognized by the diagnostic criteria for AD¹⁵. Furthermore, EMCI and MCI are highly heterogeneous since they can be easily mistaken for multiple pathological conditions, especially other neurodegenerative diseases^{16,17}. Therefore, EMCI classification requires further evaluation and approaches to optimize its efficiency.

One potential ML model to address the early AD detection challenge is extreme gradient boosting (XG-Boost). XG-Boost is a scalable tree-based ensemble learning implemented from the gradient boosting system. It introduces errors from the previous weak learner to the latter learner, improving its learning accuracy¹⁸. Since its results depend on many decision trees, XG-Boost shows high compatibility, competitive execution speed, and accuracy when applied to large data sets, making it suitable for clinical application¹⁹. While few studies have used XG-Boost for AD diagnosis, the preliminary results are promising. Ong *et al.* proposed an XG-Boost model to classify AD

and CN subjects using the FreeSurfer library to extract insight features from MRI, achieving an area under the receiver operative characteristic (ROC) curve (AUC) of 91%²⁰. Tuan *et al.* presented an XG-Boost model to classify AD and normal subjects based on the tissues segmented by a CNN and Gaussian mixture model²¹. Their highest accuracy was 89% when combined with a support vector machine (SVM) and CNN²¹. However, both models had several limitations, such as high computation cost and susceptibility to sample size and complexity. They also did not attempt to classify three classes. Therefore, future improvement is required to enhance the models' accuracy and validity.

This study used XG-Boost for three-class classification, primarily focusing on distinguishing CN, EMCI, and AD. It also evaluated and optimized three feature selection methods—backward elimination, XG-Boost importance (XGBI), and principal component analysis (PCA)—to identify the most suitable method for the XG-Boost model. When combined with the Mini Mental State Examination (MMSE) and Clinical Dementia Rating (CDR) scores, our model achieved the highest accuracy of 92% for distinguishing AD, EMCI, and CN. Only three features overlapped between the BE and XGBI feature selection methods: MMSE, CDR, and left hippocampus volume. While these results showed that the model still depends on the cognitive symptoms of AD rather than its brain structural changes, our model has great potential as an assistive tool for AD diagnosis with high performance, especially when considering its multi-class classification.

METHODS

Participants

This study obtained its data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>)²². The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD. Its primary goal has been to test whether serial MRI, positron emission tomography, biological markers, and clinical and neuropsychological assessments can be combined to measure MCI and EMCI progression²².

The data comprised 663 subjects who were equally grouped into three classes: CN, EMCI, and AD. Their demographic information is summarized in **Table 1**.

Structural MRI data

The structural MRI scans used in this study were the T1-weighted magnetization prepared-rapid gradient

Table 1: Demographic information 663 recruited subjects from ADNI

	CN	EMCI (n = 221)	AD (n = 221)	p
Age	75.28 ± 5.76	71.45 ± 7.23*	75.4 ± 7.702	< 0.0001
Sex (M/F)	120/101	118/103	120/101	0.9760
MMSE Score	29.06 ± 1.1	28.12 ± 1.66*	22.8 ± 2.63*	< 0.0001
CDR Score	0.03 ± 0.11	0.47 ± 0.16*	0.81 ± 0.32*	< 0.0001
Education (Years)	16.18 ± 3.88	16.09 ± 2.65	14.65 ± 4.35*	< 0.0001
ApoE4 (+/-)	157/64	82/139	58/163	< 0.0001

Chi-square test was conducted for gender and genotype. ANOVA one-way was conducted for age, MMSE, CDR, and educational years. * p < 0.0001, compared to CN group

Abbreviations: ns: not significant, compared to CN group (p = 0.1234) (Bonferroni 's comparison). Data is illustrated as mean ± standard deviation or number/number. **CN:** Normal Cognitive; **EMCI:** Early Mild Cognitive Impairment; **AD:** Alzheimer's disease; **MMSE:** Mini-Mental State Examination; **CDR:** Clinical Dementia Rating; (+) positive; (-) negative where available.

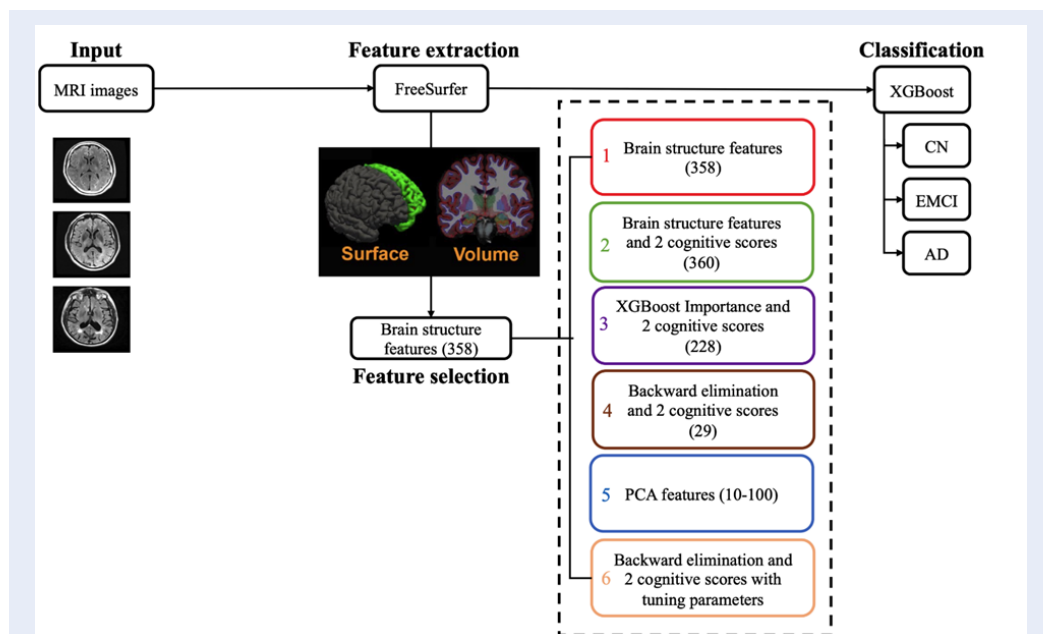


Figure 1: A study framework of AD detection, which includes three main steps. T1-weighted MRI data were collected from the ANDI database (step 1) and preprocessed through FreeSurfer software to obtain brain structure features. Sequentially they were combined with two cognitive scores and tuned by three selection methods to construct six approaches for input features (step 2). Finally, generated inputs passed through the XG-Boost model to create the decision tree for AD status (ternary classification), which are CN, EMCI, and AD in step 3. The outcome also showed the accuracies of the respective input. **Abbreviations:** **CN:** Normal Cognitive; **EMCI:** Early Mild Cognitive Impairment; **AD:** Alzheimer's disease; **PCA:** Principle Component Analysis; **XG-Boost:** Extreme Gradient Boosting.

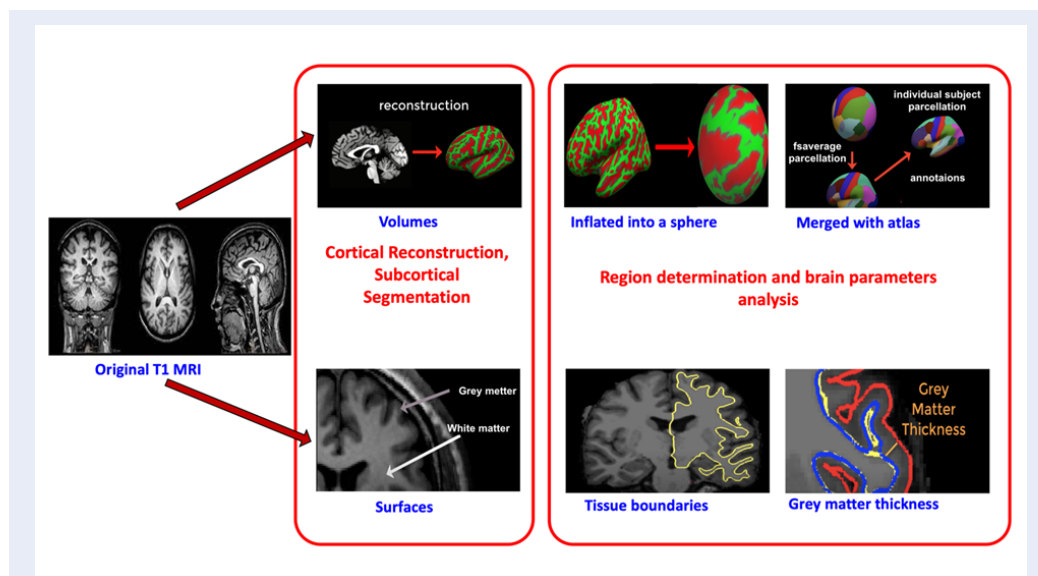


Figure 2: The process of FreeSurfer in features extraction. MRI Preprocess: Including image registration, skull stripping and intensity normalization. Cortical reconstruction and subcortical segmentation: (1) Convert a three-dimensional anatomical volume into a two-dimensional surface; (2) Segment gray matter and white matter to create the brain mask file for visualizing after. Region determination and brain parameters analysis: (1) Inflate the surfaces into a sphere and map cortical parcel actions back onto individual subjects using two atlases (Killiany and Destrieux atlas) (2) Establish the boundary between white matter and cortex and compute gray matter thickness.

