

Developing a Machine Learning Model to Detect the Early Stage of Alzheimer's Disease in Individuals with Mild Cognitive Impairment Utilizing Brain Imaging

Jatuphol Pholtaisong¹, Punnut Chirdchuvutikun², Chanodom Lisantia²,
Phuphiphat Jaikaew^{1,*}

¹Department of Biotechnology, Faculty of Science and Technology, Thammasat University (Rangsit Center), Pathum Thani 12120, Thailand

² Suankularb Wittayalai Rangsit School, Pathum Thani 12120, Thailand

*Corresponding author e-mail: pupipat@tu.ac.th

Received: 23 July 2025 / Revised: 23 February 2026 / Accepted: 30 May 2026

Abstract

Alzheimer's disease (AD), a form of dementia, significantly impairs an individual's ability to perform daily activities. The prevalence of AD patients in Thailand will increase to 1.1 million by 2030, marking AD as a critical concern in the nation's aging society. Mild Cognitive Impairment (MCI), often seen as a normal part of aging, has the potential to progress into AD. Early intervention in MCI cases could help delay the onset of AD. Our study focuses on creating a classification model leveraging neuroimaging data to distinguish between AD, MCI, MCI that has transitioned to AD (MCI_AD), and cognitively normal (CN) individuals. We utilized MRI brain images from the ADNI database, employing the FreeSurfer program to transform neuroimages into numerical data arrays for feature extraction. The study further included data preprocessing with standardization and feature selection by the scikit-learn library. Upon constructing five supervised machine learning models, we identified the one with the best performance through comparative analysis. The model's effectiveness was evaluated using ROC curves, with the Support Vector Machine (SVM) model showing superior performance, achieving an area under the curve (AUC) of 0.764. Comparatively, Logistic Regression, NuSVC, Extra Trees Classifier, and MLP Classifier scored AUCs of 0.761, 0.758, 0.745, and 0.736, respectively. In summary, the SVM model demonstrated a notable ability to classify early stages of MCI and AD from other groups. Nevertheless, for further advancements in AD and MCI detection, exploring a wider array of datasets and feature extraction methodologies is imperative to improve the model's performance.

Keywords: Machine learning, Alzheimer's disease, Mild cognitive impairment, Neuroimaging

1. Introduction

Alzheimer's disease (AD) is an age-related neurodegenerative disorder and a leading cause of dementia. AD is a fatal condition that impairs an individual's ability to perform daily tasks. The increasing prevalence of dementia in countries with low- and middle-income countries is having serious global implications. This increase is largely driven by AD cases. The ongoing pandemic is expected to alter the dynamics of dementia incidence, with a specific impact on long-term care. Older individuals with cognitive impairment or dementia are at a higher risk of viral infections (Alzheimer's Disease International & McGill University, 2021). This issue is particularly relevant to Thailand, which is currently grappling with an aging society. Studies indicate that the country is expected to have 1.1 million patients with Alzheimer's disease in 2030 (Na Chiangmai, 2023). Experts highlight that the greatest risk factor for Alzheimer's is old age, making the situation even more pronounced for Thailand as it navigates this period of population aging. Alzheimer's Disease International (ADI) states, "In the absence of a cure or a treatment that is globally accessible, risk reduction remains the most feasible and proactive way to combat dementia" (Alzheimer's Disease International & McGill University, 2021). If the disease can be detected at an early stage, it will be advantageous for preparing to handle the disease; in the early stage, patients may exhibit a condition called Mild Cognitive Impairment (MCI) which is often associated with mild forgetfulness in

elderly individuals. Clinicians consider the cognitive decline to be the result of the prodromal stage of AD (Alzheimer's Disease International & McGill University, 2021).

There are various diagnostic methods. For example, genetic testing, spinal fluid analysis, and brain imaging techniques, such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET). Of relevance to this study is MRI because other common biomarkers for MCI and AD, such as fluorodeoxyglucose or amyloid positron emission tomography and cerebrospinal fluid measures, are more invasive, more expensive, and less accessible compared to MRI scanning (Courchesne et al., 2001). MRI is a non-invasive imaging technology that produces detailed three-dimensional anatomical images. MRI assessment of the brain has demonstrated four different patterns of atrophy in patients with AD: typical AD, limbic-predominant AD, hippocampal-sparing AD, and a subtype with minimal atrophy.

Many studies have attempted to diagnose preclinical Alzheimer's disease. Some used a variety of data types other than MRI scans, consisting of FDG-PET, the APOE gene, and cerebrospinal fluid that explored new diagnostic approaches and achieved better performance at 94.86% accuracy and a 93.59% AUC score with multiclass SVM. However, in terms of actual use, they are more invasive, and more expensive (Gupta et al., 2019). There is another study using only MRI scans that explored the generalizability across acquisition protocols, including Magnetization-Prepared Rapid Gradient Echo (MPRAGE) and Inversion Recovery Spoiled Gradient Recalled Echo (IR-SPGR) data types. The generalizability of both datasets in the ADNI and Open Access Series of Imaging Studies (OASIS) databases may be enhanced to improve model performance. However, their proposed graph theory features do not show improved performance. Their performance is 62.1% balanced accuracy and an AUC score of 77.7% (Diogo et al., 2022). Moreover, a previous study used MRI and PET scans with a deep learning transformer model that achieved an accuracy of 91.18% and an F1 score of 0.90 (Altay et al., 2021).

More recent advancements from 2024 to 2026 have further expanded the state-of-the-art in automated Alzheimer's disease detection. The field has seen a notable shift toward advanced deep learning architectures, including the use of conditional diffusion models to synthesize MRI scans to improve the training of 3D convolutional networks (Turrisi & Patané, 2026). Additionally, contemporary research heavily emphasizes the use of sequential feature embeddings and regularized multikernel support vector machines to develop highly accurate prognostic models for prodromal stages (Olatunde et al., 2024). However, while these advanced deep neural architectures and complex embeddings establish high accuracy benchmarks, they are frequently bottlenecked by massive computational overhead and algorithmic opacity. Therefore, optimizing highly interpretable, computationally lightweight classical machine learning pipelines that operate on structured morphological biomarkers remains an indispensable parallel research direction, particularly for immediate deployment in resource-constrained clinical environments.

