Multiclass classification of Alzheimer's disease prodromal stages using sequential feature embeddings and regularized multikernel support vector machine

Oyekanmi O. Olatunde [a], Kehinde S. Oyetunde [b], Jihun Han [c], Mohammad T. Khasawneh [a], Hyunsoo Yoon [c,*], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] *Department of Systems Science and Industrial Engineering, Binghamton University, NY 13902, USA*
[b] *Department of Mechanical and Aerospace Engineering, Hong Kong University of Science and Technology, Hong Kong, PR China*
[c] *Department of Industrial Engineering, Yonsei University, Seoul 03722, Republic of Korea*

ARTICLE INFO

ABSTRACT

The detection of patients in the cognitive normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) stages of neurodegeneration is crucial for early treatment interventions. However, the heterogeneity of MCI data samples poses a challenge for CN vs. MCI vs. AD multiclass classification, as some samples are closer to AD while others are closer to CN in the feature space. Previous attempts to address this challenge produced inaccurate results, leading most frameworks to break the assessment into binary classification tasks such as AD vs. CN, AD vs. MCI, and CN vs. MCI. Other methods proposed sequential binary classifications such as CN vs. others and dividing others into AD vs. MCI. While those approaches may have yielded encouraging results, the sequential binary classification method makes interpretation and comparison with other frameworks challenging and subjective. Those frameworks exhibited varying accuracy scores for different binary tasks, making it unclear how to compare the model performance with other direct multiclass methods. Therefore, we introduce a classification framework comprising unsupervised ensemble manifold regularized sparse low-rank approximation and regularized multikernel support vector machine (SVM). This framework first extracts a joint feature embedding from MRI and PET neuroimaging features, which were then combined with the Apoe4, Adas11, MPACC digits, and Intracranial volume features using a regularized multikernel SVM. Using that framework, we achieved a state-of-the-art (SOTA) result in a CN vs. MCI vs. AD multiclass classification (mean accuracy: 84.87±6.09, F1 score: 84.83±6.12 vs 67.69). The methods generalize well to binary classification tasks, achieving SOTA results in all but the CN vs. MCI category, which was slightly lower than the best score by just 0.2%.

1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disease. It is the primary cause of dementia among older adults, constituting approximately 70% of all dementia in individuals aged **60 years and older** (Jena et al., 2015; "2023 Alzheimer's disease facts and figures," 2023). Its hallmark characteristics include the accumulation of β-amyloid (Aβ) plaques and neurofibrillary tangles, which are associated with synapse loss and neurodegeneration, ultimately leading to deficits in memory, language, and cognition (Klunk et al., 2004; Rowe et al., 2007; Weiner et al., 2012). During its prodromal stages, patients may experience heightened confusion, increased withdrawal from social activities, and agitation, with these symptoms escalating in severity as the disease progresses (Golde, 2022, "What Are the Signs of Alzheimer's Disease?", 2022). No interventions capable of curing AD exist. Nonetheless, treatments have been developed to provide temporary symptom relief or

alter disease progression ("Treatments for Alzheimer's," n.d.; "How Is Alzheimer's Disease Treated?," 2023). Early clinical detection is often challenging because symptoms typically manifest ≥15 years after the onset of the disease pathology (Bateman et al., 2012; Weiner et al., 2012; "2023 Alzheimer's disease facts and figures," 2023). This suggests that upon clinical confirmation of the disease, it may often be too late to restore individuals to their pre-morbid state (Fleisher et al., 2009). Moreover, given that therapy is likely to be most effective at symptom onset, early detection of AD pathology is crucial for timely interventions.

Artificial intelligence researchers have employed either single-modality approaches (Shi et al., 2021; Zhang et al., 2022) or multi-modal biomarker datasets (Lin et al., 2021a; Wolf et al., 2022; Nan et al., 2022; Chang and Brainerd, 2022) to create various machine learning (ML) frameworks, **to** address that challenge. Those frameworks are typically designed to classify Alzheimer's disease prodromal stages, including the Cognitive Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's disease (AD). Single modality **considers** only one type of biomarker, such as MRI or PET. For instance, Hsu et al., (2013) employed genetic data, Jitsuishi and Yamaguchi, (2022) used MRI data, and Shen et al., (2019) made use of FDG-PET data, all aiming to predict AD conversion. In contrast, multimodal studies incorporate multiple heterogeneous types of data to improve AD predictions. For example, Jie et al., (2015) integrated MRI and FDG-PET data; Kim and Lee, (2018) incorporated CSF with MRI and FDG-PET, and Caminiti et al., (2018) used FDG-PET and CSF. Each biomarker has distinct sensitivities to the different preclinical phases of the disease (Donohue et al., 2014; Jena et al., 2015). Hence, several machine learning studies have used multiple biomarkers to improve the classification accuracy of these stages and demonstrated that using multiple biomarkers can provide complementary insights (Jie et al., 2015; Shi et al., 2022; Ban et al., 2023). However, determining the optimal method for integrating different biomedical datasets poses a challenge because direct feature concatenation assumes uniform sensitivity to the diagnosis across different modalities. When the number of features from each modality differs significantly, the high-dimensional feature sets dominate joint feature representation and, subsequently, the fused predictor (Gray et al., 2013; Singanamalli et al., 2017). Moreover, feature concatenation can significantly increase the overall feature dimensions, thereby increasing the susceptibility of the model to overfitting problems. Moreover, combining different modality datasets increases the probability of data nonlinearity, rendering many conventional linear dimension reduction methods suboptimal, particularly as they overlook the interrelationships across modalities. Furthermore, multimodal datasets typically exhibit distinct distributions, which can render direct feature fusion suboptimal (Shi et al., 2022). Many previous studies that used MRI modality did not account for cerebral atrophy caused by normal aging (Weiner et al., 2012), potentially causing their models to learn artifacts unrelated to the disease pathology.

Furthermore, the intermediate state, MCI, encompasses groups of patients demonstrating early symptoms of cognitive decline (Singanamalli et al., 2017). However, these symptoms typically do not significantly interfere with their everyday lives. Certain MCI groups progress to AD through time while others may return to the CN state. Therefore, those characteristics make the subpopulation of MCI patients less than completely heterogeneous with some closer to AD and others to CN. This diversity poses a significant challenge in accurately classifying CN vs. MCI vs. AD in a one-shot whether using unimodal or multimodal datasets (Liu et al., 2015; Lin et al., 2021). A typical multiclass strategy involves simplifying the tasks into several binary classification tasks such as CN vs AD, CN vs MCI, and AD vs MCI (Suk et al., 2014; Li et al., 2018; Akramifard et al., 2020). Another method integrating different modalities to classify these binary subcategories (Singanamalli et al., 2017). The authors contended that the same combination of data modalities might not be optimal or most informative for all the subclinical categories. While this approach may yield promising results, the sequential binary classification approach employed for evaluation

complicates interpretation and comparison with other frameworks. Furthermore, these frameworks have reported varying accuracy scores across classes, thereby, creating ambiguity regarding how to assign a model accuracy score and compare their performances to those of other methods.

This study introduces a unified framework for the multimodal multiclass classification of AD preclinical stages at one shot that incorporates an unsupervised ensemble manifold regularized sparse low-rank approximation and a regularized multikernel SVM. We improve multiclass classification accuracy in one shot, with its **robustness** influenced by the modalities employed and the algorithms or frameworks utilized. Furthermore, because neuroimaging features are highly correlated, we show that neglecting the right-hand-side MRI features or left-hand-side PET features might serve as a straightforward solution to minimize the curse of dimensionality on our limited sample datasets. Finally, we conducted a comprehensive experiment on the ADNI1 dataset, using age-stratified cross-validation techniques across numerous replications.

2. Materials and methods

2.1. Data collection

Data used in this study 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 assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). We downloaded the shared processed MRI scans named: ADNI1 Complete 3Yr 1.5T from the ADNI database. That file comprises standardized T1-weighted MRI images obtained from the same 1.5T scanner, meeting the quality control criteria. Those images have undergone multiplanar reconstruction (MPR), GradWarp, B1 non-uniformity correction, and N4 intensity normalization (Wyman et al., 2013). Subsequently, we filtered the corresponding processed FDG-PET image of the same subjects using an advanced query. The PET images are those that underwent maximum preprocessing, which encompassed co-registration of separate image frames to one another, averaging the frames, standardizing image and voxel size to 1mm x 1mm x 1mm ensuring uniform resolution. Overall 194 subjects with MRI and corresponding PET images from the same session (specifically, session M00 or baseline) were selected. Table 1 shows their characteristics. The IDs of the individuals in our analysis are available on our GitHub page. Additionally, other features, such as genetic information (Apoe4), cognitive scores (ADAS11: Alzheimer Disease Assessment Scale & MPACC: Modified Preclinical Alzheimer's Cognitive Composite scores), and Intracranial volume (ICV) were extracted from the ADNIMERGE clinical file.

Table 1
Summary of subject characteristics.

