Journal of Alzheimer's Disease xx (2021) x–xx
DOI 10.3233/JAD-210064
IOS Press

PET Imaging of Tau Pathology and Amyloid-β, and MRI for Alzheimer's Disease Feature Fusion and Multimodal Classification

Mehdi Shojaie[a,*], Solale Tabarestani[a], Mercedes Cabrerizo[a], Steven T. DeKosky[b,f], David E. Vaillancourt[b,c,f], David Loewenstein[d,f], Ranjan Duara[e,f] and Malek Adjouadi[a,f]
[a]*Center for Advanced Technology and Education, Department of Electrical and Computer Engineering, Florida International University, Miami, FL, USA*
[b]*Department of Neurology, University of Florida, Gainesville, FL, USA*
[c]*Department of Applied Physiology and Kinesiology, University of Florida, Gainesville, FL, USA*

Accepted 16 September 2021
Pre-press 25 October 2021

Abstract.
Background: Machine learning is a promising tool for biomarker-based diagnosis of Alzheimer's disease (AD). Performing multimodal feature selection and studying the interaction between biological and clinical AD can help to improve the performance of the diagnosis models.
Objective: This study aims to formulate a feature ranking metric based on the mutual information index to assess the relevance and redundancy of regional biomarkers and improve the AD classification accuracy.
Methods: From the Alzheimer's Disease Neuroimaging Initiative (ADNI), 722 participants with three modalities, including florbetapir-PET, flortaucipir-PET, and MRI, were studied. The multivariate mutual information metric was utilized to capture the redundancy and complementarity of the predictors and develop a feature ranking approach. This was followed by evaluating the capability of single-modal and multimodal biomarkers in predicting the cognitive stage.
Results: Although amyloid-β deposition is an earlier event in the disease trajectory, tau PET with feature selection yielded a higher early-stage classification F1-score (65.4%) compared to amyloid-β PET (63.3%) and MRI (63.2%). The SVC multimodal scenario with feature selection improved the F1-score to 70.0% and 71.8% for the early and late-stage, respectively. When age and risk factors were included, the scores improved by 2 to 4%. The Amyloid-Tau-Neurodegeneration [AT(N)] framework helped to interpret the classification results for different biomarker categories.
Conclusion: The results underscore the utility of a novel feature selection approach to reduce the dimensionality of multimodal datasets and enhance model performance. The AT(N) biomarker framework can help to explore the misclassified cases by revealing the relationship between neuropathological biomarkers and cognition.

Keywords: Alzheimer's disease, amyloid-β, classification, feature selection, information theory, machine-learning, multimodal imaging, tau

Uncorrected Author Proof

INTRODUCTION

With the aging of society, Alzheimer's disease (AD) is bound to affect more people, with projections suggesting that there will be over 13.8 million

*Correspondence to: Mehdi Shojaie, 10555 W Flagler St., Miami, FL, 33174, USA. Tel.: +1 305 781 8163; E-mail: mshoj004@fiu.edu.

ISSN 1387-2877/$35.00 © 2021 – IOS Press. All rights reserved.

people with dementia by 2050 in the US [1]. A misfolding and abnormal deposition of specific proteins in the brain is recognized as the pathological cause for the initiation and progression of this neurodegenerative disease. AD is irreversible, causing significant memory and behavioral issues. Therefore, researchers are keen to identify its earliest manifestations, even at the pre-symptomatic stage, to plan for and more effectively take advantage of emerging early treatment and therapeutic interventions. Thus, effective diagnosis of AD and its early stage, i.e., mild cognitive impairment (MCI), specifically using computer-aided methods, has attracted extensive attention in recent years [2–14].

Several well-established biomarkers associated with the pathology of AD have been identified and studied by researchers for decades. Magnetic resonance imaging (MRI) as a structural indicator for brain atrophy, measures of tau and amyloid-β (Aβ) from cerebrospinal fluid (CSF), and Aβ accumulation from regional positron emission tomography (PET) and hypometabolism from fluorodeoxyglucose (FDG) PET are among the most remarkable biomarkers for AD. In recent years, several tau PET tracers such as ^{11}C-PBB3, ^{18}F-AV1451, and ^{18}F-THK have been developed, which enable *in vivo* visualization of tau pathology in brain regions. Tau imaging can help to facilitate disease staging and diagnosis. Compared to Aβ, tau is a delayed event and is more related to cognitive decline [15, 16]. The interrelatedness of these two biomarkers has been extensively studied [17–21]. Moreover, the temporal ordering of biomarkers provides added insight into AD staging. Based on such biomarkers ordering, a disease progression score has been defined in [22]. Biomarkers of Aβ plaque, i.e., amyloid PET and CSF Aβ, represent the initiating events of AD that happen during the cognitively normal stage. On the other hand, biomarkers of neurodegeneration, including MRI, FDG-PET, and CSF total tau, are later events that correlate with cognitive decline [23]. Besides the pathological biomarkers, there are other contributing variables in AD diagnosis, such as risk factors (age, gender, and *APOE* ε4) and protective factors (cognitive reserve, brain resilience, and resistance). The variability of the factors, including age, gender, *APOE* ε4 genotype, and year of education between AD subtypes, can be used to address the disease heterogeneity to some extent.