Our work explores the possibility of making a preclinical diagnosis of Alzheimer's disease by diagnosing among cognitively normal (CN), mild cognitive impairment (MCI), and mild cognitive impairment that may progress to Alzheimer's disease (MCI_AD). Our study aims to estimate AD risk and to develop an AD diagnostic method by exploring multi-model training and evaluating multiple feature selection and feature extraction techniques.

2. Materials and Methods

2.1 Dataset collection

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) (Aisen et al., 2015). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Initially, MRI scans were processed to extract morphometric brain features using automated neuroimaging analysis tools. These features include measurements related to brain structures that are commonly associated with neurodegeneration. Then, the MPAGE-type dataset for MRI neuroimaging was collected from the ADNI1 and ADNI2 projects in the ADNI database and saved in 2D DICOM formats. Lastly, based on the clinical diagnosis provided in the ADNI database, the subjects were categorized into four classes: cognitively normal (CN), mild cognitive impairment (MCI), Alzheimer's disease (AD), and individuals with mild cognitive impairment who later progressed to Alzheimer's disease (MCI_AD) (Petersen et al., 2010). These class labels were used as the target variables for the machine learning classification models, as seen in Figure 1.

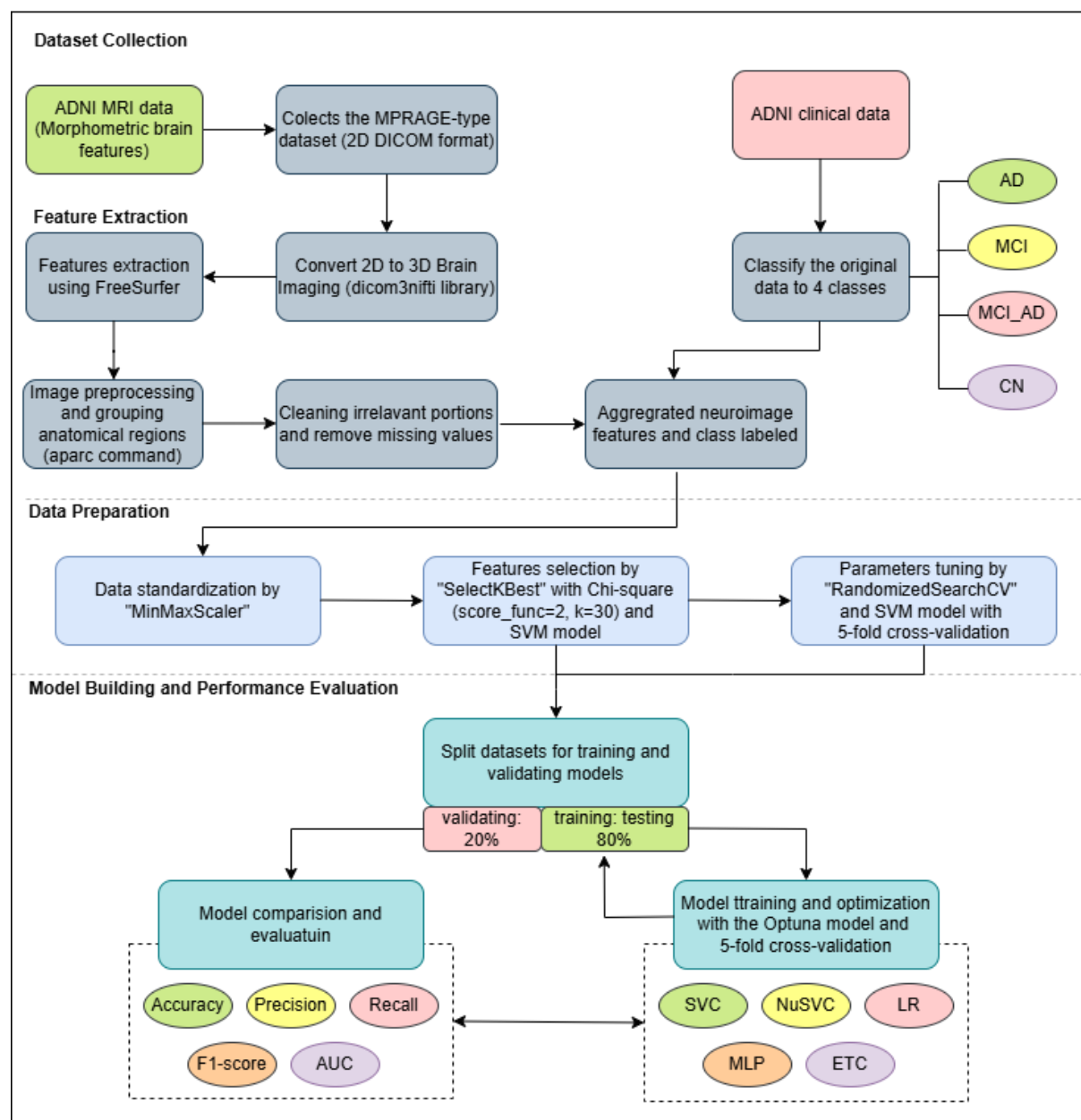


Figure 1. The workflow of machine learning model development for predicting AD and MCI includes the following steps: collecting brain imaging data from the ADNI database, extracting features with the FreeSurfer program, aggregating the input files into a table, preprocessing features and samples, training the model and tuning parameters, and validating the performance of the models.

2.2 Feature extraction

A 2D DICOM file format was converted to a 3D NIfTI file format by the dicom2nifti library. Then, the FreeSurfer program was used for feature extraction, image preprocessing, and grouping anatomical regions in neuroimages with the aparc command. The output files, rh.aparc.stats and lh.aparc.stats, were retrieved from the FreeSurfer program.