	AD	MCI	CN
N	53	81	60
Sex (M/F)	29/24	59/22	33/27
Age (y)	75.66±6.68	74.74±6.56	74.7±5.13
Education	14.58±3.04	15.98±2.65	15.53±3.27
CDR	4.69±1.54	1.44±0.75	0.04±0.14
MMSE	23.15±2.04	27.28±0.99	29.00±1.70

*CDR: Clinical Dementia Rating * MMSE: Mini-Mental State Examination
* Age, Education, CDR & MMSE scores are the mean ± standard deviation

2.2. Data preprocessing

A primary source of variability is in the choice of a preprocessing and feature extraction method (Zhou et al., 2022). Hence, we chose the CLINICA software pipeline developed by the ARAMIS Lab for neuroimage processing and feature extraction, since it can be applied directly to most public neuroimage datasets, and carefully designed for the reproducibility of results (Routier et al., 2021). That pipeline supported FSL, Statistical Parametric Mapping (SPM), and FreeSurfer software for preprocessing. First, all downloaded neuroimages were converted to the Brain Imaging Data Structure (BIDS) format using the ADNI-to-BIDS command (Gorgolewski et al., 2016). Subsequently, after which the t1-volume pipeline is used (Samper-González et al., 2018; Routier et al., 2021). This t1-volume pipeline serves as a wrapper encompassing Segmentation, Run Dartel, and Normalize to MNI Space routine, which is implemented in SPM. The procedures began with unified segmentation, which simultaneously performs tissue segmentation, bias field correction, and spatial normalization of the input MRI image (Ashburner and Friston, 2005). This tissue segmentation operation separates each MRI image into its corresponding gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Following that, Run Dartel which employed the DARTEL algorithm of diffeomorphic image registration is used to generate a group template from each tissue probability map of each subject (i.e. from GM, WM & CSF) on the native space (Ashburner, 2007). The Dartel2MNI method is then applied to register each native space image into the MNI space (Ashburner, 2007). Finally, the gray matter density of a set of anatomical regions obtained from different atlases in the MNI space is computed(Tzourio-Mazoyer et al., 2002; Shattuck et al., 2008; Gousias et al., 2008; Rolls et al., 2015; Joliot et al., 2015; Gaser et al., 2022; Hammers et al., 2003). Only the features extracted using the Hammers et al., (2003) atlas was used for this study. After executing the t1-volume pipeline to extract the MRI gray matter features, we then ran the pet-volume pipeline (Samper-González et al., 2018; Routier et al., 2021). The pipeline initially performed an intra-subject registration of each PET image into the space of the corresponding MRI image using SPM. Then partial volume correction was made using the PETPVC toolbox (Thomas et al., 2016). Subsequently, each PET image is spatially normalized into the MNI space using the DARTEL deformation model of SPM. Furthermore, the voxel intensities were normalized using the average PET uptake in the pons region. Finally, the average intensities of each region in the selected atlases were computed in the MNI space. Sixty-eight regional features were extracted from each processed MRI and PET scan using the Hammers atlas. Additionally, we included the Apoe4 genetic features, two cognitive features (ADAS11, MPACC digits), and ICV (collectively referred to as NEW in this report). While the MMSE and CDR are other frequently used features, we opted not to include them in this study because they

were used to label the ground truth ADNI data employed in this study. Finally, the MRI (68 features), PET (68 features), and NEW (4 features) features are normalized, and eventually used in this study.

2.3. Framework and algorithms

Fig. 1. Illustrates the proposed framework. The process begins with neuroimage preprocessing and feature extraction. Then, a joint low-dimensional embedding was generated from the MRI and PET features, using unsupervised ensemble manifold regularized sparse low-rank approximation, as described in Section 2.3.1. Finally, a regularized multikernel SVM was trained using the joint embedding and the NEW features as inputs. The output of the multikernel SVM determined the desired class label of each test sample.

2.3.1. Joint feature embedding

For the joint feature embedding, we used the ensemble manifold regularized sparse low-rank approximation (EMR-SLRA), originally proposed for creating a multiview embedding for facial recognition (Zhang et al., 2015). In pattern recognition such as human identification or hyperspectral data classification, objects are often represented by features from different views with each view summarizing a unique characteristic of the object. These features, though from different views, are often complementary to one another. This embedding method can be adopted for MRI and PET features because of their high feature dimensionality and complementary nature in terms of the progression of AD pathology (Zhang et al., 2011; Jie et al., 2015). Due to the high collinearity among and within neuroimage features, conventional feature selection, which is commonly employed, might not capture essential interrelationships between features. For instance, in cases where the relationship between features A and B explains feature C, conventional feature selection algorithms may not adequately capture the relationship. Particularly, this embedding method was selected because it considers the complementarity of different modalities, and also its ability to minimize the influence of noisy input on the output embeddings. The algorithm comprised three components: Multiview Matrix Low-rank Approximation, Ensemble Manifold Regularization, and Noise Reduction or Sparsity Constraint.

Given a multimodal dataset of N samples, where each modality, m dataset is represented as a feature matrix, $X^m = \{x_1^m, x_2^m, ..., x_N^m\}$ where $x_i^m \in \mathbb{R}^p \; \forall \; i = 1, ..., N$, we aim to determine a joint embedding $Y \in \mathbb{R}^{dxN}$ for all M-th modalities where $d \leq p * M$. If the multimodal feature matrix is denoted as $X = [X^1, X^2, ..., X^m] \in \mathbb{R}^{pMxN}$, the joint feature embedding Y and projection matrix U can be derived by solving the following optimization problem.
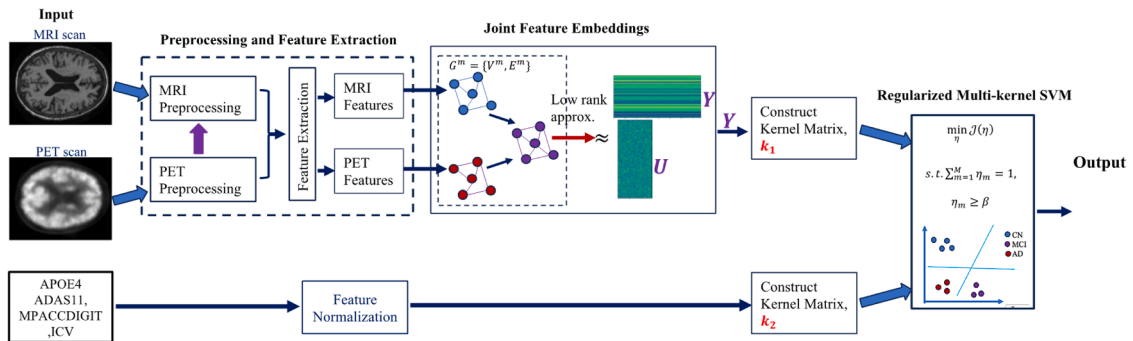
Fig. 1. A framework for CN vs. MCI vs. AD multiclass classification. First, MRI and PET scans undergo preprocessing and feature extraction. Subsequently, a multimodal joint feature embedding, Y is created from MRI & PET, where U represents the projection matrix for transforming new test samples. Finally, kernel matrices, k_1 and k_2 are constructed from the feature embeddings and the NEW feature set (Apoe4, ADAS11, MPACCDIGIT, and ICV) are used as inputs to a regularized multikernel SVM for classification. To compute the embeddings, undirected graphs G^m are constructed from each neuroimage feature set to capture their respective manifold structures.

$$\min_{U,\overline{X},Y,\beta} ||\overline{X} - UY||^2 + \lambda_1 \sum_{m=1}^{M} (\beta_m)^r tr\left(YL^mY^T\right) + \lambda_2 ||\overline{X} - X||_{2,1}$$
$$s.t.\ \mathrm{U}^{\mathrm{T}}\mathrm{U} = \mathrm{I} \qquad (1)$$
$$\sum_{m=1}^{M} \beta_m = 1$$
$$\beta_m > 0$$

Where λ_1 and λ_2 are regularization parameters for the manifold and group sparsity constraints, respectively. L^m is the Graph Laplacian Matrix of the m-th modality, I is an identity matrix, $\overline{X}$ denotes the ideal feature matrix with a reduced noise level of the input feature matrix, X. U represents the projecting matrix for new test samples to the embedding space of Y. β_m represents the combination weight for the manifold term of the m-th data modality, r denotes a weight scaling factor and M is the number of data modalities (Table 2, additional information). The first term, $|| \overline{X} - UY ||^2$ represents a low-rank approximation of the feature matrix $\overline{X}$ to the embedding Y, and projection matrix U, aiming for the approximation that minimizes the error (De La Torre, 2012). The middle term, known as the manifold regularization term, ensures that samples in the lower dimensional space preserve their similarities as they did in the original input space. This approach aligns with that of the Laplacian Eigenmaps (Belkin and Niyogi, 2003), aiming to preserve the data manifold in the m-th modality as much as possible. The final term, $|| \overline{X} - X ||_{2,1}$ denotes the noise reduction term where $|| * ||_{2,1}$ is the $\ell_{2,1}$-norm.

To determine the Laplacian matrix, L^m, a connected and undirected weighted graph, $G^m = (V^m, E^m)$, was constructed from each data modality with data points as the vertices, and edges connecting the nearest points. Each edge weight denotes the similarity value between two endpoints, S_{ij}^m, which was calculated using the heat kernel (Belkin and Niyogi, 2003) (Table 2).

The objective function in Eq (1) is non-convex, making simultaneous optimization of the variables challenging. However, alternate optimization algorithms have been used to address this problem (Lei et al., 2019; Zhang et al., 2015). The procedure is described in Appendix A.

2.3.2. Multikernel SVM

The support vector machine, popularly known as SVM, is a supervised ML algorithm developed at AT&T Bell Laboratories by Vapnik et al. in 1992 (Cortes and Vapnik, 1995). It is a popular algorithm used in neurocomputing within the context of precision psychiatry because of its effectiveness in handling high-dimensional datasets (Chen et al., 2022). Given a dataset of training examples $S = \{(x_1,y_1), (x_2,y_2),...,(x_N,$

$y_N)\}$, where $x_i \in \mathbb{R}^d$ and class label $y_i \in \{-1, +1\}$, we define the dual single kernel SVM as follows;

$$\max_{\alpha \geq 0} \sum_{i}^{N} \alpha_i - \frac{1}{2} \sum_{i}^{N} \sum_{j}^{N} \alpha_i \alpha_j y_i y_j k(x_i, x_j)$$
$$s.t. 0 \leq \alpha_i \leq C \qquad (2)$$
$$\sum_{i}^{N} \alpha_i y_i = 0 \ \forall i \in [N]$$

where α_i is the Lagrangian multiplier, and $k(x_i, x_j)$ represents the kernel function. SVM learns a linear discriminant function in Eq (3) which maximizes the minimum margin between support vectors by solving the optimization function in Eq (2).

$$y = sgn\left(\sum_{i}^{N} \alpha_i y_i k(x_i, x) + b \right) \qquad (3)$$

where sgn is the sign or signum function.