Table 2: 358 features were extracted by Freesurfer from 663 ADNI subjects, particular dimension for each brain region

No.	Subject ID	Brain Segmentati-on Volume Without Ventricles	Left Entorhinal Cortex (temporal lobe)	White Surface Total Area in the left hemisphere	Banks of Superior Temporal Sulcus in the left hemisphere	...	Number of Defect Holes in right hemispherical Surface Prior to fixin
1	135_S_4598	1076438.0	285.0	84644.5	996.0	...	17.0
2	099_S_4480	945976.0	310.0	76032.8	744.0	...	33.0
3	099_S_2146	1138086.0	453.0	88770.5	1118.0	...	46.0
...
662	082_S_1079	1131880.0	446.0	94008.6	1244.0	...	73.0
663	130_S_5059	1160101.0	601.0	85947.9	862.0	...	49.0

* Where area in mm², volume in mm³

Table 3: The results of feature selection by Approach 3, Approach 4, and Approach 5

Method	Backward Elimination (Approach 3)	XGBoost Importance (Approach 4)	PCA (Approach 5)
Number of features after selection	29	228	71
Type of features	Brain features and cognitive scores	Brain features and cognitive scores	PCA features

Abbreviations: PCA: Principle Component Analysis; XG-Boost: Extreme Gradient Boosting.

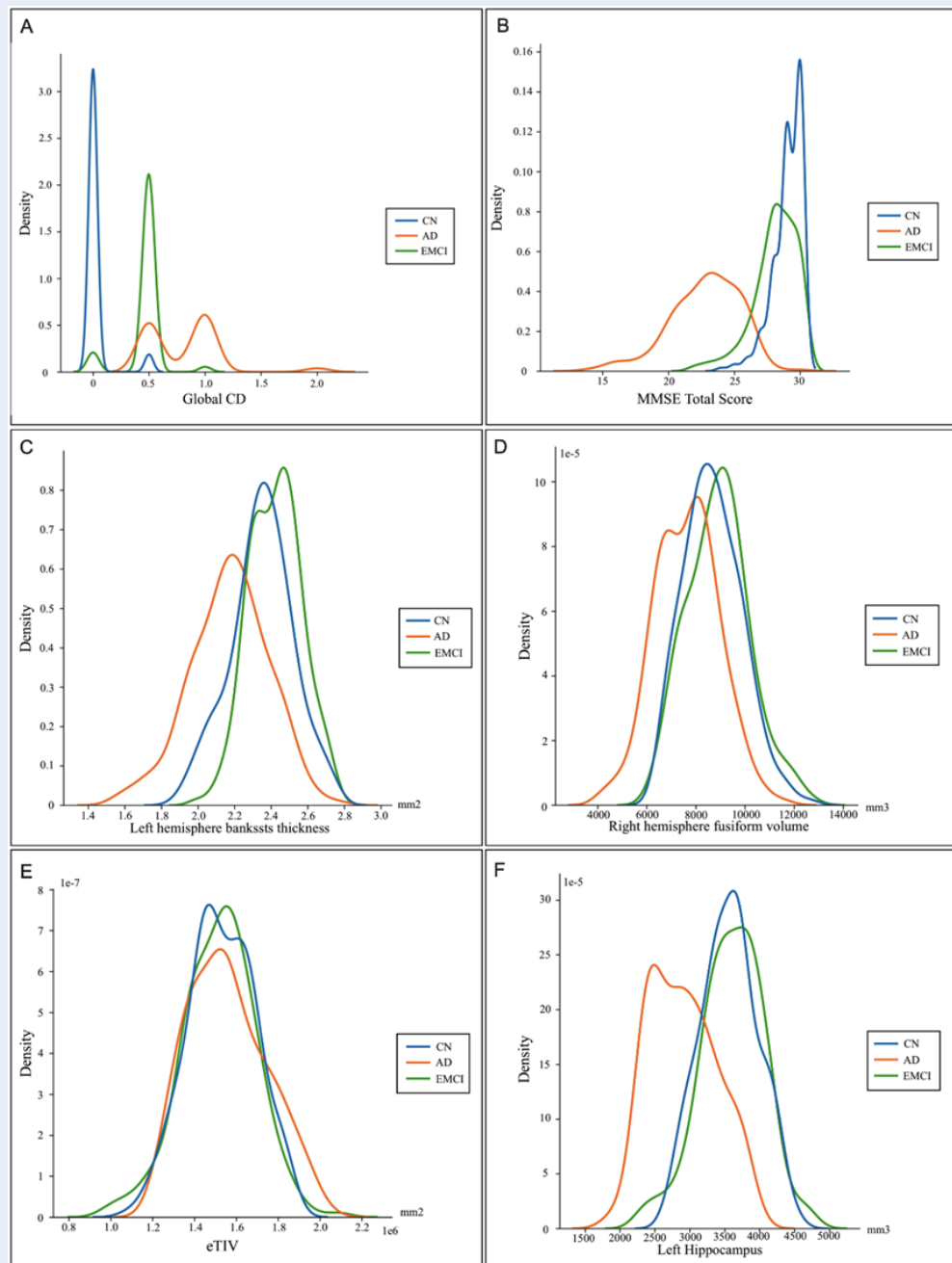


Figure 3: Density plots showing the distribution among three classes (AD, EMCI, CN) of two cognitive scores and several MRI features. (A) Global CDR Scores, (B) MMSE Scores, (C) Left hemisphere bankssts thickness, (D) Right hemisphere fusiform volume, (E) eTIV, (F) Left Hippocampus volume. Blue: AD, orange: EMCI, green: CN. Abbreviations: CN: Normal Cognitive; EMCI: Early Mild Cognitive Impairment; AD: Alzheimer's disease; PCA: Principle Component Analysis

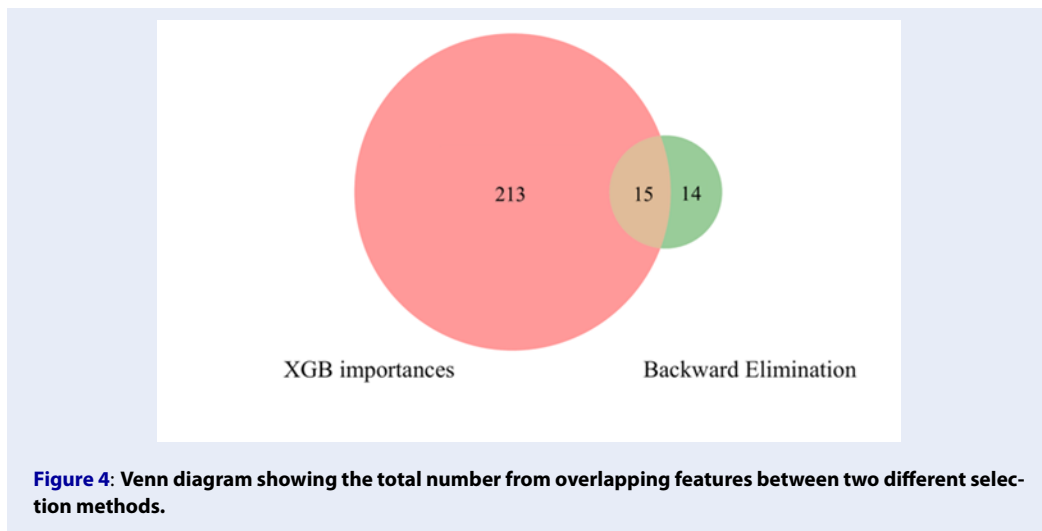


Figure 4: Venn diagram showing the total number from overlapping features between two different selection methods.