Following the feature extraction, irrelevant portions of the dataset were removed, and any missing values were eliminated. Then, the dataset was converted from its original format to a CSV format. After this format change, all dataset tables were aggregated into one, and a class label was assigned to each dataset entry.

2.3 Feature preprocessing

The feature preprocessing began with the reduction of dataset variance using the MinMaxScaler. Next, the number of features was reduced through a feature selection method, with features ranked using a default model. The best feature selection method was subsequently chosen. Then, the feature selection parameters were adjusted using RandomizedSearchCV, combined with the SVM model in the Optuna-based pipeline. Finally, the optimal hyperparameters for feature selection were determined, as seen in Figure 1.

2.4 Model training and optimization

We evaluated several supervised learning algorithms to develop a robust machine learning framework for the early detection of Alzheimer's disease using brain imaging characteristics. These included Support Vector Machine (SVM or SVC), NuSVM, Logistic Regression (LR), Multilayer Perceptron (MLP), and Extra Trees Classifier (ETC). We chose these algorithms because they are examples of different types of machine learning methods that are often used for biomedical classification tasks. These include linear models (LR), kernel-based classifiers (SVN and NuSVM), neural networks (MLP), and ensemble tree-based methods (Extra Trees Classifier).

Table 1. Optimized hyperparameters obtained using RandomizedSearchCV.

Model Name	Hyperparameter	Search Range	Optimized Value
SVC	C	[0.01, 0.1, 1, 10, 100]	10
	kernel	{linear, rbf, poly}	rbf
	gamma	{scale, auto, 0.001, 0.01, 0.1}	scale
	degree	[2-5] (poly kernel)	3
NuSVC	nu	[0.1-0.9]	0.5
	kernel	{linear, rbf, poly}	rbf
	gamma	{scale, auto, 0.01, 0.1}	scale
MLP Classifier	hidden_layer_sizes	{(50), (100), (100,50), (100,100)}	(100, 50)
	activation	{relu, tanh, logic}	relu
	solver	{adam, lbfgs}	adam
	alpha	{0.0001-0.01}	0.001
	learning_rate	{constant, adaptive}	adaptive
	penalty	{l1, l2, elasticnet}	l2
Logistic Regression	C	[0.01, 0.1, 1, 10, 100]	1
	solver	{liblinear, saga}	liblinear
	max_iter	[100-500]	200
Extra Trees Classifier	n_estimators	[100-1000]	500
	max_depth	[None, 10, 20, 30]	20
	min_samples_split	[2-10]	4
	min_samples_leaf	[1-5]	2
	max_features	{sqrt, log2, None}	sqrt

Hyperparameters were optimized using RandomizedSearchCV with 5-fold cross-validation in the Scikit-learn framework. The reported values correspond to the best-performing configuration based on cross-validated accuracy.

Support Vector Machine (SVM) models demonstrate enhanced efficacy when the number of features substantially surpasses the sample size, a common characteristic of neuroimaging-derived biomedical datasets (Bi et al., 2018; Cuingnet et al., 2011). Logistic regression (LR) was utilized as a fundamental linear model, selected for its inherent simplicity and interpretability (Arora et al., 2018), qualities that are particularly valuable in the context of medical research. The Multilayer Perceptron (MLP) classifier was employed to investigate potential nonlinear relationships among morphometric brain parameters (Bhattacharya et al., 2022). Furthermore, the Extra Trees Classifier was integrated as an ensemble tree-based method, known for its capacity to effectively model complex feature interactions and accommodate nonlinear decision boundaries (Ossai & Wickramasinghe, 2022).

This multiclassification model was built by the Optuna framework within the scikit-learn package. Initially, the hyperparameters were optimized with RandomizedSearchCV and as shown in Table 1, followed by training the model using five-fold cross-validation. To strictly prevent data leakage and ensure reproducibility, we ensured that a subject-level data split was maintained throughout the cross-validation process, meaning no individual subject's neuroimaging data appeared in both the training and testing sets simultaneously. Next, the feature selection and model training were integrated into a pipeline, culminating in the final optimized model.

2.5 Model comparison and evaluation

The performance metrics of each classification model, including accuracy, precision, recall, macro-averaged F1 score, and AUC score, were estimated. Next, the top five best multiclassification models from the scikit-learn modules were compared. Then, the performance of each optimized model was evaluated using ROC curves and AUC scores.

3. Results and Discussion

3.1 Data collection and features extraction

In total, the dataset used in this study includes 90 subjects, comprising 28 AD, 26 CN, 15 MCI, and 21 MCI_AD cases. The demographic information indicated that the study population was predominantly Caucasian. The age range of the participants was 55-90 years, and the gender was not considered in this study. The class distribution was moderately imbalanced, with approximately 23.3% belonging to the converter group (MCI_AD) and 16.67% belonging to the stable MCI group (MCI), with the remaining 31.1% and 28.9% belonging to AD and CN, respectively. To address potential bias arising from class imbalance, appropriate evaluation metrics and model validation procedures were applied during the training and testing phases.

The dataset obtained from the ADNI database consists of MPRAGE-type MRI images. The FreeSurfer program extracted 612 features from the neuroimages. After preprocessing by SelectKBest with the chi-square method with $k = 30$, we obtained 30 features.

Figure 2 shows the normalized feature distribution, comparing data before and after preprocessing, with values ranging between -2 and 2. The selected morphometric features include gray volume, average thickness, Gaussian Curvature, Mean Curvature, and Folding Index. The complete set of morphometric features comprises Number of Vertices, Surface Area, gray volume, Average Thickness, Standard Deviation Thickness, Mean Curvature, Gaussian Curvature, Folding Index, and Curvature Index. The morphometric features that have the most significant impact are the volume and size of the brain (Heckemann et al., 2011). Our research utilized the gray volume to determine volume, the average thickness to assess regional brain size, and Gaussian Curvature to evaluate brain curvature, which affects memory (Pini et al., 2016). We primarily chose the anatomical regions: inferior parietal, inferior temporal, middle temporal, and entorhinal, with some inclusion of bankssts, fusiform, lateral occipital, lateral orbitofrontal, precuneus, pars opercularis, pars triangularis, superior temporal, and superior frontal (Heckemann et al., 2011). In AD, the brain regions that contribute the most are the middle frontal, superior frontal, anterior cingulate, temporal, and precentral. While in MCI, the main affected areas are the middle frontal, superior frontal, inferior frontal, and precentral. We prioritize the temporal part of the brain for feature selection, with some consideration given to the frontal part. Our main features come from the temporal lobe, while

our minor features come from the frontal lobe and other brain regions involved in sensory processing (e.g., the lateral occipital area), speech (e.g., the pars opercularis), and memory (e.g., the precuneus). Our study showed that significant changes in both AD and MCI classes occur in the temporal lobe of the brain.