Multikernel SVM (MKSVM) combines multiple homogeneous or heterogeneous kernels corresponding to feature representations from varying or similar information sources (Gonen et al., 2011). In the case of multikernel SVM, Eq (2) is typically modified, and the kernel function becomes parameterized by its information source as shown in Eq (4).

$$k(I, \mathrm{x}_j) \rightarrow k_\eta (I_i, \mathrm{x}_j)$$
$$\Rightarrow k_\eta(I, \mathrm{x}_j) = f_\eta\left(\left\{ k_m\left(\mathrm{x}_i^m, \mathrm{x}_j^m\right) \right\}_{m=1}^{p} \Big| \eta \right) \qquad (4)$$

where p represents the number of information sources, $f_\eta : \mathbb{R}^p$ denotes a function describing how the multiple kernels are combined, and the variable η represents the kernel combination function parameter. Essentially, the kernel functions and their corresponding parameters are predetermined before multikernel model training, typically through cross-validation on different kernel types for each information source. The performance of the MKSVM model is influenced by the choice of kernel used to represent each information source; hence, allowing the algorithm to select from a set of kernels can lead to an optimal solution. Another crucial decision for MKSVM is the kernel combination function, f_η. In recent years, several methods have been proposed for combining kernels, however, the weighted sum of kernels remains the most widely used for AD-related tasks (Gonen et al., 2011; Zhang et al., 2011; Liu et al., 2014). Consequently, we employed the same linear combination function for merging kernels as follows:

$$k_\eta(x_i, x_j) = \sum_{m=1}^{M} \eta_m * k_m\left(x_i^m, x_j^m\right)$$
$$\sum_{m=1}^{M} \eta_m = 1, \quad \eta \in \boldsymbol{\eta_m} \ \mathbb{R}_+^M \qquad (5)$$
$$\boldsymbol{\eta_m} \geq \beta$$

where η_m represents the contribution weights of each kernel matrix, and β denotes the minimum contribution from each modality. Previous studies have indicated that when the weighted sum of the kernel is used, one modality tends to dominate, thereby reducing the contribution of the other modalities to almost zero, i.e., $\eta_m \rightarrow 0$. Therefore, we introduced a new regularization constraint, $\eta_m \geq \beta$ as shown in Eq (5) which sets a minimum contribution from each modality, where β denotes a tunable hyperparameter. This new constraint ensures that each modality is compelled to contribute to the model's performance. Furthermore, previous approaches have restricted kernels from each modality to be homogeneous, with the same hyperparameter values (Liu et al., 2014; Shi et al., 2022; Zhang et al., 2011). However, in multimodal datasets, each modality typically exhibits a different data distribution. Hence, we relaxed the constraints on the kernels, enabling the SVM algorithm to

Table 2
Notations and their corresponding definitions.

Notation	Meaning	Definitions
L^m	Graph Laplacian Matrix of m-th modality	$L^m = D^m - S^m$
S^m	Symmetric matrix of similarity between all samples in m-th modality	$S_{ij}^m = \begin{cases} e^{-\frac{||x_i^m - x_j^m||^2}{t}}, & x_j^m \in N(x_i^m) \ or \ x_i^m \in N(x_j^m) \\ 0, & otherwise \end{cases}$ **Where** S_{ij}^m is the similarity between i-th & j-th sample, and t is a hyperparameter
D^m	Diagonal square matrix with rows, i and column j of m-th modality	$D_{ii}^m = \sum_{J}^{N} S_{ij}^m$
$x_j^m \in N(x_i^m)$	x_j^m is among the k-nearest neighbors of x_i^m	Determine $N(x_i^m)$ using k-NN algorithm where k is a hyperparameter
$\| X \|_{2,1}$	$\ell_{2,1}$-norm of matrix M	$\sum_{i=1}^{N} \sqrt{\sum_{j=1}^{p} X_{ij}^2}$ for p features and N samples

select the optimal kernel from sets of predefined kernels for each input into the multikernel SVM. Similarly, we jointly optimized the kernel combination weights with the SVM parameters instead of specifying them as hyperparameters. The formulation of the regularized multi-kernel SVM is as follows:

$$\min_{\eta} \mathscr{J}(\eta)$$
$$s.t. \sum_{m=1}^{M} \eta_m = 1, \tag{6}$$
$$\eta_m \geq \beta$$

where the objective function, $\mathscr{J}(\eta)$ is defined in Eq (7), and α^* represents the optimal alpha value of the single kernel SVM problem defined in Eq (2) after replacing $k(x_i, x_j)$ with the $k_\eta(x_i, x_j)$ value from Eq (5).

$$\mathscr{J}(\eta) = -\frac{1}{2} \sum_{ij}^{N} \alpha_i^* \alpha_j^* y_i y_j \sum_{m=1}^{M} \eta_m k_m(x_i, x_j) + \sum_{i=1}^{N} \alpha_i^* \tag{7}$$

We employed the SimpleMKL (Rakotomamonjy et al.,2008) method to solve Eq (7) restricting the bounds of η_m to the range from β to $1 - \beta$ {see GitHub link for implementation details}.

For the multiclass MKL, Rakotomamonjy et al.,(2008) extend the SimpleMKL by introducing an objective function that jointly optimizes all the pairwise decision functions as shown in Eq (8).

$$\mathscr{J}(\eta) = \sum_{p \in \mathscr{P}} \mathscr{J}_p(\eta) \tag{8}$$

Where $\mathscr{P}=3$ in this case since we used a one-vs-rest SVM solver which creates three pairwise binary SVM with objective function denoted as, $\mathscr{J}_p(\eta)$.

2.4. Evaluation metric

Models are primarily evaluated using accuracy scores, which measure the ratio of correctly classified samples across all class labels. Additionally, the precision, recall, and F1 scores of the best model are reported. Precision scores help in assessing the accuracy of the positive predictions across all classes, while recall scores evaluate the ability of a model to identify all instances of each class within the dataset. The F1 score represents the harmonic mean of precision and recall. In multiclass cases, we used average scores weighted by the number of instances for each class.

2.5. Experiments

We created a new feature called "age_category," to categorize subjects into two groups based on whether their ages are greater or less than 76 years. The reference age of 76 years was selected because it approximately represents the median age of the subjects, effectively dividing them into nearly equal groups. Subsequently, we split the dataset into training and test sets in an 80%:20% ratio, respectively. The data splitting process was stratified based on the class labels, age_category, and sex. Cerebral atrophy, which could occur because of normal aging, can influence the performance of the model (Wyman et al., 2013; Lin et al., 2021). Hence, we used the age-stratified split recommended by (Wyman et al., 2013) to minimize any potential effect of normal aging on our results. For each experiment, we performed 10 or 20 replications using randomly generated seeds and reported the mean and standard deviation of the accuracy scores. Additionally, we determined the precision, recall scores, and F1 scores for the proposed model. All experiments were conducted using the 5-fold cross-validation approach. Our experiment initially involved only the MRI and PET modalities. Subsequently, we introduced the Apoe4, ADAS11, MPACCdigit, and ICV features, which we referred to as "NEW Features" or "NEW" for the convenience of readers. We trained all unimodal and concatenated

features on five candidate models: Logistic Regression (LOGREG), SVM, Random Forest (RF), XGBoost (XGB), and LightGBM. The highest mean accuracy across these models was reported. Following that, principal component analysis (PCA) was performed on the overall concatenated features. Moreover, multiple feature selection methods (RFE-SVM, MRMR-RF, MRMR-KS (Zhao et al., 2019), and Embedded XGBoost) were performed individually. Features selected by at least two models ($n\_suports \geq 2$) were used to retrain each candidate model. Table 3 shows the best accuracy scores for unimodal and multimodal features.

3. Results and discussions

Table 3 presents the model performance for the multiclass classification task on the independent test set across 10 replications. Symbol notations and their corresponding meanings are described in Table 4.

The individual MRI and PET features yielded the lowest accuracy scores of 0.5500±0.0546 and 0.5282±0.0566, respectively. However, the combined modality accuracy score became 0.5782±0.063069, greater than either of the individual scores. Multiclass classification poses challenges, particularly stemming from the heterogeneity in MCI samples (Singanamalli et al., 2017). Consequently, identifying other useful modalities may be equally crucial as the selection of algorithms for multiclass classification. The NEW features alone yielded a mean accuracy score of 0.7926±0.0475, suggesting a high predictive power relative to the combined MRI and PET modalities. Furthermore, the new features contain the Genetic (Apoe4), Cognitive (Adas11, MPACC digits), and ICV features, rendering them multimodal. The accuracy scores were further increased to 0.8282±0.0401 and 0.8282±0.0567, respectively, when the new features were concatenated with MRI or PET features. However, concatenating additional features, especially from neuroimaging, significantly increased the overall feature dimensions. In this case, MRI (68 features), PET (68 features), and NEW (4 features) increased the overall feature size to 140. Because of the small sample size dataset (194) employed in this study, models became susceptible to the curse of dimensionality, leading to potential performance degradation. Previous studies used PCA, feature selection, and multikernel SVM to overcome those challenges (Lin et al., 2021; Liu et al., 2015). Therefore, each candidate model was trained on the selected features from MRI, PET, and concatenated features, following the feature selection methodology outlined in Section 2.5 to establish a reasonable

Table 3
Multiclass classification test scores across 10 replications.

Feature Types	AD vs. MCI vs. CN
MRI	0.5500±0.0546
PET	0.5282±0.0566
MRI + PET$_{Concat}$	0.5782±0.0630
MRI + PET$_{MKSVM}$	0.5692±0.0449
NEW	0.7926±0.0475
NEW + PET$_{Concat}$	0.8282±0.0567
NEW + MRI$_{Concat}$	0.8282±0.0401
MRI + PET + NEW$_{Concat}$	0.7846±0.0756
MRI + PET + NEW$_{PCA}$	0.7759±0.0518
MRI + PET + NEW$_{MKSVM}$	0.7615±0.0554
MRI + PET + NEW$_{MKSVM\_Updated}$	0.7879±0.0589
LHSMRI + RHSPET + NEW$_{Concat}$	0.8295±0.0556
LHSMRI + RHSPET + NEW$_{MKSVM}$	0.7848±0.0540
LHSMRI + RHSPET + NEW$_{RMKSVM}$	0.8076±0.0530
LHSMRI + ALLPET + NEW$_{MKSVM}$	0.7794±0.0556
LHSMRI + ALLPET + NEW$_{RMKSVM}$	0.8076±0.0804
$\mathscr{H}_{LHSMRI+RHSPET}$ + **NEW**$_{RMKSVM}$	0.8230±0.0371
$\mathscr{H}_{LHSMRI+ALLPET}$ + **NEW**$_{RMKSVM}$	0.8282±0.0651
$\mathscr{H}_{ALLMRI+ALLPET}$ + **NEW**$_{RMKSVM}$	0.8333±0.0675
$\mathscr{H}_{ALLMRI+ALLPET}$ + **NEW**$_{RMKSVM}$ ($\beta = 0.05$)	**0.8487±0.0609**

**NEW: Apoe4, MPACC digits, Adas11, ICV **Best model is highlighted in bold letters.

$\beta = 0.2$ unless stated otherwise.