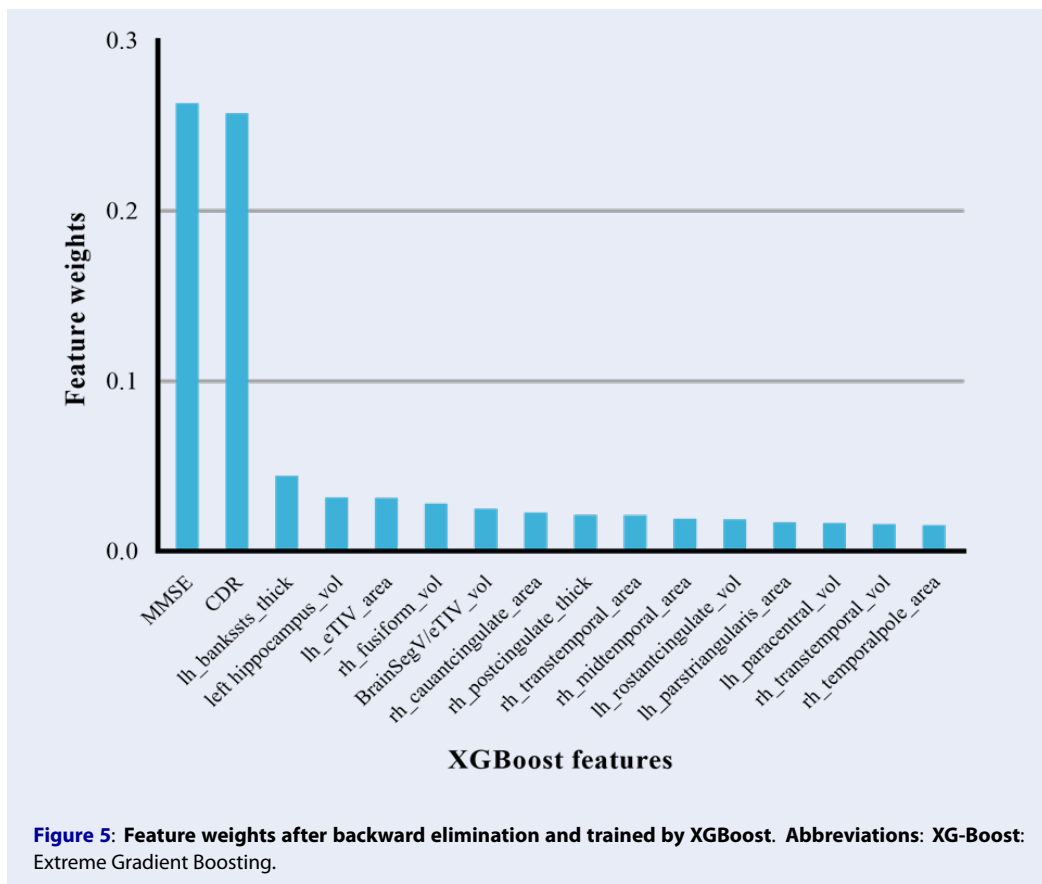


Figure 5: Feature weights after backward elimination and trained by XGBoost. Abbreviations: XG-Boost: Extreme Gradient Boosting.

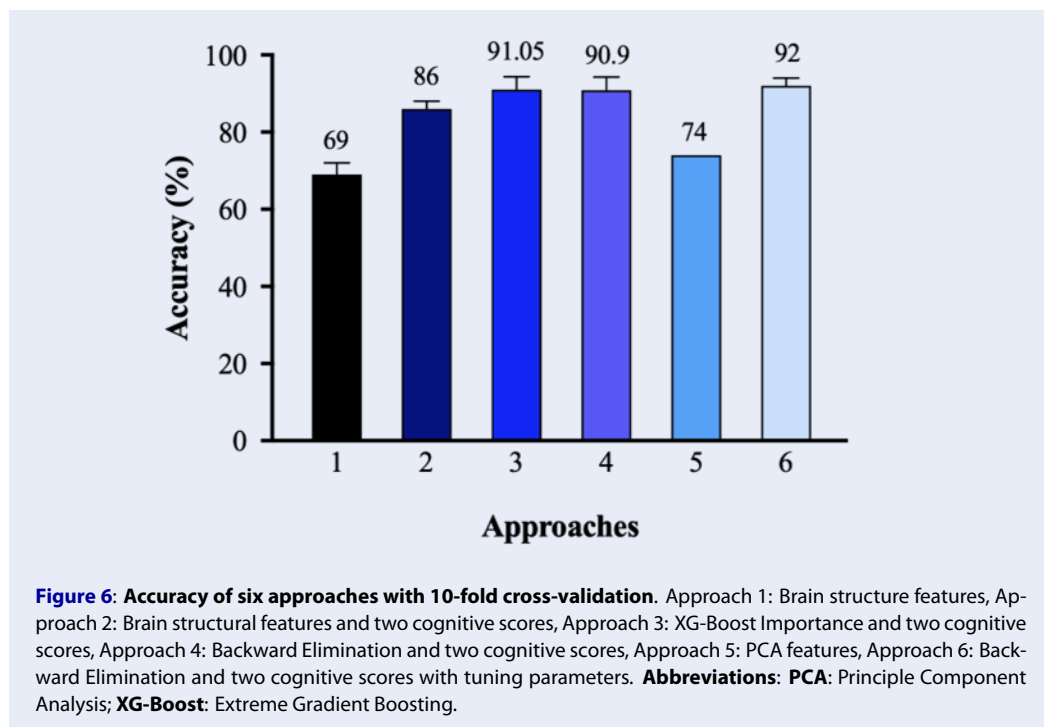


Table 4: The performance results of six approaches for three-class classification

Approach	Class	Accuracy	Precision	Recall	F1 score
1	CN	68.8 %	64 %	56 %	60 %
	EMCI		64 %	75 %	69 %
	AD		79 %	74 %	77 %
2	CN	86 %	80 %	97 %	88 %
	EMCI		97 %	71 %	82 %
	AD		83 %	98 %	90 %
3	CN	91.05 %	89 %	98 %	93 %
	EMCI		95 %	83 %	89 %
	AD		91 %	95 %	93 %
4	CN	90.9 %	91 %	98 %	95 %
	EMCI		92 %	79 %	85 %
	AD		90 %	96 %	93 %
5	CN	74 %	68 %	59 %	63 %
	EMCI		75 %	77 %	76 %
	AD		78 %	86 %	82 %
6	CN	92 %	88 %	97 %	93 %
	EMCI		91 %	85 %	88 %
	AD		96 %	94 %	95 %

The data included sMRI from 663 subjects, which were divided into 80 % for training and 20 % for testing.

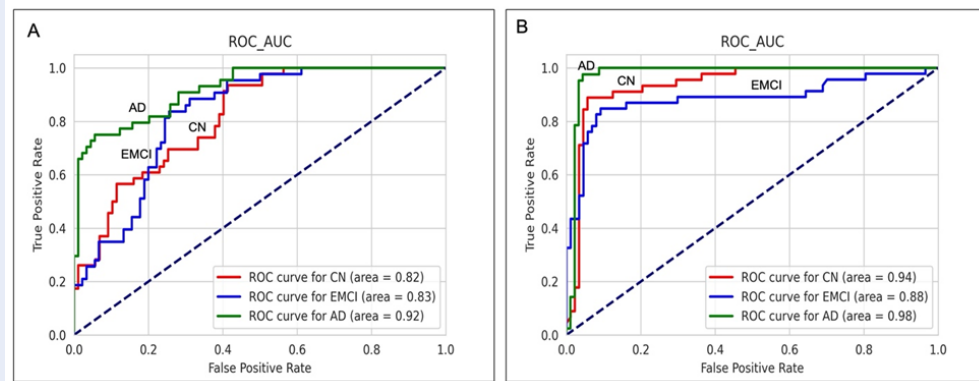


Figure 7: Receiver Operating Characteristic (ROC) curves of Approach 1 and Approach 6 for three classes classification. The green line corresponds to AD, the blue line represents for EMCI, and the red line shows CN. **Abbreviations:** CN: Normal Cognitive; EMCI: Early Mild Cognitive Impairment; AD: Alzheimer's disease; PCA: Principle Component Analysis

echo scans from ADNI 1 and ADNI GO/2. Various MRI scanner models were used for MRI acquisition; details of the acquisition protocol for the MRI data can be found on the ADNI website (<http://adni.loni.usc.edu>)²².