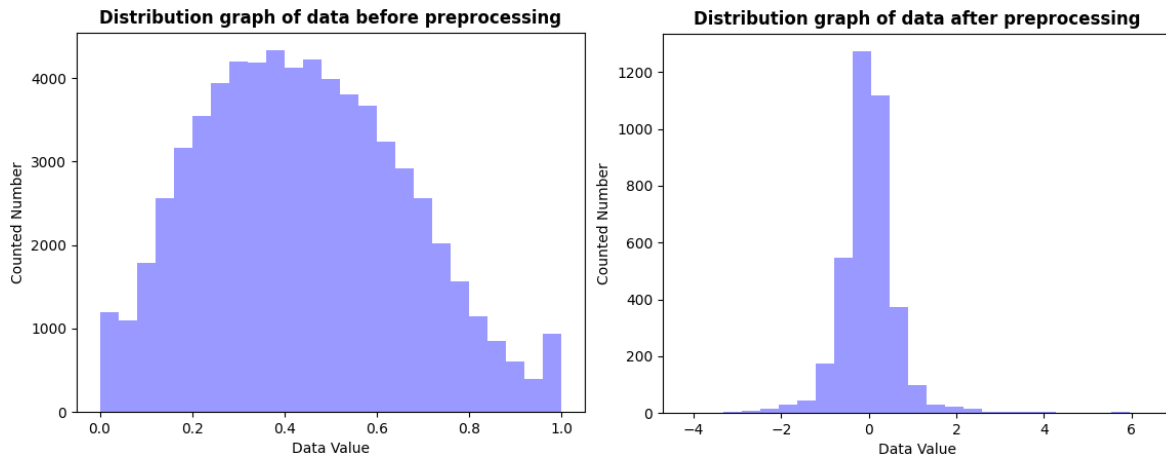


Figure 2. A distribution graph of data before and after preprocessing. The x-axes show data values of each feature. The y-axes show the counted number of features related to each data value.

3.2 Comparison area under curves (AUC) of feature selection methods with feature dimension reduction methods

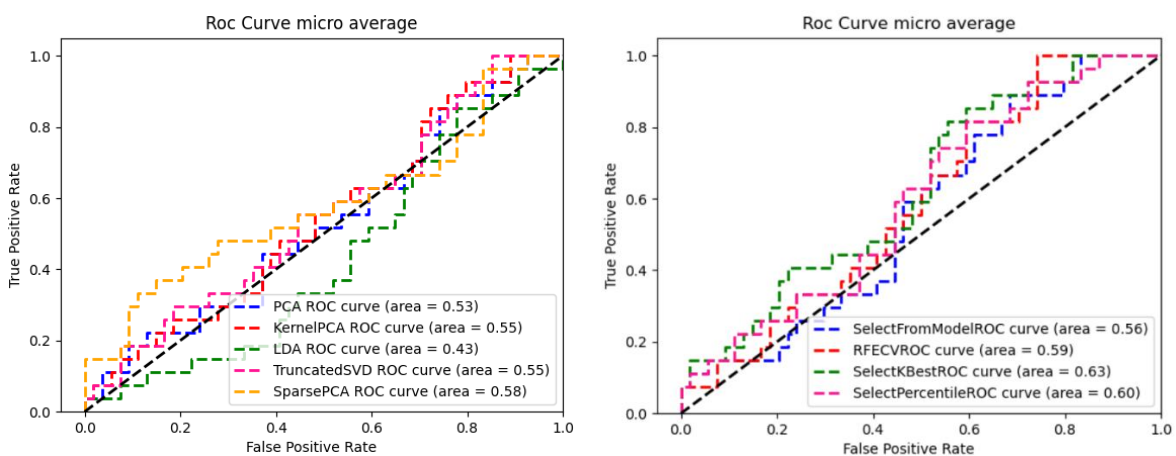


Figure 3. The ROC curve of feature dimensionality reduction methods compared with feature selection methods. The x-axes represent the false positive rate, and the y-axes represent the true positive rate. The left figure shows the five algorithms from features dimension reduction methods. The right figure shows the four algorithms from features selection methods.

The best data scaler that is used to reduce the variance of a dataset is the MinMaxScaler. The difference between the feature selection group and the feature extraction group is that feature selection has a higher AUC score but lower accuracy while feature extraction has a lower AUC score but higher accuracy (illustrated in Figure 3). We preferred the AUC score for ranking since it provides a more comprehensive measure of overall performance compared to accuracy. Feature selection is used to create a feature-reduced group. The best feature selection method is SelectKBest with hyperparameters `score_func=chi2`, `k=30`, which was optimized by Pipeline and RandomizedSearchCV. In a study by Khedher et al. (2015), PLS feature extraction was employed and utilized a similar SVM model to ours. However, their model included three output class labels, one of which was the absence of the MCI_AD class. They achieved 85.41% accuracy, 85.95% sensitivity, and 84.86% specificity due

to a better ability to distinguish between MCI and AD compared to other PLS-based feature extraction methods. Another approach is the use of truncated SVD for their four types of features (Gupta et al., 2019). The truncated SVD is used because it preserves features from each data type by merging them, rather than eliminating them as in feature selection.