Table 4
Notations and their corresponding meanings.

Notation	Description
A	Single Modality Feature, A
$A + B_{Concat}$	Multimodal Features A and B are concatenated
$A + B_{MKSVM}$	Multimodal Features A and B using Multikernel SVM
$A + B_{PCA}$	Joint feature embeddings of A and B from Principal Component Analysis
$A + B_{RMKSVM}$	Multimodal Features A and B using our regularized Multikernel SVM
$\mathcal{H}_{A+B}$	Joint feature embeddings of A and B features using Algorithm 1 (see Appendix A)

baseline. We further performed PCA on the concatenated features, reducing the number of features from its initial 140 features to 60 principal components accounting for 96% of the variance. However, the accuracy score of the best model using PCA features is lower than that achieved with feature concatenation. With multikernel SVM, models are susceptible to performance degradation when modalities with highly unequal predictive powers are used, resulting in the contribution from lower predictive modality approaching zero. Experiment results show that $MRI + PET_{MKSVM}$ and $MRI + PET + NEW_{MKSVM}$ exhibited lower mean accuracy scores of 0.5692±0.0449 and 0.7615±0.0554 respectively compared to 0.5782±0.063069 and 0.7846±0.0756 when the features were concatenated. In this work, all reported accuracy scores from concatenated features represent the highest realized from all the candidate models considered, which may not necessarily be an SVM model. Nonetheless, our experiment results indicates that the commonly used multikernel SVM provided suboptimal accuracy scores. Neuroimaging features from MRI and PET contributed the most to the overall feature dimensions, however, they individually provided the least increase in accuracy scores compared to NEW features. Neuroimaging features typically exhibit high feature collinearity between features from the left-hand side (LHS) and right-hand side (RHS) of the brain. Therefore, we explored the predictive powers of these regional features to identify potential regional redundancies. Our experiments indicate that the LHS regional features from MRI (LHS acc: 0.5269 vs RHS acc: 0.5205) and the RHS features from PET (LHS acc: 0.4910 vs RHS acc: 0.5243) were the most predictive neuroimaging regions (see Fig. 2a and b).

Similarly, various combinations of regional features were employed for training candidate models as shown in Fig. 3a and b. The combination of features from the LHS MRI features + RHS PET Features or LHS MRI features + ALL PET Features produce the highest mean accuracy scores, with the latter slightly outperforming the former. Considering these findings, several potential methods were also evaluated using these combinations of regional features (Table 3).

To address these performance degradation problems associated with MKSVM, a regularized MKSVM was developed, as described in Section 2.3.2. Further experiments revealed that this regularized MKSVM (denoted as R$MKSVM$ in Table 3), with a minimum modality

contribution of $\beta = 0.2$ outperformed the previous MKSVM across all different feature combinations. This improvement led to increased accuracy scores from 0.7615±0.0554 to 0.7879±0.0589 for $MRI + PET + NEW$, from 0.7848±0.0540 to 0.8076±0.0530 for $LHSMRI + RHSPET + NEW$, and 0.7794±0.0556 to 0.8076±0.0804 for $LHSMRI + ALLPET + NEW$ feature types. The improvement is from the regularization effect of imposing a minimum contribution constraint on each modality, which facilitates improved model stability across replications. Finally, using the proposed framework in Fig. 1, we first created a joint feature embedding of size 60 from MRI and PET features employing Algorithm 1. Subsequently, we trained this embedding simultaneously with the NEW features using our regularized multikernel SVM. The embedding was derived by setting r=1.5, k = 8, and ξ=0.5, selected through a heuristic approach. Other hyperparameters, such as t, λ_1 and λ_2 were fine-tuned using Hyperopt (Bergstra et al., 2013). We chose the same embedding length as that of PCA to facilitate easy comparison. Experiment results show that the classification performance denoted as $\mathcal{H}_{ALLMRI+ALLPET} + NEW_{RMKSVM}$ with a mean accuracy score of 0.8333 ± 0.0675, outperforms others across all classification tasks. The distribution of F1, Precision, and Recall scores across the experiment replications is displayed in Fig. 4. Note that the embedding sizes of 40 and 50 were used for $\mathcal{H}_{LHSMRI+RHSPET}$ & $\mathcal{H}_{LHSMRI+ALLPET}$ experiments, respectively. Following the tuning of the regularization parameter (Fig. 6), a higher accuracy score of 0.8487±0.0609 was achieved when $\beta = 0.05$.

While our framework is primarily for multiclass classification, we also evaluated its performance in binary classification tasks. Table 5 shows the model performance on AD vs. CN, AD vs. MCI, and CN vs. MCI, on the independent test set.

The results show that models trained solely on MRI or PET features independently have lower accuracies compared to their combination across all binary classification tasks, consistent with observation in the multiclass scenario. However, upon the inclusion of NEW features, mean accuracy scores significantly improved across all tasks. Significantly, these NEW features alone outperform combined MRI and PET features (AD vs. CN: 0.9988 vs. 0.8826, AD vs. MCI: 0.9000 vs. 0.7555, CN vs. MCI: 0.8655 vs. 0.6620 for NEW vs. MRI+PET). This implies that the selection of biomarkers is significant for preclinical AD diagnosis, aligning with previous studies (Weiner et al., 2012, 2015). When these two different multimodal features are concatenated (represented as $MRI + PET + NEW_{Concat}$), the mean accuracy decreased for CN vs. MCI (accuracy score: 0.8655 to 0.8241) and slightly for AD vs. CN (accuracy score: 0.9988 to 0.9913) compared to the NEW features. This reduction may be attributed to the significant increase in the overall feature dimension, leading to model instability and performance degradation across replications. When PCA features are used, model accuracy declined across all experimental categories (denoted as $MRI + PET + NEW_{PCA}$ in Table 5). Comparatively, MKSVM demonstrated slight improvement over direct feature concatenation in the AD vs. CN (accuracy score: 0.9913 to 0.9956) and CN vs. MCI (accuracy score: 0.8241
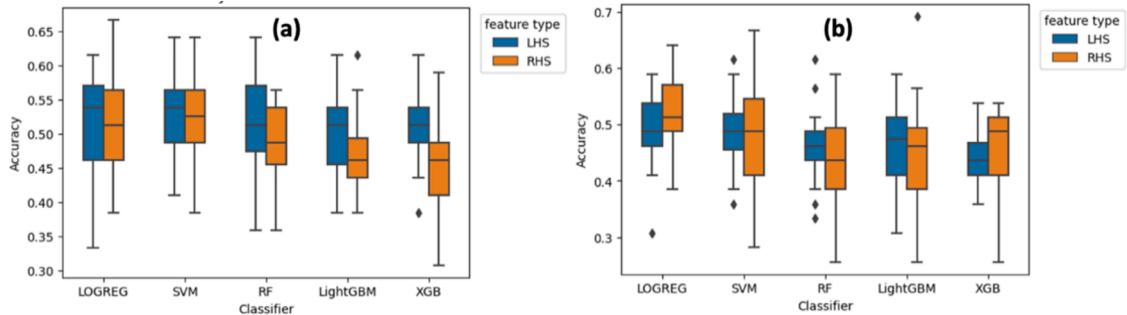
Fig. 2. The distribution of accuracy scores of candidate models across 20 replications when trained on the regional MRI and PET features. Boxes (a) and (b) depict the MRI and PET results, respectively.
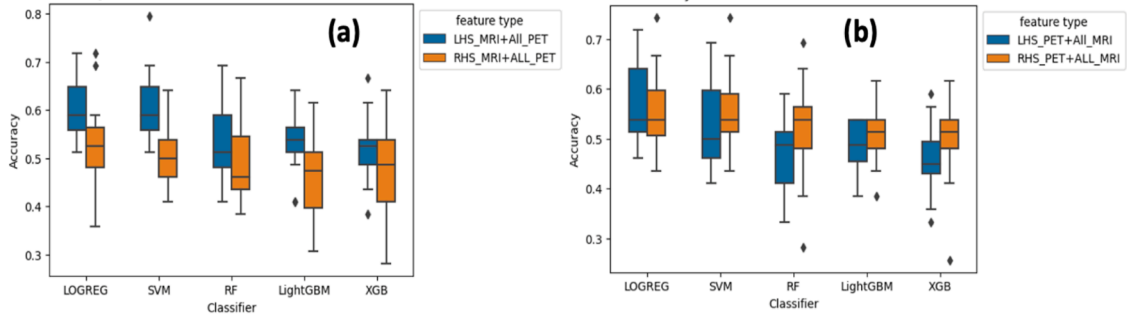
Fig. 3. Distribution of accuracy scores of five candidate models across 20 replications when trained on combinations of different regional features from MRI and PET.
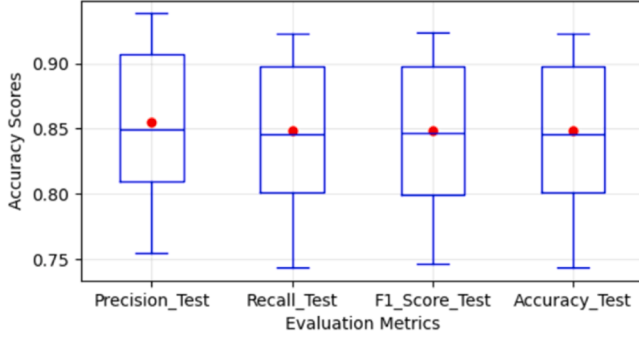

Fig. 4. The distribution of accuracy, precision, recall, and F1 scores for our proposed classification framework on CN vs. MCI vs. AD classification depicted across 10 replications. The red point indicates mean values.

Table 5
Mean test accuracy scores of different binary classification tasks across 10 replications.