In an effort to present a biological definition of AD, biomarkers are pathologically grouped into three classes. This scheme is known as AT(N) with "A", "T", and "(N)" representing Aβ, tau, and neurodegeneration biomarker groups, respectively. Based on this system, each biomarker class is labeled as positive or negative through defined cut-points to determine the overall pathology status [24]. The AT(N) framework attempts to reflect the interactions between neuropathological changes (characterized by biomarkers profiles) and the cognitive stage (determined clinically through symptoms). This framework can serve as a helpful supplementary tool when interpreting the results of a computer-aided diagnosis system.

While each neuroimaging modality provides distinct features and measures for AD diagnosis, their fusion consolidates their unique strengths when using effective machine learning and deep learning models [25–29]. In retrospect, few multimodal studies include tau imaging for computer-aided diagnosis of AD.

An initial step required for the machine learning-based diagnosis is the optimal data representation through a feature extraction procedure. Feature extraction methods can be categorized as voxel-based, region of interest (ROI)-based, and patch-based techniques. Among them, ROI-based features are more common due to their consistency and lower dimensionality [25, 30]. In AD studies, the sample size is typically small, and the dimensionality of voxel-based and even ROI-based features is high. This makes it difficult for the machine learning model to generalize to unseen data while avoiding overfitting. Therefore, to reduce the model complexity and enhance its performance, removing redundant and extraneous features by selecting the most informative ones is a critical step [31–34]. Also, feature selection can be used to understand the process under study by identifying disease-prone regions that contribute best to AD diagnosis and disease progression.

In some feature selection methods, the selection process is embedded in the learning algorithm, and the model accuracy or loss is then used to evaluate different subsets of features. With the use of these methods, an optimized combination of features can be achieved; however, these approaches are subject to the curse of dimensionality. Another category of techniques known as filter methods uses a criterion such as Pearson's correlation, ANOVA, *t*-test, chi-square test, and mutual information, among others, to evaluate the many features and determine their relevance to the target variable [35, 36]. In [31], the similarity between samples was computed, and their consistency metrics

Uncorrected Author Proof

have been used for multimodal feature selection. In [37], a feature selection method was developed based on the receiver operating characteristic (ROC) curve for each volumes-of-interest (VOI) where the classification true positive rate is plotted versus the false positive rate using only that specific VOI. In [38], the linear discriminant analysis and locality preserving projection learning methods have been combined with a sparse regression model to determine discriminative features. Most filter methods use univariate metrics in which features are evaluated independently, and the interaction between them is often overlooked. Also, filter methods focus mainly on the linear relationship between variables, and any nonlinear dependencies are neglected. Concerning the associations between variables, there exists some research endeavors for incorporating the correlation and redundancy of the features. However, due to the nature of the used metrics, these approaches are mainly unsupervised, and the detected relationships are not necessarily connected to the target variable and may not be valuable concerning the classification problem. Another group of methods uses embedded regularization for sparse feature learning in which the interaction of all variables is considered [39–41]. However, in these models, the variable selection is less interpretable, limiting the flexibility and ability to further explore the discriminative features.

In this study, we aimed to implement a multimodal feature fusion approach for the machine learning-based diagnosis of AD. A feature selection technique was proposed based on the multivariate mutual information (MMI) criterion. We attempted to handle feature redundancy and complementarity in a supervised manner where the shared information between features is evaluated in terms of its capability in predicting the target variable. MRI, Amyloid-β PET, and tau PET data from the ADNI cohort were used in this multimodal study. The effect of modalities on the disease staging was evaluated both individually and combined. Machine learning models, including support vector machine, random forest (RF), and eXtreme gradient boosting (XGB), were used for the classification of different stages of the disease and the effect of the proposed feature selection method on the classification performance was evaluated. Lastly, the AT(N) biomarkers framework was used to investigate the interconnection between the biomarkers' profile and the cognitive stage to assess the classification performance degradation due to biomarker insufficiency.

MATERIALS AND METHODS

Participants

The clinical data used for our analysis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). ADNI was launched in 2003 as a public-private partnership, directed by Principal Investigator Michael W. Weiner, MD. The primary objective of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. For up-to-date information, see http://www.adni-info.org.