Study design

An overview of the study design is shown in **Figure 1**. Firstly, the MRI images were preprocessed with FreeSurfer to extract 358 features, including volumetric and thickness measurements. Three feature selection methods were used, and their efficiencies were compared. This step determined the optimal features from the 360 elements (FreeSurfer features, MMSE score, and CDR score). The data were divided into two sets with a ratio of 80% training to 20% testing using Python's Scikit-learn library. Finally, the proposed models were evaluated using the performance metrics of accuracy, precision, recall, F1-score, and ROC curves with AUCs to identify the most efficient classification algorithm.

Feature extraction

Six hundred sixty-three MRI images were reconstructed and segmented using FreeSurfer (version 5.3; <http://surfer.nmr.mgh.harvard.edu>). This open-source software measures and visualizes the human brain's functional, connective, and structural characteristics to extract brain structural features²³. This software's processing operations have two major stages (**Figure 2**).

Feature selection

Feature selection plays a significant role in ML and pattern recognition. Pearson's product-moment correlation coefficient (r) was first applied to remove all linearly related features with a $r > 0.9$. The reason for using this method is that several features extracted by FreeSurfer are sub-regions or different measurements of the same brain region. Therefore, including highly relevant features in a particular brain-diagnosed area is redundant from a neuroscience perspective. Moreover, highly correlated features may lead to overfitting, impacting model performance. Therefore, applying non-linear feature selection can improve model performance and reduce training time efficiently. The next step was performed with three feature selection methods to compare their efficiency.

PCA

PCA is a multivariate exploratory analysis approach that reduces the complexity of multidimensional data while preserving trends and key patterns^{24,25}. PCA was applied using Python's Scikit-learn library with different numbers of principal components (PCs; 1–321) to determine the optimal set of features for the classification model. Then, in each model, the PCs were incrementally included in 10 PC increments to observe changes in accuracy with Python's Matplotlib library.

BE

BE is a feature selection strategy that excludes characteristics strongly associated with the exposure without significantly influencing dependent variables or

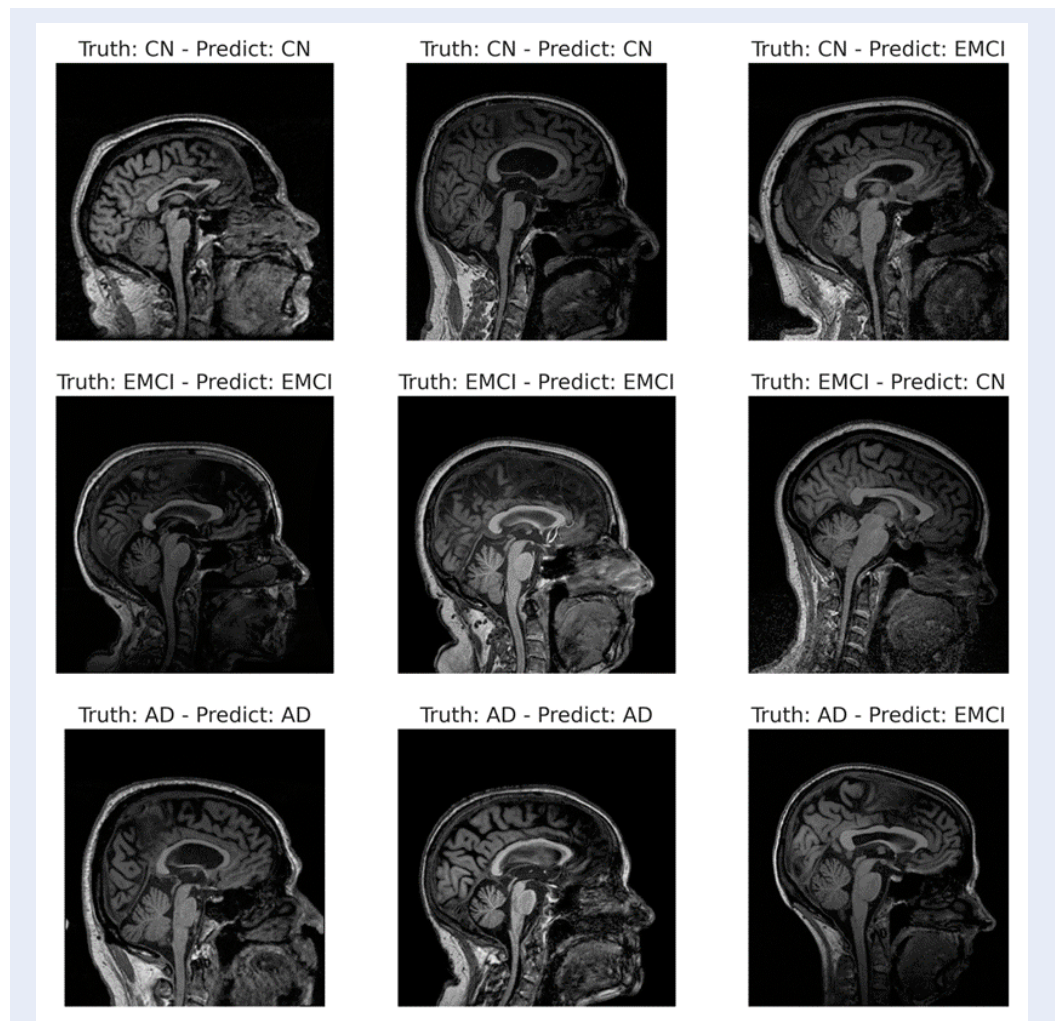


Figure 8: Visualization results for the ground truths and the corresponding predictions in three classes (CN, EMCI, AD). The first and second columns illustrate the correctly-predicted examples, while the last column shows the wrongly-predicted ones. **Abbreviations:** CN: Normal Cognitive; EMCI: Early Mild Cognitive Impairment; AD: Alzheimer's disease; PCA: Principle Component Analysis

predicted outputs^{26,27}. BE was applied in five main steps: (i) select a significance level (SL) that is suitable for the model (SL = 0.05), (ii) calculate original least squares with Python's Statsmodels library before determining the p -values of all features, (iii) compare the calculated p -value with the SL, (iv) remove features and predictors with a p -value greater than the SL, and (v) modify it to fit the model with the remaining variables.

XGBI

XG-Boost has the advantage of extracting importance scores for each feature in the predictive problem, enabling the determination of the highest importance score. The next step removes all unusable features

with zero importance coefficients depending on their ranking. This action is repeatedly performed until stable accuracy and non-zero importance coefficients are achieved.

Six features selection approaches

This study investigated six approaches for feature selection. Feature selection was not applied in the first and second approaches. The first approach used all 358 features extracted by Freesurfer to train the model. The second approach added the two cognitive scores to the 358 Freesurfer features. The third approach used XGBI to filter the Freesurfer features and included the two cognitive scores when training the model. The fourth and sixth approaches used BE

for feature selection and included the two cognitive scores; however, the sixth approach also applied parameter tuning. Finally, the fifth approach used PCA for feature selection.

Classification

XG-Boost is a scalable and efficient gradient-boosting framework used to combine a series of weak base learners (small decision trees) into a single powerful learner (a big tree)^{28,29}. The enhanced performance of XG-Boost has been shown in several major areas. Firstly, XG-Boost introduces a regularization component into the objective function, making the model less prone to overfitting. Secondly, it conducts a second-order rather than first-order Taylor expansion on the objective function, enabling it to specify the loss function more accurately. Thirdly, XG-Boost has a fast training speed due to data compression, multithreading, and GPU acceleration^{30,31}.