Features	AD vs. CN	AD vs. MCI	CN vs. MCI
MRI	0.8303±0.0935	0.6761±0.0947	0.6135±0.0665
PET	0.7897±0.0348	0.6698±0.1075	0.6004±0.0529
$MRI + PET_{Concat}$	0.8478±0.0656	0.7555±0.0822	0.6620±0.0534
$MRI + PET_{MKSVM}$	0.8826±0.0460	0.7296±0.0553	0.6517±0.0954
NEW	0.9988±0.001	0.9000±0.0351	0.8655±0.0770
$MRI + PET + NEW_{Concat}$	0.9913±0.0183	0.9000±0.0351	0.8241±0.0638
$MRI + PET + NEW_{PCA}$	0.9826±0.0224	0.8888±0.0907	0.8103±0.0675
$MRI + PET + NEW_{MKSVM}$	0.9956±0.0137	0.8740±0.0475	0.85175±0.0564
$MRI + PET + New_{RMKSVM}$	1.0000±0.0000	0.9185±0.0517	0.8620±0.0363
$\mathscr{H}_{MRI+PET} + NEW_{RMKSVM}$	**1.0000±0.0000**	**0.9296±0.0507**	**0.8689±0.0422**

the CN vs. MCI case, which is slightly lower than that reported by Kim and Lee (2018) (accuracy score: 87.09) by 0.2%. While we reported the mean score across 10 replications of random data splits, Kim and Lee (2018) reported the accuracy score based on a single experiment without replication. During the CN vs. MCI experiment replications, our model achieved a peak accuracy of 91%. Consequently, we contend that our model accuracy surpasses that of previous methods across all experiment categories.

3.1. Effect of the regularization parameter, β on multiclass classification performance

The selection of β-value significantly influences the overall classification performance of our proposed framework. Therefore, in this section, we show how the mean accuracy scores vary across different values of β (Fig. 5). Table 3 shows that the NEW feature set has more predictive power than MRI and PET combined. Increasing the $\beta - value$ implies imposing more constraint on the contribution of the MRI+PET modality to the RMKSVM performance, consequently decreasing the influence of the NEW features, which possess a higher predictive power. The results suggest that the best accuracy score (0.8487) is realized when $\beta = 0.05$ while the second-highest score (0.8461) is realized when $\beta = 0.14$, both have the same median scores of 0.8461. The latter ($\beta = 0.14$) demonstrates greater stability in terms of accuracy scores across replications compared to the former ($\beta = 0.05$), with a standard deviation of 0.0592 and 0.0609, respectively. Significantly, setting $\beta = 0.23$ resulted in the lowest mean accuracy score of 0.7923; however, this lower score was primarily influenced by an outlier accuracy score of 0.4102 observed during one experimental replication.

Conversely, setting $\beta = 0$ (that is, without regularization) produced the second-lowest mean accuracy score and the lowest median accuracy score, with the potential to be the lowest overall if the outlier score for $\beta = 0.23$ is excluded. Overall, the figure illustrates a pattern of accuracy increasing followed by convergence and then decreasing as the $\beta - value$ increases.

### 3.2. Topological meaning of the joint MRI and PET embeddings

To understand the topological meaning of the joint MRI and PET embedding, we apply the permutation feature importance method on the joint MRI and PET embedding to select the top 3 most important features (Altmann et al., 2010). Since our embedding method in Section 2.3 is based on the least square component analysis, each principal component is an approximate orthogonal combination of features from the various brain regions (Zhang et al., 2015). Then, we determine the top 5 most important brain regions based on the magnitude of their correlation with each of these three feature embeddings. When both left and right neuroimaging features of a particular brain region are selected from the same modality and typically show correlation, the side exhibiting a lower correlation is discarded. We then substitute it with features from the next region that demonstrate stronger correlations.

to 0.8518) classification tasks. However, the accuracy score decreased from 0.9000 to 0.8740 in the AD vs. MCI tasks. By using the RMKSVM and setting $\beta = 0.2$ (denoted as $MRI + PET + New_{RMKSVM}$), the model accuracy increased from 0.9956 to 1.0000 for AD vs. CN, 0.8740 to 0.9185 for AD vs. MCI, and 0.85175 to 0.8620 for CN vs. MCI task. Setting $\beta = 0.2$ outperforms $\beta = 0.05$ in the binary classification categories. Finally, using our classification framework, we generated a joint feature embedding of size 60 from MRI and PET data, the same length as that of the multiclass feature set. Subsequently, we performed simultaneous training of this feature embedding with the NEW features using RMKSVM. This method produces accuracy scores of 1.0000±0.0000 for AD vs CN, 0.9296±0.0507 for AD vs MCI, and 0.8689±0.0422 for CN vs MCI tasks (denoted as $\mathscr{H}_{MRI+PET} + NEW_{RMKSVM}$), demonstrating superior results across all binary classification tasks. Table 6 shows that we achieved SOTA results across the different classification tasks, except in

Table 6
Performance comparison of previous multimodal binary and multiclass frameworks on the ADNI dataset.

Articles	Samples	AD vs. CN	AD vs. MCI	CN vs. MCI	AD vs. MCI vs. CN	Some or all NEW features present?
Zhang et al., (2011)	51 AD, 99 MCI, 52 CN	93.2	-	76.4	-	NO
Gray et al., (2013)	37 AD, 75MCI, 35 CN	89.0 ± 0.7	-	74.6 ± 0.8	-	NO
Liu et al., (2014)	51 AD, 99 MCI, 52 CN	94.37	-	78.80	-	NO
Suk et al., (2014)	51 AD, 99 MCI, 52 CN	96.18 ± 6.58	73.21	81.45		NO
Jie et al., (2015)	51 AD, 99 MCI, 52 CN	95.38	-	82.99	-	NO
Xu et al., (2015)	113 AD, 110 MCI, 117 CN	94.8	-	74.5	-	NO
Liu et al., (2015)	85 AD, 102 ncMCI, 67cMCI, 77 CN	92.89 ± 6.17		82.10 ± 4.91	53.79 ± 4.7	NO
Zu et al., (2016)	51 AD, 99 MCI, 52 CN	95.95	-	80.26	-	NO
Ye et al., (2016)	51 AD, 99 MCI, 52 CN	95.92	-	82.13	-	NO
Li et al., (2017)	113 AD, 110 MCI, 117 CN	97.36	-	77.66	-	NO
Tong et al., (2017)	37 AD, 75 MCI, 35 CN	91.8	-	79.5	60.2	YES
Singanamalli et al., (2017)	52 AD, 71 MCI, 26 CN	87	-	86	75.16**	YES
Wang et al., (2018)	99 MCI, 52 CN	-		75.5	-	NO
Kim and Lee, (2018)	51 AD, 99 MCI, 52 CN	97.12	-	87.09	-	NO
Li et al., (2018)	113 AD, 110 MCI, 117 CN	98.18	74.47	78.5	-	NO
Hao et al., (2020)	ADNI1–51 AD, 99 MCI, 52 CN	97.60± 5.03	-	84.47 ± 6.83	-	NO
Akramifard et al., (2020)	156 AD, 338 MCI, 211 CN	98.81	81.40	81.61	-	NO
Fang et al., (2020)	162 AD, 493 MCI, 251 CN	-	87.69	-	67.69	NO
Lin et al., (2020)	102 AD, 315 MCI, 200 CN		84.7	-		YES
Lin et al., (2021)	105 AD, 441 MCI, 200 CN	93.4	-		66.7± 1.4	YES
Ning et al., (2021)	ADNI1–93 AD, 200 MCI, 99 CN	96.9 ± 3.5	–	86.6± 3.4	–	NO
Shi et al., (2022)	51 AD, 99 MCI, 52 CN	96.76± 0.05	-	80.73± 0.09	-	NO
Ban et al.,(2023)	144 AD, 203 MCI, 181 CN	98.78	86.47	78.15	-	NO
OURS	53 AD, 81 MCI, 60 CN	**100.00±0.00**	**92.96±5.07**	**86.89±4.22**	**84.87±6.09**	

** Accuracy score from sequential multiclass classification method. The accuracy score was calculated using the weighted sum of the accuracy score reported for each class. The last column on the right indicates whether the modalities used contain any or all of the features in NEW.
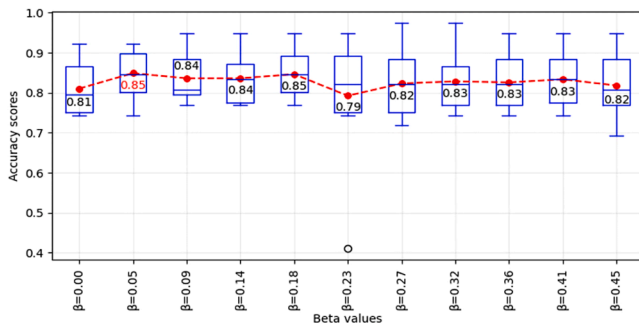
Fig. 5. A boxplot showing the distribution of the multiclass classification accuracy scores across various β values. A linear red broken line is drawn connecting the mean point for each box. The number annotated within each box represents the mean score, with the highest score in red.
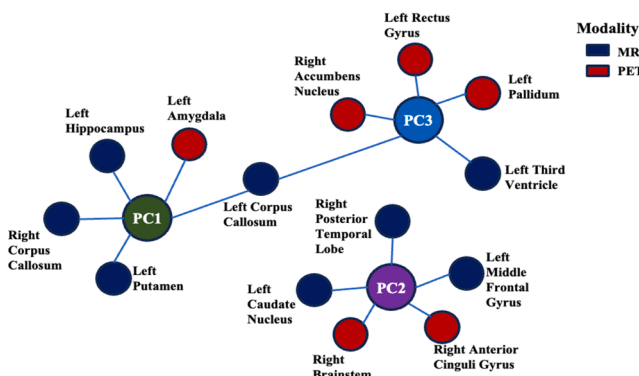
Fig. 6. Top 3 most important feature embeddings and their corresponding top 5 most correlated brain regions. Red-colored circles are features from PET modality, while blue circles are from MRI. The first principal component (PC1) is predominantly influenced by MRI features, whereas the third principal component (PC3) shows a stronger association with PET features.

Fig. 6 shows the top 3 most important neuroimaging feature embeddings and their corresponding top 5 most correlated brain regions.

In Fig. 6, the top 5 brain regions influencing PC1 (the most important embedding) are the left hippocampus, Left Amygdala, Right Corpus Callosum, Left Putamen, and Left Corpus Callosum. Four of the five features are from the MRI modality. However, for PC3, three features are from the PET modality and two from MRI. The top 5 useful regions for PC3 embedding are the Left Pallidum, Left Rectus Gyrus, Right Accumbens Nucleus, Left Third Ventricle, and Left Corpus Callosum, the common feature between PC1 and PC3. PC2 contains three MRI features (Right Posterior Temporal Lobe, Left Middle Frontal Gyrus, and Left Caudate Nucleus), and two PET features (Right Brainstem & Right Anterior Cinguli Gyrus). The Hippocampus, Amygdala, and Temporal Lobes are some of the regions known to indicate the presence of Alzheimer's disease (Weiner et al., 2012, 2015). These features are also captured in the PC1 and PC2 embeddings.