3.3 Comparison of five machine learning models with area under curve (AUC)

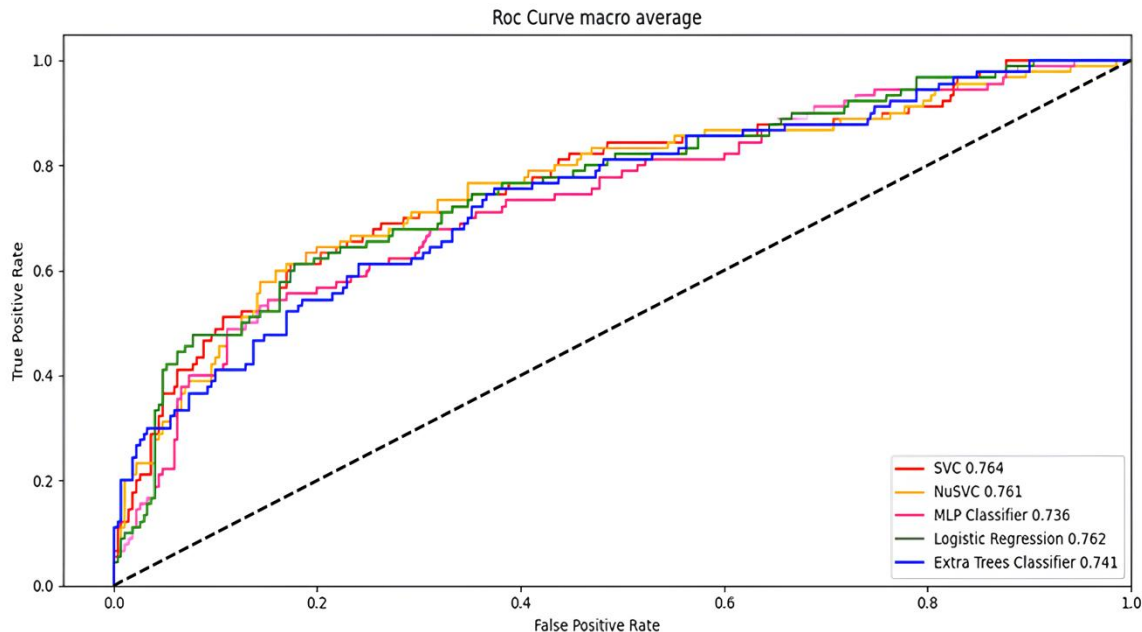


Figure 4. The figure shows the micro-averaged ROC curve and AUC score ranking of five models. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. ‘AUC’ stands for area under the curve, with the best value approaching 1.0 and the least desirable value nearing 0.0.

The AUC score is used to explain the ability to classify classes of the model (Hajian-Tilaki, 2013). It is shown on the right-hand side in Figure 4. The SVC or SVM model achieves the best AUC score out of five models, followed by Logistic Regression, NuSVC, Extra Trees Classifier, and MLP Classifier, respectively. Overall, AUC score levels are 73.5%-76.5%. The results are similar to other work that focuses on diagnosing preclinical Alzheimer's Disease getting a 0.777 (77.7%) AUC score (Diogo et al., 2022).

Confusion matrices were created for the best classifier (SVC) to better examine its classification behavior, as illustrated in Figure 5. The confusion matrices describe the proportion of true positives, true negatives, false positives, and false negatives for each class. This study offers further insight into the performance of the models across different classes and is useful in identifying potential misclassification tendencies across Alzheimer's disease (AD), mild cognitive impairment (MCI), MCI that may progress to AD (MCI_AD), and cognitively normal (CN) groups.

Figure 5 presents the normalized confusion matrix for the four-class classification task. The matrix illustrates the percentage distribution of predictions for each class. The results indicate that the highest classification performance was observed for the CN group (73.08%). Moderate performance was obtained for the MCI group (60.00%). However, some misclassification occurred between MCI and CN, as well as between MCI_AD and CN, indicating overlap in structural brain characteristics among these clinical stages. The results are similar to Cuingnet et al. (2011), who reported that automatic classifiers on ADNI structural MRI showed high CN vs. AD accuracy (81-95%) but lower sensitivity for MCI_AD vs. CN (~70%), due to shared gray matter density in the hippocampus/entorhinal cortex and white matter integrity (Cuingnet et al., 2011).

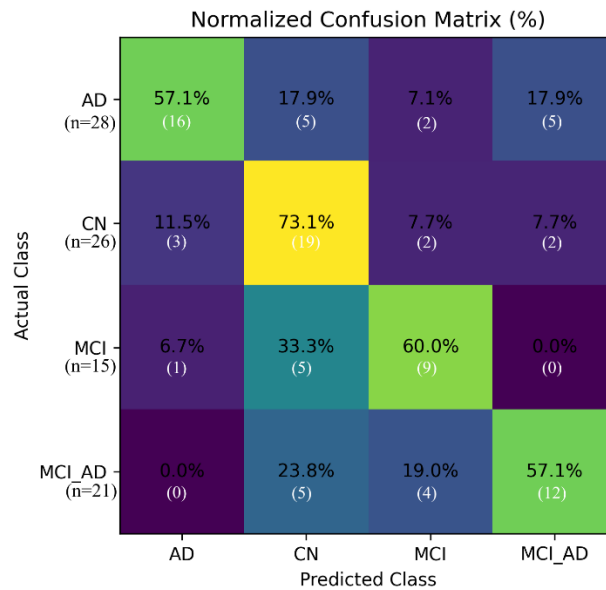


Figure 5. Normalized confusion matrix (%) for the four-class classification problem (AD, CN, MCI, and MCI_AD). Rows represent the actual class labels, while columns represent the predicted class labels. The values indicate the percentage of predictions for each class.

Table 2. Performance comparison of five machine learning models with metrics values.