The objective function is defined as:

$$Obj^{(t)} = \sum_{i=1}^n L(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^n \Omega(f_i)$$

where $\hat{y}_i^{(t)}$ represents the prediction for the t^{th} round, f_i represents the structure of a decision tree, and $\Omega(f_i)$ represents the regularization component. $\Omega(f_i)$ is given by:

$$\Omega(f_i) = \gamma T + \frac{1}{2} \lambda \sum_{i=1}^T \omega_j^2$$

where λ represents the penalty coefficient and $\frac{1}{2} \lambda \sum_j^T = 1 \omega_j^2$ represents the L2 norm of leaf scores. After t iterations, the model's function is added to a new decision tree:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$$

and the objective function is updated:

$$Obj^{(t)} = \sum_{i=1}^n L(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \sum_{i=1}^n \Omega(f_t)$$

with the Taylor expansion specification:

$$Obj^{(t)} \simeq \sum_{i=1}^n \left[L(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

where g_i represents the first derivative and h_i represents the second derivative of the loss function. g_i and h_i are given by³¹:

$$g_i = \frac{\partial}{\partial \hat{y}_i^{(t-1)}} (L(y_i, \hat{y}_i^{(t-1)}))$$

$$h_i = \frac{\partial^2}{\partial \hat{y}_i^{(t-1)^2}} (L(y_i, \hat{y}_i^{(t-1)}))$$

This study applied the model from the open-source XG-Boost library. The algorithm also applies the softmax parameter and the cross-entropy function. After fitting the data, the Matplotlib library visualizes the fitting process and stops the process early to prevent overfitting.

Tenfold cross-validation³²

Grid Search cross-validation (GridSearchCV) is an object provided by Python's Scikit-learn library that generates a set of hyperparameters for tenfold cross-validation to achieve a maximally accurate model (estimator). GridSearch evaluates the grid of indicated parameters based on the estimator during the call to fit, including predicting, scoring, or transforming methods. Then, it returns the best-performing combination of hyperparameters with a maximum score (the scoring strategy of the basic estimator). Any other estimator can be applied to this object in this manner. Lastly, all modifiers and an estimator are assembled by a pipeline, resulting in a combined estimator that can implement several reductions afterward, such as tuning dimensions before fitting.

RESULTS

Feature extraction

After preprocessing and extraction, 358 features were exported. **Table 2** shows a portion of the extraction results. From the extraction results, we assessed the discriminative power of several features and two additional cognitive scores (CDR and MMSE) using the point distributions between three classes: AD, CN, and EMCI (**Figure 3**). We selected the top four weighted features according to XGBI and BE: left hemisphere banks of superior temporal sulcus thickness, right hemisphere fusiform volume, left hemisphere estimated total intracranial volume (eTIV), and left hippocampus volume. The two scores of the dementia tests (CDR and MMSE) showed a distinctive distribution in the density plots between the three classes (**Figure 3A, B**). In contrast, a significant overlap existed between classes in the eTIV distribution (**Figure 3E**). Nevertheless, the AD group separated relatively well from the CN and EMCI groups in the distributions of the other three Freesurfer features, especially the left hippocampus volume (**Figure 3F**). Overall, the density plots in **Figure 3** showed the great

potential of CDR and MMSE to enhance model accuracy when combined with the extracted features. These plots also highlight the challenges in distinguishing the CN and EMCI groups.

Feature selection

Several primary factors, such as redundancy (feature-feature) and relevance (feature-class), must be considered during feature selection³³. For redundancy minimization, this study used Pearson's product-moment correlation coefficient to measure the association between features and remove all linearly related features³⁴. This phase reduced the features from 360 to 324. Next, PCA, a popular feature selection method, was used to reduce dimensionality and identify highly effective and minimally redundant features. PCA created 33 feature sets; the first contained one feature, the second 11 features, and so on until the final set contained 321 features. Then, the performance of these feature sets was compared to investigate the efficiency of the PCA method.

Besides PCA, **Table 3** and **Figure 4** summarize the results with the other two feature selection methods (XGBoost and BE) to maximize relevance. The XGBoost library identified several features with unimportant values during the training process. Consequently, Approach 4 selected 228 features with non-zero importance coefficients to ensure that every feature benefits the training model. In addition, BE was applied for its speed and simplicity in removing irrelevant features with p -values > 0.05 . Interestingly, it only identified 29 features, of which 15 were shared with XGBoost, including the two cognitive scores and 13 brain structure features (**Figure 4**).

After selection, XGBoost continued to train on the features, resulting in the best performance with Approach 4 (see the **Classification results** section). **Figure 5** shows the weights of top-ranked features with Approach 4. The two cognitive scores were most influential in the prediction since their weights are approximately sixfold higher than those of the brain structure features (0.263 and 0.257, respectively). Moreover, the thickness of the left superior temporal sulcus was the most informative brain structure feature. The temporal lobe was also the most informative brain region because several features extracted from it had high weights, including the superior temporal sulcus, fusiform gyrus, transverse temporal gyrus, middle temporal gyrus, the temporal pole from the right hemisphere, and hippocampus from the left hemisphere. In conclusion, the temporal lobe shows the most significant changes in patients with AD.

Classification

The accuracies of all approaches and the details of each approach are summarized in **Figures 1 and 6**. The accuracies of these three-class classification models were assessed by the proportion of correct expected observations to all actual class observations with tenfold cross-validation. Approach 1, using 358 brain features, had the lowest accuracy ($69.00\% \pm 3.00\%$). The accuracy improved with Approach 2, which added the two cognitive scores to the feature set ($86.00\% \pm 2.00\%$). The accuracy improved again with Approach 3, which used XGBoost to select the features ($91.05\% \pm 3.34\%$). However, the accuracy decreased with Approaches 4 ($90.90\% \pm 3.35\%$) and 5 (74.00%). In Approach 5, the accuracies ranged from 63% to 74%, corresponding to 1 to 321 PCA features; the highest accuracy is shown in **Figure 6**. Approach 6, using BE for feature selection and tuning model parameters with grid search, achieved 92.00% accuracy. The performance of the six approaches is summarized in **Table 4**. In Approach 1, the AD class had the highest precision (79%), recall (74%), and F1 score (77%), while the CN class had the lowest precision, recall, and F1 score. In Approach 6, the AD class also achieved the highest precision (96%) and F1 score (95%). However, the CN class had the highest recall (97%) and a higher F1 score (93%) than the EMCI class (88%).

Figure 7 presents ROC curves showing the classification performance of Approaches 1 and 6. The ROC curve for Approach 1 showed that the model had poor performance in classifying CN and EMCI subjects (**Figure 7A**). The AUC of the EMCI class (0.83) was slightly higher than that of the CN class (0.82). However, Approach 1 performed well in identifying the AD class (AUC = 0.92). The ROC curve for Approach 6 showed that the final model classified the EMCI class less accurately than the CN and AD classes (AUC = 0.88; **Figure 7B**). Nevertheless, the ROC curves of all three classes were significantly improved with Approach 6 compared to Approach 1. The ROC curves for the CN (AUC = 0.94) and AD (AUC = 0.98) classes demonstrated excellent performance. The ground truths and their corresponding predictions in three classes are illustrated in **Figure 8**.

DISCUSSION

This study's primary aim was to implement the XGBoost algorithm in early AD detection at the EMCI stage. The model performance significantly improved from 68.8% to 92.0% after adding two cognitive scores (MMSE and CDR) and selecting features (**Figure 6**

and **Table 4**). The final model achieved the highest accuracy of 92% by combining Pearson's correlations with BE for feature selection, reducing the number of features from 360 to 29 (**Figure 4** and **Table 3**). In addition, BE was explicitly recognized as the most suitable selection method (**Figure 6** and **Table 4**). The ROC curve illustrated excellent performance for Approach 6 (**Figure 8 B**), with the AD class having the highest AUC (0.98), followed by the CN class (0.94) and the EMCI class (0.88).