Furthermore, the model also detects some regions that are not commonly mentioned in previous studies such as the Brainstem, Rectus Gyrus, and so on. We believe that while these features may not hold much significance individually, they become important due to the interaction effects between multiple features within each embedding. Interpreting the topological representation of the multiclass classification task is challenging since feature extraction typically precedes the classification task, hence, making it difficult to understand the meanings of each embedding (Liu et al., 2015; Singanamalli et al., 2017; Tong et al., 2017; Fang et al., 2020; Lin et al., 2021). We will elaborate on this in our future study.

4. Conclusion

This study introduces a novel classification framework aimed at distinguishing AD, MCI, and CN samples in one shot. Previous works have highlighted the challenge of this multiclass classification task owing to the heterogeneity observed in MCI samples, with some demonstrating similarities to CN while others align more closely with AD within the feature space. This framework starts by extracting a low-dimensional joint embedding from MRI and PET features using the ensemble manifold regularized sparse low-rank approximation algorithm. It then combines this embedding with NEW features (Apoe4,

MPACC digits, Adas11, and ICV) using a regularized multikernel SVM. However, conventional multikernel SVMs tend to degrade in performance when trained on features with uneven predictive power, reducing the contribution from less informative sources to near zero. Therefore, we introduced a regularized MKSVM with a minimum modality contribution constraint to address these challenges. Our framework outperformed existing methods in various multiclass and binary classification tasks, achieving SOTA results across all categories. This study has a limitation by solely focusing on the ADNI1 dataset cohorts and using neuroimaging features extracted using the anatomical atlas proposed by Hammers et al. (2003). This atlas is widely used for similar problems; however, further investigation should explore features from alternative atlases, as they may produce features of varying dimensions. Additionally, while ADNI1 is a commonly used data cohort for such problems, future research should consider studying other cross-center Alzheimer datasets to enhance the robustness of this framework. This issue will be addressed in future works.

Ethics statement

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As per ADNI protocols, all procedures involving human participants in this study were approved by the institutional review board of each participating institution and the United States Code of Federal Regulations. We obtained approval from the ADNI Data and Publications Committee before the initiation and submission of this study.

CRediT authorship contribution statement

Oyekanmi O. Olatunde: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Kehinde S. Oyetunde:** Visualization, Validation, Investigation, Data curation, Conceptualization. **Jihun Han:** Visualization, Validation, Methodology, Formal analysis, Data curation. **Mohammad T. Khasawneh:** Writing – review & editing, Supervision, Investigation, Conceptualization. **Hyunsoo Yoon:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Appendix A

The following describes the formulation process of the EMR-SLRA algorithm. We provided this formulation here because the one in the original paper by (Zhang et al., 2015) couldn't be reproduced. Hence, we made a slight change to the algorithm. This is to encourage reproducibility of results. The alternate optimization steps are the following;

Step 1: Set Y, β, and $\overline{X}$ as constants, then optimize U

The objective function in Eq (1) becomes.

$$\min_{U} \; ||\overline{X} - UY||^2$$
$$s.t. \, U^T U = I \tag{A.1}$$

Our objective function was then combined with the constraint by introducing a LaGrange multiplier, ρ Such that,

$$L(U, \rho) = ||\overline{X} - UY||^2 + \rho(U^T U - I) = tr(\overline{X}^T \overline{X}) + tr(Y^T Y) - 2tr(U^T \overline{X} Y^T) + \rho(U^T U - I) \tag{A.2}$$

Taking the partial derivatives of L w.r.t. U.

$$\frac{\partial L}{\partial U} = -2\overline{X}Y^T + 2U\rho = 0$$
$$\Rightarrow U = \overline{X} \, Y^T \rho^{-1}$$
$$\Rightarrow U = Z\rho^{-1}, \; where \; Z = \overline{X} \, Y^T \tag{A.3}$$

Such that constraint $U^T U = I$ becomes,

$$(Z\rho^{-1})^{\mathrm{T}}(Z\rho^{-1}) = I$$
$$\Rightarrow Z^T Z (\rho^{-1})^T \rho^{-1} = I$$
$$\Rightarrow \rho = (Z^T Z)^{1/2} \tag{A.4}$$

Substitute Eq (4) into Eq (3), we have

$$U = Z\rho^{-1}$$
$$\Rightarrow U = Z(Z^T Z)^{-1/2} \tag{A.5}$$

Through singular value decomposition, we can factorize matrix Z, such that $Z = GDV^T$. Following that, Eq (5) becomes.

$$U = GDV^T \left((GDV^T)^T GDV^T \right)^{-1/2}$$
$$\Rightarrow U = GDV^T * (VDG^T * GDV^T)^{-1/2}$$
$$Since \, G^T G = \, U^T U = 1, \, since \, they \, are \, orthogonal \, matrices, \, then$$
$$\Rightarrow U = \, GV^T \tag{A.6}$$

Step 2: Set U, β, and, $\overline{X}$ as constants, then optimize Y

To optimize the embedding, Y, the objective function was reduced to Eq (7), as seen below.

$$\min_Y ||\overline{X} - UY||^2 + \lambda_1 tr(YLY^T)$$
$$where \, L = \sum_{m=1}^{M} (\beta_m)^r L^m \tag{A.7}$$

Taking the derivative of Eq (7) w.r.t. Y and set to zero, we have

$$2Y - 2U^T\overline{X} + 2\lambda_1 YL = 0$$
$$\Rightarrow Y = U^T\overline{X}\gamma^{-1} \tag{A.8}$$

where $\gamma = I + \lambda_1 L$ and I is the identity matrix

Step 3: Set U, Y, and $\overline{X}$ as constants, then optimize β

To optimize β, we considered a reduced minimization problem, as defined below

$$\min_{\beta} \sum_{m=1}^{M} (\beta_m)^r tr(YL^mY^T)$$
$$s.t. \sum_{m=1}^{M} \beta_m = 1$$
$$\beta_m > 0 \tag{A.9}$$

Like Eq (2), we merged the objective function and constraint in Eq (9) by introducing a LaGrange multiplier, α as shown below:

$$L(\beta_m, \alpha) = \sum_{m=1}^{M} (\beta_m)^r tr(YL^mY^T) - \alpha \left(\sum_{m=1}^{M} \beta_m - 1 \right) \tag{A.10}$$

By taking the partial derivative of $L(\beta_m, \alpha)$ w.r.t β_m and setting the value to zero, we derived

$$\beta_m = \left(\frac{\alpha}{r\phi^m} \right)^{\frac{1}{r-1}} \tag{A.11}$$
$$where \, \phi^m = tr(YL^mY^T)$$

since $\sum_{m=1}^{M} \beta_m = 1$, then we normalize β_m such that

$$\beta_m = \left(\frac{\alpha}{r\phi^m} \right)^{\frac{1}{r-1}} \Big/ \sum_{m=1}^{M} \left(\frac{\alpha}{r\phi^m} \right)^{\frac{1}{r-1}}$$
$$\Rightarrow \beta_m = (r\phi^m)^{\frac{1}{1-r}} \Big/ \sum_{m=1}^{M} (r\phi^m)^{\frac{1}{1-r}} \tag{A.12}$$

Step 4: Set U, Y, and β as constant, then optimize $\overline{X}$

To optimize the ideal feature matrix, $\overline{X}$, we reduced the objective function to Eq (13).

$$\min_{\overline{x}} ||\overline{X} - UY||^2 + \lambda_2 ||\overline{X} - X||_{2,1} \tag{A.13}$$

The $\ell_{2,1}$-norm minimization problem was further simplified for fast convergence to Eq (14) (Nie et al., 2010) such that we obtained;

$$\min_{\overline{X}} \ ||\overline{X} - UY||^2 + \lambda_2 \ tr\left((\overline{X} - X)^T D(\overline{X} - X)\right) \tag{A.14}$$

Where D is a diagonal matrix defined as follows

$$D = \begin{pmatrix} \dfrac{1}{2\left|\left|(\overline{X} - X)^i\right|\right|_2} & & \\ & \ddots & \\ & & \dfrac{1}{2\left|\left|(\overline{X} - X)^i\right|\right|_2} \end{pmatrix} \tag{A.15}$$

By taking the derivative of Eq (14) w.r.t. $\overline{X}$, and setting the value to zero, we obtained

$$\Rightarrow 2\overline{X} - 2U^T Y^T + 2\lambda_2 \ D \ (\overline{X} - X) = 0$$
$$\Rightarrow \overline{X}(1 + \lambda_2 D) = U^T Y^T + \lambda_2 DX \tag{A.16}$$
$$\Rightarrow \overline{X} = \frac{U^T Y^T + \lambda_2 \ DX}{(I + \lambda_2 \ D)}$$

Algorithm 1

Parameters: Given $X^m = \{x_1^m, x_2^m, \ldots, x_N^m\}$, $x_i^m \in \mathbb{R}^p \ \forall m = 1, 2 \ldots, M$, construct a multimodal feature matrix, $X = [X^1, X^2, \ldots, X^m] \in \mathbb{R}^{pMxN}$; Embedding dimension, d; Regularization parameters λ_1 & λ_2 ; nearest neighbor, **k**; scaling parameter, **r**; Similarity scaling parameter, **t**; Number of iterations, **Iter**, and threshold, **ξ**.
Initialize:
    Set random_seed value for reproducibility
    Joint Feature Embedding, $Y_i \in \mathbb{R}^{d \ x \ N}$
    Ideal Feature Matrix, $\overline{X}_i = X + \text{random} \ (pM, N)$
    Laplacian Combination Weight, $\beta_m = \left[\dfrac{1}{M}\right] * M$
Diagonal Weight Matrix, D_i for noise reduction term using Eq (15)
Compute S^m from $G^m = (V^m, E^m)$, then L^m based on Table 2
For I = 1 to Iter
    Determine $Z_i = \overline{X}_i Y_i^T$
    Perform SVD on Z_i such that $G_i, \sum_i, V_i = svd \ (Z_i)$
    Calculate $U_i = G_i V_i^T$
    Calculate Obj_i using Eq (1)
Set $L_i = \sum_{m=1}^{M} (\beta_m)_i^r L^m$, $\gamma = I_N + \lambda_1 L_i$, then update $Y_{i+1} = U_i^T \overline{X}_i \gamma^{-1}$ where $I_N \in \mathbb{R}^N$ is a column vector whose elements are 1.
    Set $\phi_i^m = \text{tr}(Y_i L^m Y_i^T)$, then update $(\beta_m)_{i+1} = (r\phi_i^m)^{\frac{1}{1-r}} / \sum_{m=1}^{M} (r\phi_i^m)^{\frac{1}{1-r}}$
    Update $\overline{X}_{i+1} = \dfrac{U_i^T Y_i + \lambda_2 \ D_i X}{(I_{pM} + \lambda_2 \ D_i)}$, then D_{i+1} by substituting $\overline{X}_{i+1}$ in Eq (15)
    Update Obj_{i+1} using Eq (1), break iteration if $|Obj_{i+1} - Obj_i| < \xi$
Endfor
Return multimodal feature embedding, Y and Projection matrix, U

Data availability

No data was used for the research described in the article.