Model Name	Accuracy	Precision	Recall	F1 score	AUC score
SVC	0.62	0.63	0.62	0.62	0.764
NuSVC	0.60	0.60	0.62	0.60	0.761
MLP Classifier	0.53	0.52	0.53	0.53	0.736
Logistic Regression	0.54	0.53	0.54	0.54	0.762
Extra Trees Classifier	0.54	0.53	0.52	0.52	0.741

From Table 2, the best performing model with the highest scores in both accuracy and AUC score is the SVC or SVM model with an accuracy of 0.62 and an AUC score of 0.764. NuSVC classifier performs slightly lower in AUC score compared to Logistic Regression, but it achieves higher accuracy and F1 score, indicating greater specificity and sensitivity, resulting in more accurate classifications than Logistic Regression. The difference in its AUC score and other scores places it third overall. The fourth is Extra Trees Classifier, which has higher specificity than the MLP classifier but has lower sensitivity, but in terms of AUC score it gets clearly higher than MLP classifier. From Table 2 and Figure 4, the SVM model is the best-performing model in this study.

Our work, while possibly not optimal for real-world applications, achieves an acceptable classification rate and introduces a novel approach in machine learning methodology to address the challenge of distinguishing between cognitively normal individuals and mild cognitive impairment that progresses into Alzheimer's disease (Markesbery, 2010). We suggest that the multiclass model will be more useful in practical applications. A similar study (Hwang et al., 2023) incorporated three types of data of MRI scans, genetic variables, and demographic variables, which are age, sex, years of education, intracranial volume, and handedness. The study utilized genetic variables, specifically the number of *APOE* $\epsilon 4$ alleles from the ADNI database, in conjunction with the HexaGAN deep learning model. This approach yielded an AUC score of 0.8609 and an F1 score of 0.7596. The HexaGAN model effectively addressed the issue of class imbalance, unlike the MLP+CNN model discussed in the paper, which struggled with this problem and obtained an AUC score of 0.7343 and an F1 score of 0.5607 (Hwang et al., 2023). Another study (Gupta et al., 2019) utilized an SVM model similar to our study but with a specific emphasis on utilizing four distinct data types: MRI scans, FDG PET scans, cerebrospinal fluid, and the APOE gene. The feature extractor used for FDG-PET scans and MRI scans was the Niftyreg software, achieving

a performance accuracy of 0.9486 and an AUC score of 0.9359. They utilized several features from various datasets; however, the acquisition of four separate data types is too expensive for practical applications.

There are many attempts to diagnose preclinical Alzheimer's disease. Some used a variety of data types other than MRI scans consisting of FDG-PET, the APOE gene, and cerebrospinal fluid, that explored new diagnostic approaches and achieved better performance at 94.86% accuracy with a 93.59% AUC score with multiclass SVM, but in terms of actual use, it is too expensive for patients to undergo multiple diagnostic tests (Gupta et al., 2019). Another study used only MRI scans (Diogo et al., 2022) that explored generalizability across acquisition protocols of MPRAGE and IR-SPGR and generalizability across datasets of ADNI and OASIS with improved performance. However, their proposed Graph theory features do not show any sign of better performance. Their performance is 62.1% balanced accuracy and an AUC score of 77.7%. Another study used MRI and PET scans with a deep learning transformer model that achieved an accuracy of 91.18% and an F1 score of 0.90 (Altay et al., 2021).

During cross-validation, the Support Vector Machine (SVM) model showed the best overall performance, as measured by accuracy and the Area Under the Curve (AUC) score, compared to the other algorithms that were evaluated. Specifically, the SVM model obtained the highest AUC score (0.764) compared with the other models evaluated in this study. These results suggest that SVM is particularly effective for capturing the structural brain differences associated with Alzheimer's disease progression. Consistency with a previous study confirms that automatic classification algorithms (such as SVM) could be useful in distinguishing AD patients from healthy controls (Oliveira et al., 2010).

The dataset used in this study consists of morphometric features extracted from MRI brain images using the FreeSurfer software. Such neuroimaging datasets typically contain relatively high-dimensional numerical features compared to the number of samples. Algorithms such as SVM are particularly well suited for this type of data because they are effective in high-dimensional feature spaces and can construct robust decision boundaries through kernel functions, which are similar to the methodology proposed by Zhang et al. (2018). This research employed recursive feature elimination (RFE) in conjunction with Support Vector Machine (SVM) analysis, and leave-one-out cross-validation (LOOCV) was utilized to ascertain the most suitable feature dimensions, thereby supporting the viability of this methodology for diagnostic image analysis in clinical settings (Zhang et al., 2018).

Furthermore, while SVM models are generally effective when the feature set is extensive compared to the sample size, this finding contrasts with Vichianin et al. (2021), who observed diminished classification accuracy when brain volumetric features were incorporated into SVM models. In contrast, the use of clinical parameters, the mental state examination score, controlled oral word association tests, learning and memory tests, the clock-drawing test, and construction praxis as features for SVM models provided good accuracy (Vichianin et al., 2021).

The experimental findings indicate that the suggested machine learning model may accurately differentiate individuals with mild cognitive impairment (MCI) who are susceptible to early-stage Alzheimer's disease. The performance metrics collected demonstrate that the integration of brain imaging features with machine learning algorithms can identify nuanced structural patterns linked to early neurodegenerative alterations. These results align with prior research indicating the efficacy of MRI-based biomarkers and machine learning methodologies for the early identification of Alzheimer's disease (Oliveira et al., 2010; Zhang et al., 2018). Recent studies have demonstrated that structural brain abnormalities in areas such as the hippocampus and temporal lobe might act as significant predictors for disease progression (Nie et al., 2017; Planche et al., 2022).

The main problem in our research is the feature extraction method and the limitations in feature performance. For features that remain constant over time, their inherent limitations restrict the extent to which our model tuning can have a significant impact. Another main problem is class imbalance in the dataset that limits our ability to incorporate additional data into the model and leads to overfitting in the MCI and CN classes.