Feature weights

The BE method in Approach 4 showed that the hippocampus and temporal lobe features were the most important. This result is expected since structural changes in these regions are considered early indicators of MCI and AD³⁵. During the earliest stages of AD, brain atrophy typically follows the hippocampal pathway (entorhinal cortex, hippocampus, and posterior cingulate cortex) and is associated with early memory deficits³⁶. Furthermore, the variations in structural measures, including hippocampus and temporal lobe volumes, sulcus width and thickness, and subcortical nuclei volume, correlate with cognitive performance³⁷⁻⁴⁰.

Our study found that the two cognitive scores (MMSE and CDR) had substantially higher weights than the brain features. We conclude that the ML architecture designed in this study remains insufficiently effective. Clinically, these two scores are used as parts of the preferred standard diagnosis procedure for AD. Moreover, MMSE and CDR mainly depend on general cognitive and behavioral states rather than the underlying biological changes in the nervous system^{41,42}. Consequently, while the final model still shows considerable performance, it remains too dependent on symptom testing rather than brain structure changes.

Roles of cognitive scores and feature selection

Performance differed significantly between the first approach excluding the cognitive scores and the other approaches including them. Specifically, after adding MMSE and CDR to the feature set, the accuracy increased drastically by nearly 20%, from $69\% \pm 3\%$ to $86\% \pm 2\%$. We suggest that future model development should minimize the influences of the two scores in the prediction to make applying the model in the clinical setting less dependent on the availability of well-trained neurologists to conduct such cognitive tests. There has been a recent increase in the

number of studies completing this task. For example, Liu *et al.* reported a multi-model DL framework with accuracies of 88.9% for classifying AD and CN and 76.2% for classifying MCI and CN⁴³. Farooq *et al.* compared GoogLeNet, ResNet-18, and ResNet-152, reporting accuracies of 98% for all three models⁴⁴. However, most recent studies only used a DL approach, which could hinder technology acceptance by medical doctors⁴⁵.

Our study also illustrated that feature selection, especially BE and XGBI, plays a crucial role in the classification model. Both methods led to significant increases in model performance, which surpassed the results of other approaches. The reason is that, from a biological perspective, not all brain features contribute to AD pathology⁴⁶⁻⁴⁸. Several studies suggest that several brain regions are affected by AD-related atrophy, including the frontal, temporal, and parietal lobes or cerebellum brain regions⁴⁶⁻⁴⁸. Other feature selection methods also showed outstanding accuracy. For example, Fang *et al.* proposed several ML algorithms combined with goal-directed conceptual aggregation to demonstrate the effectiveness of this method compared to other approaches (PCA, least absolute shrinkage and selection operator, and univariate feature selection). They achieved 79.25 % accuracy in classifying CN vs. EMCI and 83.33% in classifying CN vs. LMCI⁴⁹. Khagi *et al.* combined SVM and K-nearest neighbors with one of four feature selection methods (ReliefF, Laplacian, UDFS, and Mutinffs), reporting accuracies of approximately 99% for AD classification⁵⁰.

Model selection and comparison

While the models in Approaches 3, 4, and 6 performed relatively similarly, Approach 6 was chosen to be the final model. Firstly, this approach achieved the highest accuracy (92%). Secondly, this model had a shorter training time (45.5 seconds) than Approach 3 (242.6 seconds). Moreover, in the feature selection step, Approach 6 selected features automatically, while Approach 4 required manual feature selection. In addition, by running GridSearch, Approach 6 could obtain optimal parameters compared to Approach 4 (without GridSearch).

Approaches 1 and 6 had greater difficulty classifying EMCI than the other classes. The AUC for the CN class was the lowest in Approach 1 (0.82) but increased significantly in Approach 6 (0.94). This increase indicates that feature selection may eliminate misleading features, which remained significant for CN classification⁵¹. However, the AUC of the EMCI

class increased slightly from 0.83 to 0.88; therefore, EMCI is the most challenging class for the model to identify. Brain structural changes in patients with EMCI are likely not prominent enough for the model to recognize easily. Moreover, the EMCI classification remains challenging, and this class often showed low accuracy in previous studies. For example, Goryawala *et al.* only achieved an accuracy of 0.616 for distinguishing CN and EMCI and 0.814 for distinguishing EMCI and AD⁵².

Overall, three-way classification in the AD diagnosis model still performs poorly. The proposed model is compared to current models in **Table 5**. However, most current models using three-way classification focus on the MCI class, while the EMCI class is more important in facilitating early AD diagnosis. This oversight underscores the distinctiveness of this study, which introduces novelty by addressing three-class classification involving EMCI, AD, and CN categories. Therefore, the proposed method shows substantial promise in its performance compared to other methods. Compared with state-of-the-art models for three-way classification, the method proposed in this study achieves promising performance with 92% accuracy. However, Ahmed *et al.* developed a multi-class deep CNN framework for early AD diagnosis, achieving 93.86% accuracy for three-way AD/MCI/CN classification⁵³. It is important to note that their focus was on MCI, whereas our study focuses on the more challenging EMCI classification. Consequently, our model offers a more sophisticated approach and, therefore, has a competitive advantage.

CONCLUSIONS

This study developed an ML model for early AD diagnosis based on structural MRI scans using XG-Boost to classify three classes: CN, EMCI, and AD. We also evaluated three feature selection methods (BE, XGBI, and PCA) to identify the optimal method for our model. The final model using BE with tuning parameters achieved the highest accuracy of 92%. The AUCs for the AD, CN, and EMCI classes were 0.98, 0.94, and 0.88, respectively. Compared to previous three-class classification methods, the proposed method appears promising for early AD detection.

While the XG-Boost model attained high accuracy with the aid of BE, several technical issues remain unsolved. Firstly, the AUC was lower for the EMCI class than for the CN and AD classes. Therefore, additional interventions in fitting parameters to enhance the performance of EMCI accuracy are essential. In addition,

the model should be modified to reduce its dependence on MMSE and CDR scores. Finally, the model should be tested on multi-datasets to optimize its performance.

ABBREVIATIONS

ADNI: Alzheimer's Disease Neuroimaging Initiative; **AD**: Alzheimer's disease; **AI**: Artificial Learning; **BE**: Backward Elimination; **CAD**: Computer-Aided Diagnosis; **CDR**: Clinical Dementia Rating; **CN**: Cognitive Normal; **CNN**: Convolutional Neural Network; **DL**: Deep Learning; **eTIV**: estimated Total Intracranial Volume; **EMCI**: Early MCI; **GMM**: Gaussian Mixture Model; **GridSearchCV**: Grid Search cross-validation; **GDCA**: Goal-Directed Conceptual Aggregation; **GLCM**: Gray Level Co-occurrence Matrix; **KNN**: K Nearest Neighbor; **LMCI**: Late MCI; **ML**: Machine Learning; **MCI**: Mild Cognitive Impairment; **MMSE**: Mini-Mental State Examination; **MRI**: Magnetic Resonance Imaging; **OLS**: Ordinary Least Square; **RELM**: Rough Extreme Learning Machine; **ROC-AUC**: Area Under The ROC Curve; **PET**: Positron Emission Tomography; **PCA**: Principle Component Analysis; **PC**: Principle Components; **sMRI**: structural MRI; **SVM**: Support Vector Machine; **SL**: Significance Level; **XGBI**: XG-Boost Importance.

ACKNOWLEDGMENTS

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) and the Alzheimer's Disease Metabolomics Consortium (ADMC). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf and <https://sites.duke.edu/adnimetab/team>

AUTHOR'S CONTRIBUTIONS

All authors contributed to the ideas, designed, did the experiments. All authors read and approved the final manuscript.

FUNDING

This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number NCM2020-28-01.