References

Akramifard, H., Balafar, M., Razavi, S., Ramli, A.R., 2020. Emphasis learning, features repetition in width instead of length to improve classification performance: case study—Alzheimer's disease diagnosis. Sensors 20, 941. https://doi.org/10.3390/s20030941.

Altmann, A., Toloşi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. Bioinformatics 26, 1340–1347. https://doi.org/10.1093/bioinformatics/btq134.

2023 Alzheimer's disease facts and figures. Alzheimer. Dement. 19, 2023, 1598–1695. https://doi.org/10.1002/alz.13016.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. NeuroImage 38, 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007.

Ban, Y., Lao, H., Li, B., Su, W., Zhang, X., 2023. Diagnosis of Alzheimer's disease using hypergraph p-Laplacian regularized multi-task feature learning. J. Biomed. Inform. 140, 104326. https://doi.org/10.1016/j.jbi.2023.104326.

Bateman, R.J., Xiong, C., Benzinger, T.L.S., Fagan, A.M., Goate, A., Fox, N.C., Marcus, D. S., Cairns, N.J., Xie, X., Blazey, T.M., Holtzman, D.M., Santacruz, A., Buckles, V., Oliver, A., Moulder, K., Aisen, P.S., Ghetti, B., Klunk, W.E., McDade, E., Martins, R. N., Masters, C.L., Mayeux, R., Ringman, J.M., Rossor, M.N., Schofield, P.R., Sperling, R.A., Salloway, S., Morris, J.C., 2012. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N. Engl. J. Med. 367, 795–804. https://doi.org/10.1056/NEJMoa1202753.

Belkin, M., Niyogi, P., 2003. Laplacian Eigenmaps for dimensionality reduction and data representation. Neur. Comput. 15, 1373–1396. https://doi.org/10.1162/089976603321780317.

Bergstra, J., Yamins, D., Cox, D.D. (2013) Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. TProc. of the 30th International Conference on Machine Learning (ICML 2013), June 2013, pp. I-115 to I-23.

Caminiti, S.P., Ballarini, T., Sala, A., Cerami, C., Presotto, L., Santangelo, R., Fallanca, F., Vanoli, E.G., Gianolli, L., Iannaccone, S., Magnani, G., Perani, D., Parnetti, L., Eusebi, P., Frisoni, G., Nobili, F., Picco, A., Scarpini, E., 2018. FDG-PET and CSF biomarker accuracy in prediction of conversion to different dementias in a large

multicentre MCI cohort. NeuroImage Clin. 18, 167–177. https://doi.org/10.1016/j.nicl.2018.01.019.

Chang, M., Brainerd, C.J., 2022. Predicting conversion from mild cognitive impairment to Alzheimer's disease with multimodal latent factors. J. Clin. Exp. Neuropsychol. 44, 316–335. https://doi.org/10.1080/13803395.2022.2115015.

(Param) Chen, Z.S., Kulkarni, P, Galatzer-Levy, I.R., Bigio, B., Nasca, C., Zhang, Y., 2022. Modern views of machine learning for precision psychiatry. Patterns 3, 100602. https://doi.org/10.1016/j.patter.2022.100602.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018.

De La Torre, F., 2012. A least-squares framework for component analysis. IEEE Trans. Pattern Anal. Mach. Intell. 34, 1041–1055. https://doi.org/10.1109/TPAMI.2011.184.

Donohue, M.C., Sperling, R.A., Salmon, D.P., Rentz, D.M., Raman, R., Thomas, R.G., Weiner, M., Aisen, P.S., 2014. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. JAMA Neurol. 71, 961–970. https://doi.org/10.1001/jamaneurol.2014.803.

Fang, C., Li, C., Forouzannezhad, P., Cabrerizo, M., Curiel, R.E., Loewenstein, D., Duara, R., Adjouadi, M., 2020. Gaussian discriminative component analysis for early detection of Alzheimer's disease: a supervised dimensionality reduction algorithm. J. Neurosci. Methods 344, 108856. https://doi.org/10.1016/j.jneumeth.2020.108856.

Fleisher, A.S., Donohue, M., Chen, K., Brewer, J.B., Aisen, P.S., 2009. Applications of neuroimaging to disease-modification trials in Alzheimer's disease. Behav. Neurol. 21, 129–136. https://doi.org/10.1155/2009/836437.

Gaser, C., Dahnke, R., Thompson, P.M., Kurth, F., Luders, E., 2022. CAT – a computational anatomy toolbox for the analysis of structural MRI data (preprint). Neuroscience. https://doi.org/10.1101/2022.06.11.495736.

Golde, T.E., 2022. Alzheimer's disease – the journey of a healthy brain into organ failure. Mol. Neurodegener. 17, 18. https://doi.org/10.1186/s13024-022-00523-1.

Gonen, M., Alpaydın, E., Tr, B.E., Tr, B.E., 2011. Multiple kernel learning algorithms. J. Mach. Learn. Res. 12, 2211–2268.

Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., Poline, J.-B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J.A., Varoquaux, G., Poldrack, R.A., 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci. Data 3, 160044. https://doi.org/10.1038/sdata.2016.44.

Gousias, I.S., Rueckert, D., Heckemann, R.A., Dyet, L.E., Boardman, J.P., Edwards, A.D., Hammers, A., 2008. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. NeuroImage 40, 672–684. https://doi.org/10.1016/j.neuroimage.2007.11.034.

Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D., 2013. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. NeuroImage 65, 167–175. https://doi.org/10.1016/j.neuroimage.2012.09.065.

Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum. Brain Mapp. 19, 224–247. https://doi.org/10.1002/hbm.10123.

Hao, X., Bao, Y., Guo, Y., Yu, M., Zhang, D., Risacher, S.L., Saykin, A.J., Yao, X., Shen, L., 2020. Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease. Med. Image Anal. 60, 101625. https://doi.org/10.1016/j.media.2019.101625.

How Is Alzheimer's Disease Treated? [WWW Document], 2023. Natl. Inst. Aging. URL https://www.nia.nih.gov/health/alzheimers-treatment/how-alzheimers-disease-treated (accessed 12.1.23).

Hsu, W.-C., Denq, C., Chen, S.-S., 2013. A diagnostic methodology for Alzheimer's disease. J. Clin. Bioinforma. 3, 9. https://doi.org/10.1186/2043-9113-3-9.

Jena, A., Renjen, P.N., Taneja, S., Gambhir, A., Negi, P., 2015. Integrated 18F-fluoro-deoxyglucose positron emission tomography magnetic resonance imaging (18F-FDG PET/MRI), a multimodality approach for comprehensive evaluation of dementia patients: A pictorial essay. Indian J. Radiol. Imaging 25, 342–352. https://doi.org/10.4103/0971-3026.169449.

Jie, B., Zhang, D., Cheng, B., Shen, D., 2015. Manifold Regularized Multitask Feature Learning for Multimodality Disease Classification: Manifold Regularized Multitask Feature Learning. Hum. Brain Mapp. 36, 489–507. https://doi.org/10.1002/hbm.22642.

Jitsuishi, T., Yamaguchi, A., 2022. Searching for optimal machine learning model to classify mild cognitive impairment (MCI) subtypes using multimodal MRI data. Sci. Rep. 12, 4284. https://doi.org/10.1038/s41598-022-08231-y.

Joliot, M., Jobard, G., Naveau, M., Delcroix, N., Petit, L., Zago, L., Crivello, F., Mellet, E., Mazoyer, B., Tzourio-Mazoyer, N., 2015. AICHA: an atlas of intrinsic connectivity of homotopic areas. J. Neurosci. Methods 254 (46), 59. https://doi.org/10.1016/j.jneumeth.2015.07.013.

Kim, J., Lee, B., 2018. Identification of Alzheimer's disease and mild cognitive impairment using multimodal sparse hierarchical extreme learning machine. Hum. Brain Mapp. 39, 3728–3741. https://doi.org/10.1002/hbm.24207.

Klunk, W.E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D.P., Bergström, M., Savitcheva, I., Huang, G.-F., Estrada, S., Ausén, B., Debnath, M.L., Barletta, J., Price, J.C., Sandell, J., Lopresti, B.J., Wall, A., Koivisto, P., Antoni, G., Mathis, C.A., Långström, B., 2004. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann. Neurol. 55, 306–319. https://doi.org/10.1002/ana.20009.

Lei, B., Yang, P., Zhuo, Y., Zhou, F., Ni, D., Chen, S., Xiao, X., Wang, T., 2019. Neuroimaging retrieval via adaptive ensemble manifold learning for brain disease

diagnosis. IEEE J. Biomed. Health Inform. 23, 1661–1673. https://doi.org/10.1109/JBHI.2018.2872581.

Li, Q., Wu, X., Xu, L., Chen, K., Yao, L., 2018. Classification of Alzheimer's disease, mildx cognitive impairment, and cognitively unimpaired individuals using multi-feature kernel discriminant dictionary learning. Front. Comput. Neurosci. 11, 117. https://doi.org/10.3389/fncom.2017.00117.

Lin, W., Gao, Q., Yuan, J., Chen, Z., Feng, C., Chen, W., Du, M., Tong, T., 2020. Predicting Alzheimer's Disease Conversion From Mild Cognitive Impairment Using an Extreme Learning Machine-Based Grading Method With Multimodal Data. Front. Aging Neurosci. 12, 77. https://doi.org/10.3389/fnagi.2020.00077.