Machine learning may be simpler, but deep learning is more successful in this field because of its ability to learn flexible features during training, which allow for improved feature representation learning. Most machine learning methods require a feature extraction program that takes a long time to complete, making them challenging to use in practice. In contrast, deep learning does not require this process (Zhao et al., 2023). The imbalanced class dataset is likely to cause overfitting due to the unequal distribution of important features across

classes, especially classes that are more frequently observed in real-world settings, like CN and MCI. The most effective way to use the preprocessed features and tuned model that we used is to train the model on features before preprocessing and then to integrate the preprocessed features and tuned model using the pipeline library and then train them on the original dataset. Finally, our research introduced a novel approach to detecting preclinical Alzheimer's disease by addressing the challenge of distinguishing between normal mild cognitive impairment and preclinical Alzheimer's disease-related mild cognitive impairment using machine learning. This differs from previous studies that primarily focused on diagnosing Alzheimer's disease (Arya et al., 2023). This model may not achieve performance comparable to deep learning methods, but it is simpler and less resource-consuming for model development. Our model focuses primarily on utilizing MRI scans, which are widely used in hospitals and relatively cost-effective, contrasting with many models that utilize multiple data types for improved performance but are less practical for real-world applications.

4. Conclusion

Here we present an optimized classical machine learning framework for classification of cognitively normal (CN), mild cognitive impairment (MCI), and MCI progressing to Alzheimer's disease (MCI_AD) using features from structural MRI. An integrated pipeline comprising FreeSurfer-based feature extraction, MinMax normalization, SelectKBest feature selection, and Support Vector Machine (SVM) classification was developed to ensure methodological consistency and reproducibility. The proposed SVM model reported a mean accuracy of 0.62 ± 0.11 and AUC of 0.761 ± 0.084 , which is an acceptable baseline performance for a computationally efficient approach. The current results demonstrate the continued relevance of classical ML methods for neuroimaging applications, particularly in situations with limited data availability and computational resources. The implementation of optimized hyperparameter tuning also helped ensure stable model performance across cross-validation settings. But there are some limitations that need to be recognized. Static morphometric features from FreeSurfer can constrain the model from learning complex spatial dependencies in brain structure. Also, dataset characteristics such as residual class imbalance may introduce bias and lead to overfitting. Using chi-square feature selection on continuous MRI-derived features can also restrict the optimal feature representation. Future work should focus on improving both model performance and clinical applicability. We suggest that ensemble learning techniques such as XGBoost and LightGBM serve as more robust baselines and that deep learning approaches should be explored, especially 3D convolutional neural networks, to better utilize volumetric imaging data. Moreover, more suitable feature selection techniques (e.g., ANOVA F-test or mutual information) and interpretability frameworks like SHAP should be applied to increase predictive power and clinical confidence. To conclude, this work provides a systematic and efficient baseline for the classification of AD with classical machine learning, and although further validation on larger and multi-center datasets is needed, the proposed framework offers a realistic basis for building scalable and interpretable neuroimaging-based diagnostic support systems.

Acknowledgements

This project was supported by Science Classroom in University Affiliated School (SCiUS) and Bioinformatics Laboratory in the Department of Biotechnology, Thammasat University. The funding of SCiUS is provided by the Ministry of Higher Education, Science, Research, and Innovation.

Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer

Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provides funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Conflict of Interest

The authors declare that there are no conflicts of interest.

References

- Aisen, P. S., Petersen, R. C., Donohue, M., Weiner, M. W., & Alzheimer's Disease Neuroimaging Initiative. (2015). Alzheimer's Disease Neuroimaging Initiative 2 clinical core: Progress and plans. *Alzheimer's & Dementia*, 11(7), 734–739.
- Altay, F., Sánchez, G. R., James, Y., Faraone, S. V., Velipasalar, S., & Salekin, A. (2021). Preclinical stage Alzheimer's disease detection using magnetic resonance image scans. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(17), 15088–15097. <https://doi.org/10.1609/aaai.v35i17.17772>
- Alzheimer's Disease International, & McGill University. (2021). *World Alzheimer Report 2021: Journey through the diagnosis of dementia*. Alzheimer's Disease International. <https://www.alzint.org/resource/world-alzheimer-report-2021/>
- Arora, A., Lin, J. J., Gasperian, A., Maldjian, J., Stein, J., Kahana, M., & Lega, B. (2018). Comparison of logistic regression, support vector machines, and deep learning classifiers for predicting memory encoding success using human intracranial EEG recordings. *Journal of Neural Engineering*, 15(6), 066028. <https://doi.org/10.1088/1741-2552/aae131>
- Arya, A. D., Verma, S. S., Chakarabarti, P., Chakarabarti, T., Elngar, A. A., Kamali, A. M., & Nami, M. (2023). A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. *Brain Informatics*, 10(1), 17. <https://doi.org/10.1186/s40708-023-00195-7>
- Bhattacharya, S., Bennet, L., Davidson, J. O., & Unsworth, C. P. (2022). Multi-layer perceptron classification & quantification of neuronal survival in hypoxic-ischemic brain image slices using a novel gradient direction, grey level co-occurrence matrix image training. *PLoS One*, 17(12), e0278874. <https://doi.org/10.1371/journal.pone.0278874>
- Bi, X., Shu, Q., Sun, Q., & Xu, Q. (2018). Random support vector machine cluster analysis of resting-state fMRI in Alzheimer's disease. *PLoS One*, 13(3), e0194479. <https://doi.org/10.1371/journal.pone.0194479>
- Courchesne, E., Karns, C. M., Davis, H. R., Ziccardi, R., Carper, R. A., Tigue, Z. D., Chisum, H. J., Moses, P., Pierce, K., Lord, C., Lincoln, A. J., Pizzo, S., Schreibman, L., Haas, R. H., Akshoomoff, N. A., & Courchesne, R. Y. (2001). Unusual brain growth patterns in early life in patients with autistic disorder: An MRI study. *Neurology*, 57(2), 245–254. <https://doi.org/10.1212/WNL.57.2.245>
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.-O., Chupin, M., Benali, H., & Colliot, O. (2011). Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, 56(2), 766–781. <https://doi.org/10.1016/j.neuroimage.2010.06.013>
- Diogo, V. S., Ferreira, H. A., Prata, D., & Alzheimer's Disease Neuroimaging Initiative. (2022). Early diagnosis of Alzheimer's disease using machine learning: A multi-diagnostic, generalizable approach. *Alzheimer's Research & Therapy*, 14(1), 107. <https://doi.org/10.1186/s13195-022-01047-y>
- Gupta, Y., Lama, R. K., Kwon, G. R., & Alzheimer's Disease Neuroimaging Initiative. (2019). Prediction and classification of Alzheimer's disease based on combined features from apolipoprotein-E genotype, cerebrospinal fluid, MR, and FDG-PET imaging biomarkers. *Frontiers in Computational Neuroscience*, 13, 72. <https://doi.org/10.3389/fncom.2019.00072>
- Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, 4(2), 627–635.