Table 5: Model performance of three-way classification in early diagnosis of Alzheimer

Study	Sample size	Method	Model performance
54	224 CN, 133 MCI, 85 AD	Modified Tresnet	63.2 %
55	200 CN, 441 MCI, 105 AD	Decision tree with linear discriminant analysis	66.7 %
56	197 CN, 330 MCI, 279 AD	3D CNN with 8 instance normalization layers	66.9 %
57	CN vs. MCI vs. AD	XG-Boost	66.8 %
58	229 CN, 398 MCI, 192 AD	VGG-16 (Visual Geometry Group 16)	80.66 %
59	115 CN, 133 MCI, 58 AD	ResNet-18 with Weighted Loss and Transfer Learning and Mish Activation	88.3 %
60	229 CN, 382 MCI, 187 AD	Combined Graph convolutional networks and CNN	89.4 %
Proposed method	221 CN, 221 MCI, 221 AD	XG-Boost and BE	92 %

AVAILABILITY OF DATA AND MATERIALS

The data that support the findings of this study are available in ADNI at <http://adni.loni.usc.edu/data-samples/access-data/>

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

- Wong W. Economic burden of Alzheimer disease and managed care considerations. *The American Journal of Managed Care*. 2020;26(8):177–83. PMID: 32840331.
- Kumar A, et al. Alzheimer Disease. 2021 Aug 11. StatPearls. Treasure Island (FL): StatPearls Publishing, 2022; 2022.
- Bi XA, Xu Q, Luo X, Sun Q, Wang Z. Analysis of progression toward Alzheimer's disease based on evolutionary weighted random support vector machine cluster. *Frontiers in Neuroscience*. 2018;12:716. PMID: 30349454. Available from: <https://doi.org/10.3389/fnins.2018.00716>.
- Tatiparti K, Sau S, Rauf MA, Iyer AK. Smart treatment strategies for alleviating tauopathy and neuroinflammation to improve clinical outcome in Alzheimer's disease. *Drug Discovery Today*. 2020;25(12):2110–29. PMID: 33011341. Available from: <https://doi.org/10.1016/j.drudis.2020.09.025>.
- Kang L, Jiang J, Huang J, Zhang T. Identifying early mild cognitive impairment by multi-modality mri-based deep learning. *Frontiers in Aging Neuroscience*. 2020;12:206. PMID: 33101003. Available from: <https://doi.org/10.3389/fnagi.2020.00206>.
- Zhang F, Pan B, Shao P, Liu P, Shen S, Yao P, et al. A single model deep learning approach for Alzheimer's disease diagnosis. *Neuroscience*. 2022;491:200–14. PMID: 35398507. Available from: <https://doi.org/10.1016/j.neuroscience.2022.03.026>.
- Xing X, Liang G, Zhang Y, Khanal S, Lin AL, Jacobs N. Advit: Vision transformer on multi-modality pet images for alzheimer disease diagnosis. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). 2022;2022:1–4. Available from: <https://doi.org/10.1109/ISBI52829.2022.9761584>.
- Diogo VS, Ferreira HA, Prata D, undefined Alzheimer's Disease Neuroimaging Initiative. Early diagnosis of Alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach. *Alzheimer & Research & Therapy*. 2022;14(1):107. PMID: 35922851. Available from: <https://doi.org/10.1186/s13195-022-01047-y>.
- Tábuas-Pereira M, Baldeiras I, Duro D, Santiago B, Ribeiro MH, Leitão MJ. Prognosis of early-onset vs. late-onset mild cognitive impairment: comparison of conversion rates and its predictors. *Geriatrics (Basel, Switzerland)*. 2016;1(2):11. PMID: 31022805. Available from: <https://doi.org/10.3390/geriatrics1020011>.
- Mirzaei G, Adeli H. Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia. *Biomedical Signal Processing and Control*. 2022;72:103293. Available from: <https://doi.org/10.1016/j.bspc.2021.103293>.
- Kolahkaj S, Zare H. A connectome-based deep learning approach for Early MCI and MCI detection using structural brain networks. *Neuroscience Informatics (Online)*. 2023;3(1):100118. Available from: <https://doi.org/10.1016/j.neuri.2023.100118>.
- Rice L, Wong E, Kolter Z. Overfitting in adversarially robust deep learning. In: *International Conference on Machine Learning*; 2020. p. 8093–8104. Available from: <http://proceedings.mlr.press/v119/rice20a>.
- Techa C, et al. Alzheimer's disease multi-class classification model based on CNN and StackNet using brain MRI data. In: and others, editor. *International Conference on Advanced Intelligent Systems and Informatics*. Springer; 2022. p. 248–259.

- Available from: https://doi.org/10.1007/978-3-031-20601-6_23.
14. Alorf A, Khan MU. Multi-label classification of Alzheimer's disease stages from resting-state fMRI-based correlation connectivity data and deep learning. *Computers in Biology and Medicine*. 2022;151:106240. PMID: 36423532. Available from: <https://doi.org/10.1016/j.compbiomed.2022.106240>.
 15. Alfalahi H, Dias SB, Khandoker AH, Chaudhuri KR, Hadjileontiadis LJ. A scoping review of neurodegenerative manifestations in explainable digital phenotyping. *NPJ Parkinson & Disease*. 2023;9(1):49. PMID: 36997573. Available from: <https://doi.org/10.1038/s41531-023-00494-0>.
 16. Garre-Olmo J. [Epidemiology of Alzheimer's disease and other dementias]. *Revista de neurologia*. 2018;66(11):377-86. PMID: 29790571. Available from: <https://europepmc.org/article/med/29790571>.
 17. Riek HC, Brien DC, Coe BC, Huang J, Perkins JE, Yep R, et al. Cognitive correlates of antisaccade behaviour across multiple neurodegenerative diseases. *Brain Communications*. 2023;5(2). PMID: 36970045. Available from: <https://doi.org/10.1093/braincomms/fcad049>.
 18. Jayasudha M, Elangovan M, Mahdal M, Priyadarshini J. Accurate estimation of tensile strength of 3D printed parts using machine learning algorithms. *Processes (Basel, Switzerland)*. 2022;10(6):1158. Available from: <https://doi.org/10.3390/pr10061158>.
 19. Sun X. Application and Comparison of Artificial Neural Networks and XGBoost on Alzheimer's Disease. In *Proceedings of the 2021 international conference on bioinformatics and intelligent computing*. 2021;2021:101-105. Available from: <https://doi.org/10.1145/3448748.3448765>.
 20. Ong H, et al. A Machine Learning Framework Based on Extreme Gradient Boosting for Intelligent Alzheimer's Disease Diagnosis Using Structure MRI. In: and others, editor. *International Conference on the Development of Biomedical Engineering in Vietnam*. vol. 2020. Springer; 2020. p. 815-827.
 21. Tuan TA, et al. Alzheimer's diagnosis using deep learning in segmenting and classifying 3D brain MR images. *The International Journal of Neuroscience*. 2020;132(7):689-98. PMID: 33045895. Available from: <https://doi.org/10.1080/00207454.2020.1835900>.
 22. Tuan TA, Pham TB, Kim JY, Tavares JM. Alzheimer's diagnosis using deep learning in segmenting and classifying 3D brain MR images. *International Journal of Neuroscience*. 2021;132(7):689-98. Available from: <https://doi.org/10.1080/00207454.2020.1835900>.
 23. FreeSurfer. 6 Aug 2021; Available from: <https://surfer.nmr.mgh.harvard.edu>; 2021.
 24. Geladi P, Linderholm J. Principal Component Analysis; 2020. Available from: <https://doi.org/10.1016/B978-0-12-409547-2.14892-9>.
 25. Lever J, Krzywinski M, Altman N. Principal component analysis. *Nature Methods*. 2017;14(7):641-2. Available from: <https://doi.org/10.1038/nmeth.4346>.
 26. Dunkler D, Plischke M, Leffondré K, Heinze G. Augmented backward elimination: a pragmatic and purposeful way to develop statistical models. *PLoS One*. 2014;9(11):e113677. PMID: 25415265. Available from: <https://doi.org/10.1371/journal.pone.01113677>.
 27. Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. vol. 777. John Wiley & Sons; 2008. Available from: <https://doi.org/10.1002/9780470770771>.
 28. Chen T, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*. 2015. 1(4): p. 1-4; 2015.
 29. Liu Y, Liu L, Yang L, Hao L, Bao Y. Measuring distance using ultra-wideband radio technology enhanced by extreme gradient boosting decision tree (XGBoost). *Automation in Construction*. 2021;126:103678. Available from: <https://doi.org/10.1016/j.autcon.2021.103678>.
 30. Mitchell R, et al. Xgboost: Scalable GPU accelerated learning. *arXiv preprint arXiv:1806.11248*, 2018.
 31. Guo J, Yang L, Bie R, Yu J, Gao Y, Shen Y. An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring. *Computer Networks*. 2019;151:166-80. Available from: <https://doi.org/10.1016/j.comnet.2019.01.026>.
 32. Pedregosa F, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*. 2011;2011:2825-2830. Available from: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https://>.
 33. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: A new perspective. *Neurocomputing*. 2018;300:70-9. Available from: <https://doi.org/10.1016/j.neucom.2017.11.077>.
 34. Liu J, Li R, Wu R. Feature selection for varying coefficient models with ultrahigh-dimensional covariates. *Journal of the American Statistical Association*. 2014;109(505):266-74. PMID: 24678135. Available from: <https://doi.org/10.1080/01621459.2013.850086>.
 35. Wisse LE, Biessels GJ, Heringa SM, Kuijff HJ, Koek DH, Luijckx PR, et al. Hippocampal subfield volumes at 7T in early Alzheimer's disease and normal aging. *Neurobiology of Aging*. 2014;35(9):2039-45. PMID: 24684788. Available from: <https://doi.org/10.1016/j.neurobiolaging.2014.02.021>.
 36. Scahill RI, Schott JM, Stevens JM, Rossor MN, Fox NC. Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99(7):4703-7. PMID: 11930016. Available from: <https://doi.org/10.1073/pnas.052587399>.
 37. Ridha BH, Anderson VM, Barnes J, Boyes RG, Price SL, Rossor MN. Volumetric MRI and cognitive measures in Alzheimer disease: comparison of markers of progression. *Journal of Neurology*. 2008;255(4):567-74. PMID: 18274807. Available from: <https://doi.org/10.1007/s00415-008-0750-9>.
 38. Hua X, Lee S, Yanovsky I, Leow AD, Chou YY, Ho AJ, et al. Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. *NeuroImage*. 2009;48(4):668-81. PMID: 19615450. Available from: <https://doi.org/10.1016/j.neuroimage.2009.07.011>.
 39. Visser PJ, Scheltens P, Verhey FR, Schmand B, Launer LJ, Jolles J. Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment. *Journal of Neurology*. 1999;246(6):477-85. PMID: 10431775. Available from: <https://doi.org/10.1007/s004150050387>.
 40. Dickerson BC, Bakkour A, Salat DH, Feczko E, Pacheco J, Greve DN. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral Cortex (New York, NY)*. 2009;19(3):497-510. PMID: 18632739. Available from: <https://doi.org/10.1093/cercor/bhn113>.
 41. Brossard B. Objectifying Dementia: the Use of the Mini-Mental State Exam in Medical Research and Practice, in *Measuring Mental Disorders*. 2018, Elsevier. p. 127-154; 2018. Available from: <https://doi.org/10.1016/B978-1-78548-305-9.50004-3>.
 42. Sinha A, Sinha A. Mild Cognitive Impairment and its Diagnosis to Progression to Dementia with Several Screening Measures. *The Open Psychology Journal*. 2018;11(1):142-7. Available from: <https://doi.org/10.2174/1874350101811010142>.
 43. Liu M, Li F, Yan H, Wang K, Ma Y, Shen L, et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. *NeuroImage*. 2020;208:116459. PMID: 31837471. Available from: <https://doi.org/10.1016/j.neuroimage.2019.116459>.
 44. Farooq A, et al. A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: and others, editor. 2017 *IEEE International Conference on Imaging systems and tech-*