Lin, W., Gao, Q., Du, M., Chen, W., Tong, T., 2021. Multiclass diagnosis of stages of Alzheimer's disease using linear discriminant analysis scoring for multimodal data. Comput. Biol. Med. 134, 104478. https://doi.org/10.1016/j.compbiomed.2021.104478.

Liu, F., Wee, C.-Y., Chen, H., Shen, D., 2014. Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification. NeuroImage 84, 466–475. https://doi.org/10.1016/j.neuroimage.2013.09.015.

Liu, Siqi, Liu, Sidong, Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M.J., 2015. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Trans. Biomed. Eng. 62, 1132–1140. https://doi.org/10.1109/TBME.2014.2372011.

Nan, F., Li, S., Wang, J., Tang, Y., Qi, J., Zhou, M., Zhao, Z., Yang, Y., Yang, P., 2022. A multi-classification accessment framework for reproducible evaluation of multimodal learning in Alzheimer's disease. IEEE/ACM Trans. Comput. Biol. Bioinform. 1–14. https://doi.org/10.1109/TCBB.2022.3204619.

Nie, F., Huang, H., Cai, X., Ding, C., 2010. Efficient and robust feature selection via joint \mathscr2,1-norms minimization. In: Lafferty, J., Williams, C., Shawe-Taylor, J., Zemel, R., Culotta, A. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.

Ning, Z., Xiao, Q., Feng, Q., Chen, W., Zhang, Y., 2021. Relation-Induced Multi-Modal Shared Representation Learning for Alzheimer's Disease Diagnosis. IEEE Trans. Med. Imaging 40, 1632–1645. https://doi.org/10.1109/TMI.2021.3063150.

Rakotomamonjy, Alain &, Bach, Francis &, Canu, Stephane &, Grandvalet, Yves, 2008. Simplemkl. J. Mach. Learn. Res. 9.

Rolls, E.T., Joliot, M., Tzourio-Mazoyer, N., 2015. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. NeuroImage 122, 1–5. https://doi.org/10.1016/j.neuroimage.2015.07.075.

Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., Fontanella, S., Gori, P., Guillon, J., Guyot, A., Hassanaly, R., Jacquemont, T., Lu, P., Marcoux, A., Moreau, T., Samper-González, J., Teichmann, M., Thibeau-Sutre, E., Vaillant, G., Wen, J., Wild, A., Habert, M.-O., Durrleman, S., Colliot, O., 2021. Clinica: an open-source software platform for reproducible clinical neuroscience studies. Front. Neuroinformat. 15, 689675. https://doi.org/10.3389/fninf.2021.689675.

Rowe, C.C., Ng, S., Ackermann, U., Gong, S.J., Pike, K., Savage, G., Cowie, T.F., Dickinson, K.L., Maruff, P., Darby, D., Smith, C., Woodward, M., Merory, J., Tochon-Danguy, H., O'Keefe, G., Klunk, W.E., Mathis, C.A., Price, J.C., Masters, C.L., Villemagne, V.L., 2007. Imaging β-amyloid burden in aging and dementia. Neurology 68, 1718–1725. https://doi.org/10.1212/01.wnl.0000261919.22630.ea.

Samper-González, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., Routier, A., Guillon, J., Bacci, M., Wen, J., Bertrand, A., Bertin, H., Habert, M.-O., Durrleman, S., Evgeniou, T., Colliot, O., 2018. Reproducible evaluation of classification methods in Alzheimer's disease: framework and application to MRI and PET data. NeuroImage 183, 504–521. https://doi.org/10.1016/j.neuroimage.2018.08.042.

Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W., 2008. Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39, 1064–1080. https://doi.org/10.1016/j.neuroimage.2007.09.031.

Shen, T., Jiang, J., Lu, J., Wang, M., Zuo, C., Yu, Z., Yan, Z., 2019. Predicting Alzheimer disease from mild cognitive impairment with a deep belief network based on 18F-FDG-PET images. Mol. Imaging 18. https://doi.org/10.1177/1536012119877285, 153601211987728.

Shi, H., Ma, D., Nie, Y., Faisal Beg, M., Pei, J., Cao, J., 2021. Early diagnosis of Alzheimer's disease on ADNI data using novel longitudinal score based on functional principal component analysis. J. Med. Imaging 8. https://doi.org/10.1117/1.JMI.8.2.024502.

Shi, Y., Zu, C., Hong, M., Zhou, L., Wang, L., Wu, X., Zhou, J., Zhang, D., Wang, Y., 2022. ASMFS: adaptive-similarity-based multi-modality feature selection for classification of Alzheimer's disease. Pattern Recognit. 126, 108566. https://doi.org/10.1016/j.patcog.2022.108566.

Singanamalli, A., Wang, H., Madabhushi, A., 2017. Cascaded multi-view canonical correlation (CaMCCo) for early diagnosis of Alzheimer's disease via fusion of clinical, imaging and Omic features. Sci. Rep. 7, 8137. https://doi.org/10.1038/s41598-017-03925-0.

Suk, H.-I., Lee, S.-W., Shen, D., 2014. Subclass-based multi-task learning for Alzheimer's disease diagnosis. Front. Aging Neurosci. 6. https://doi.org/10.3389/fnagi.2014.00168.

Thomas, B.A., Cuplov, V., Bousse, A., Mendes, A., Thielemans, K., Hutton, B.F., Erlandsson, K., 2016. PETPVC: a toolbox for performing partial volume correction techniques in positron emission tomography. Phys. Med. Biol. 61, 7975. https://doi.org/10.1088/0031-9155/61/22/7975.

Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D., 2017. Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. Pattern Recognit. 63, 171–181. https://doi.org/10.1016/j.patcog.2016.10.009.

Treatments for Alzheimer's [WWW Document], n.d.. Alzheimers Dis. Dement. URL https://alz.org/alzheimers-dementia/treatments (accessed 12.1.23).

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15, 273–289. https://doi.org/10.1006/nimg.2001.0978.

Wang, B., Lu, K., Zheng, X., Su, B., Zhou, Y., Chen, P., Zhang, J., 2018. Early-Stage Identification of Alzheimer's Disease Using a Two-stage Ensemble Classifier. Curr. Bioinforma. 13, 529–535. https://doi.org/10.2174/1574893613666180328093114.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., Morris, J.C., Petersen, R.C., Saykin, A.J., Schmidt, M.E., Shaw, L., Siuciak, J.A., Soares, H., Toga, A.W., Trojanowski, J.Q., 2012. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimer. Dement. 8. https://doi.org/10.1016/j.jalz.2011.09.172.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Cedarbaum, J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Luthman, J., Morris, J.C., Petersen, R. C., Saykin, A.J., Shaw, L., Shen, L., Schwarz, A., Toga, A.W., Trojanowski, J.Q., 2015. 2014 update of the Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimer. Dement. 11. https://doi.org/10.1016/j.jalz.2014.11.001.

What Are the Signs of Alzheimer's Disease? [WWW Document], 2022. Natl. Inst. Aging. URL https://www.nia.nih.gov/health/alzheimers-symptoms-and-diagnosis/what-are-signs-alzheimers-disease (accessed 12.23.23).

Wolf, T.N., Pölsterl, S., Wachinger, C., 2022. DAFT: A universal module to interweave tabular data and 3D images in CNNs. NeuroImage 260, 119505. https://doi.org/10.1016/j.neuroimage.2022.119505.

Wyman, B.T., Harvey, D.J., Crawford, K., Bernstein, M.A., Carmichael, O., Cole, P.E., Crane, P.K., DeCarli, C., Fox, N.C., Gunter, J.L., Hill, D., Killiany, R.J., Pachai, C., Schwarz, A.J., Schuff, N., Senjem, M.L., Suhy, J., Thompson, P.M., Weiner, M., Jack, C.R., Alzheimer's Disease Neuroimaging Initiative, 2013. Standardization of analysis sets for reporting results from ADNI MRI data. Alzheimer's Dement. 9, 332–337. https://doi.org/10.1016/j.jalz.2012.06.004.

Xu, L., Wu, X., Chen, K., Yao, L., 2015. Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment. Comput.

Methods Programs Biomed. 122, 182–190. https://doi.org/10.1016/j.cmpb.2015.08.004.

Ye, T., Zu, C., Jie, B., Shen, D., Zhang, D., 2016. Discriminative multi-task feature selection for multi-modality classification of Alzheimer's disease. Brain Imaging Behav. 10, 739–749. https://doi.org/10.1007/s11682-015-9437-x.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 55, 856–867. https://doi.org/10.1016/j.neuroimage.2011.01.008.

Zhang, Lefei, Zhang, Q., Zhang, Liangpei, Tao, D., Huang, X., Du, B., 2015. Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding. Pattern Recognit. 48, 3102–3112. https://doi.org/10.1016/j.patcog.2014.12.016.

Zhang, J., He, X., Qing, L., Xu, Y., Liu, Y., Chen, H., 2022. Multi-scale discriminative regions analysis in FDG-PET imaging for early diagnosis of Alzheimer's disease. J. Neural Eng. 19, 046030. https://doi.org/10.1088/1741-2552/ac8450.

Zhao, Z., Anand, R., Wang, M., 2019. Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform, in: 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA). Presented at the 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), IEEE, Washington, DC, USA, pp. 442-452. doi:10.1109/DSAA.2019.00059.

Zhou, X., Wu, R., Zeng, Y., Qi, Z., Ferraro, S., Xu, L., Zheng, X., Li, J., Fu, M., Yao, S., Kendrick, K.M., Becker, B., 2022. Choice of Voxel-based Morphometry processing pipeline drives variability in the location of neuroanatomical brain markers. Commun. Biol. 5, 913. https://doi.org/10.1038/s42003-022-03880-1.

Zu, C., Jie, B., Liu, M., Chen, S., Shen, D., Zhang, D., the Alzheimer's Disease Neuroimaging Initiative, 2016. Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment. Brain Imaging Behav. 10, 1148–1159. https://doi.org/10.1007/s11682-015-9480-7.

Further reading

Pisner, D.A., Schnyer, D.M., 2020. Support vector machine. Machine Learning. Elsevier, pp. 101–121. https://doi.org/10.1016/B978-0-12-815739-8.00006-7.