- Heckemann, R. A., Keihaninejad, S., Aljabar, P., Gray, K. R., Nielsen, C., Rueckert, D., Hajnal, J. V., Hammers, A., & Alzheimer's Disease Neuroimaging Initiative. (2011). Automatic morphometry in Alzheimer's disease and mild cognitive impairment. *NeuroImage*, 56(4), 2024–2037.
<https://doi.org/10.1016/j.neuroimage.2011.03.014>
- Hwang, U., Kim, S. W., Jung, D., Kim, S., Lee, H., Seo, S. W., Seong, J. K., Yoon, S., & Alzheimer's Disease Neuroimaging Initiative. (2023). Real-world prediction of preclinical Alzheimer's disease with a deep generative model. *Artificial Intelligence in Medicine*, 144, 102654.
<https://doi.org/10.1016/j.artmed.2023.102654>
- Khedher, L., Ramírez, J., Góriz, J. M., Brahim, A., Segovia, F., & Alzheimer's Disease Neuroimaging Initiative. (2015). Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing*, 151, 139–150.
<https://doi.org/10.1016/j.neucom.2014.09.072>
- Markesbery, W. R. (2010). Neuropathologic alterations in mild cognitive impairment: A review. *Journal of Alzheimer's Disease*, 19(1), 221–228. <https://doi.org/10.3233/JAD-2010-1220>
- Na Chiangmai, N. (2023). *Spontaneous speech analysis for detecting mild cognitive impairment and Alzheimer's disease in Thai older adults* [Doctoral dissertation, Burapha University]. CINECA IRIS Institutional Research Information System. <https://iris.unitn.it/handle/11572/389289>
- Nie, X., Sun, Y., Wan, S., Zhao, H., Liu, R., Li, X., Wu, S., Nedelska, Z., Hort, J., Qing, Z., Xu, Y., & Zhang, B. (2017). Subregional structural alterations in hippocampus and nucleus accumbens correlate with the clinical impairment in patients with Alzheimer's disease clinical spectrum: Parallel combining volume and vertex-based approach. *Frontiers in Neurology*, 8, 399. <https://doi.org/10.3389/fneur.2017.00399>
- Olatunde, O. O., Oyetunde, K. S., Han, J., Khasawneh, M. T., & Yoon, H. (2024). Multiclass classification of Alzheimer's disease prodromal stages using sequential feature embeddings and regularized multikernel support vector machine. *NeuroImage*, 304, 120929. <https://doi.org/10.1016/j.neuroimage.2024.120929>
- Oliveira, P. P. M., Jr., Nitrini, R., Busatto, G., Buchpiguel, C., Sato, J. R., & Amaro, E., Jr. (2010). Use of SVM methods with surface-based cortical and volumetric subcortical measurements to detect Alzheimer's disease. *Journal of Alzheimer's Disease*, 19, 1263–1272. <https://doi.org/10.3233/JAD-2010-1322>
- Ossai, C. I., & Wickramasinghe, N. (2022). GLCM and statistical features extraction technique with Extra-Tree Classifier in Macular Oedema risk diagnosis. *Biomedical Signal Processing and Control*, 73, 103471. <https://doi.org/10.1016/j.bspc.2021.103471>
- Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack, C. R., Jr., Jagust, W. J., Shaw, L. M., Toga, A. W., Trojanowski J. Q., & Weiner, M. W. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI) clinical characterization. *Neurology*, 74(3), 201–209.
- Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavado, E., Galluzzi, S., Marizzoni, M., & Frisoni, G. B. (2016). Brain atrophy in Alzheimer's disease and aging. *Ageing Research Reviews*, 30, 25–48.
<https://doi.org/10.1016/j.arr.2016.01.002>
- Planche, V., Manjon, J. V., Mansencal, B., Lanuza, E., Tourdias, T., Catheline, G., & Coupé, P. (2022). Structural progression of Alzheimer's disease over decades: The MRI staging scheme. *Brain Communications*, 4(3). <https://doi.org/10.1093/braincomms/fcac109>
- Turrisi, R., & Patané, G. (2026). Generating synthetic MRI scans for improving Alzheimer's disease diagnosis. *Medical Image Analysis*, 110, 103947. <https://doi.org/10.1016/j.media.2026.103947>
- Vichianin, Y., Khummongkol, A., Chiewvit, P., Raksthaput, A., Chaichanettee, S., Aoonkaew, N., & Senanarong, V. (2021). Accuracy of support-vector machines for diagnosis of Alzheimer's disease, using volume of brain obtained by structural MRI at Siriraj Hospital. *Frontiers in Neurology*, 12, 640696.
<https://doi.org/10.3389/fneur.2021.640696>
- Zhang, Y., & Liu, S. (2018). Analysis of structural brain MRI and multi-parameter classification for Alzheimer's disease. *Biomedizinische Technik*, 63(4), 427–437. <https://doi.org/10.1515/bmt-2016-0239>
- Zhao, Z., Chuah, J. H., Lai, K. W., Chow, C. O., Gochoo, M., Dhanalakshmi, S., Wang, N., Bao, W., & Wu, X. (2023). Conventional machine learning and deep learning in Alzheimer's disease diagnosis using neuroimaging: A review. *Frontiers in Computational Neuroscience*, 17, 1038636.
<https://doi.org/10.3389/fncom.2023.1038636>