- niques (IST). IEEE.; 2017. p. 1–6. Available from: <https://doi.org/10.1109/IST.2017.8261460>.
45. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ*. 2019;7:e7702. PMID: 31592346. Available from: <https://doi.org/10.7717/peerj.7702>.
 46. Patel H, Dobson RJ, Newhouse SJ. A meta-analysis of Alzheimer's disease brain transcriptomic data. *Journal of Alzheimer's Disease*. 2019;68(4):1635–56. PMID: 30909231. Available from: <https://doi.org/10.3233/JAD-181085>.
 47. Jagust W. Imaging the evolution and pathophysiology of Alzheimer disease. *Nature Reviews Neuroscience*. 2018;19(11):687–700. PMID: 30266970. Available from: <https://doi.org/10.1038/s41583-018-0067-3>.
 48. Gautam P, Cherbuin N, Sachdev PS, Wen W, Anstey KJ. Relationships between cognitive function and frontal grey matter volumes and thickness in middle aged and early old-aged adults: the PATH Through Life Study. *NeuroImage*. 2011;55(3):845–55. PMID: 21255657. Available from: <https://doi.org/10.1016/j.neuroimage.2011.01.015>.
 49. Fang C, Li C, Forouzaneshad P, Cabrerizo M, Curiel RE, Loewenstein D, et al. Gaussian discriminative component analysis for early detection of Alzheimer's disease: A supervised dimensionality reduction algorithm. *Journal of Neuroscience Methods*. 2020;344:108856. PMID: 32663548. Available from: <https://doi.org/10.1016/j.jneumeth.2020.108856>.
 50. Khagi B, Kwon GR, Lama R. Comparative analysis of Alzheimer's disease classification by CDR level using CNN, feature selection, and machine-learning techniques. *International Journal of Imaging Systems and Technology*. 2019;29(3):297–310. Available from: <https://doi.org/10.1002/ima.22316>.
 51. Khaire UM, Dhanalakshmi R. Stability of feature selection algorithm: A review. *Journal of King Saud University Computer and Information Sciences*. 2019;34(4):1060–73. Available from: <https://doi.org/10.1016/j.jksuci.2019.06.012>.
 52. Goryawala M, et al. Inclusion of neuropsychological scores in atrophy models improves diagnostic classification of Alzheimer's disease and mild cognitive impairment. *Computational intelligence and neuroscience*. 2015;2015:865265. Available from: <https://doi.org/10.1155/2015/865265>.
 53. Ahmed HM, Elsharkawy ZF, Elkorany AS. Alzheimer disease diagnosis for magnetic resonance brain images using deep learning neural networks. *Multimedia Tools and Applications*. 2023;82(12):17963–77. Available from: <https://doi.org/10.1007/s11042-022-14203-1>.
 54. Xu Z, Deng H, Liu J, Yang Y. Diagnosis of Alzheimer's Disease Based on the Modified Tresnet. *Electronics (Basel)*. 2021;10(16):1908. Available from: <https://doi.org/10.3390/electronics10161908>.
 55. Lin W, Gao Q, Du M, Chen W, Tong T. Multiclass diagnosis of stages of Alzheimer's disease using linear discriminant analysis scoring for multimodal data. *Computers in Biology and Medicine*. 2021;134:104478. PMID: 34000523. Available from: <https://doi.org/10.1016/j.compbiomed.2021.104478>.
 56. Liu S, Yadav C, Fernandez-Granda C, Razavian N, et al. On the design of convolutional neural networks for automatic detection of Alzheimer's disease. In *Machine Learning for Health Workshop*. 2020;2020:184–201. Available from: <http://proceedings.mlr.press/v116/liu20a>.
 57. Stubblefield J, et al. Study the combination of brain MRI imaging and other datatypes to improve Alzheimer's disease diagnosis. *MedRxiv*, 2022: p. 2022; 2022.
 58. Lim BY, Lai KW, Haiskin K, Kulathilake KA, Ong ZC, Hum YC. Deep learning model for prediction of progressive mild cognitive impairment to Alzheimer's disease using structural MRI. *Frontiers in Aging Neuroscience*. 2022;14:876202. PMID: 35721012. Available from: <https://doi.org/10.3389/fnagi.2022.876202>.
 59. Oktavian MW, Yudistira N, Ridok A. Classification of Alzheimer's Disease Using the Convolutional Neural Network (CNN) with Transfer Learning and Weighted Loss. *arXiv preprint arXiv:2207.01584*, 2022; 2022.
 60. Lin L, Xiong M, Zhang G, Kang W, Sun S, Wu S, et al. A Convolutional Neural Network and Graph Convolutional Network Based Framework for AD Classification. *Sensors (Basel)*. 2023;23(4):1914. PMID: 36850510. Available from: <https://doi.org/10.3390/s23041914>.