In this study, the data were collected from three modalities in the ADNI 3 cohort, including amyloid PET (agent: ^{18}F-AV45), tau PET (agent: ^{18}F-AV1451), and MRI. For each participant, all modalities have been collected from the same visit. The MRI scan is a T1 weighted image that has gone through preprocessing steps, including gradient wrapping, scaling, B1 correction, and inhomogeneity correction. For the florbetapir and flortaucipir data, four preprocessing steps have been followed, including co-registered dynamic, averaged, standardized image and voxel size, and uniform resolution. T1 MRI scans have been processed through FreeSurfer for skull-stripping and segmentation of cortical and subcortical regions. In the next step, florbetapir and flortaucipir images have been co-registered to the subject's MRI from the same visit. Finally, volume-weighted florbetapir and flortaucipir average are defined in each cortical and subcortical region of interest, and regional standardized uptake value ratio (SUVR) is then calculated. More information about the preprocessing steps and processing methods can be found at http://ida.loni.usc.edu. The florbetapir (^{18}F-AV45) dataset analysis comprises reference region options of the whole cerebellum, cerebellar grey matter, and brain stem in addition to cortical and summary of SUVR measurements. The participant demographics and Mini-Mental State Examination (MMSE) score for each group (mean and standard deviation) are reported in Table 1. Figure 1 illustrates the distribution of average SUVRs (among all regions) for the sample set. Since not all participants have undergone all tests, the dataset contains multiple instances with missing values which are dropped in some scenarios depending on the objective of the analysis.

In this study, different types of variables, including cortical thickness and SUVR values, non-tissue

Uncorrected Author Proof

Table 1
Participant demographics and mini-mental state examination (MMSE) score for different diagnosis groups of the ADNI3 cohort. *P*-value is reported between MCI-CN and AD-CN populations

Groups	Subject (f/m)	Age (y), [p]	Education, (y) [p]	MMSE, [p]
CN	277 (153/124)	71.80 ± 5.70, [–]	16.67 ± 2.47, [–]	28.63 ± 2.12, [–]
MCI	378 (155/223)	71.26 ± 7.66, [0.179]	16.25 ± 2.61, [0.027]	26.87 ± 4.20, [< 0.001]
AD	67 (26/41)	73.41 ± 8.78, [0.075]	16.43 ± 2.35, [0.290]	22.37 ± 2.39, [< 0.001]

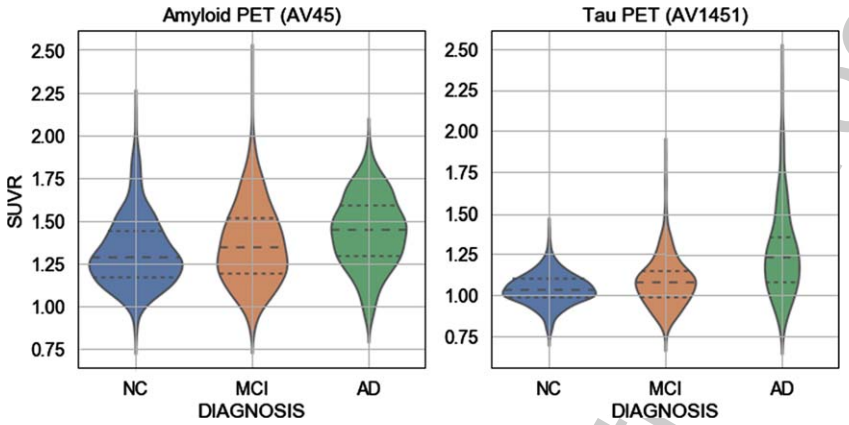

Fig. 1. Distribution of the mean value of amyloid-β and tau SUVRs in each disease group for ADNI3 cohort participants; CN, Cognitively Normal; MCI, Mild Cognitive Impairment; AD, Alzheimer's Disease.

SUVR values, and AD risk factors, were used as features for the machine learning algorithm. In the preprocessing stage, the feature set is normalized to a common scale before feeding it to the classification model. It is worth noting that the SUVR values in non-brain areas represent off-target binding by the ligand and are not related to AD pathophysiology. Such SUVR values could still be potentially beneficial for the machine learning-based classification task despite the fact that they are not interpretable as biomarkers of AD.

Feature selection

The high dimensionality of multimodal regional AD data relative to the sample size can diminish the model performance. The purpose of feature selection is to find a feature subset that yields an optimal classification score. This selection process can also help to enhance the generalization ability and interpretability of the model. The objective is to come up with a subset of features with minimum size and maximum possible information about the class variable. This can be achieved by preserving the most relevant features and dismissing the irrelevant and the redundant ones. Redundant features may not necessarily damage the system's performance. However, to limit the feature space size and complexity, it is beneficial to remove the redundant features and keep the complementary ones to maximize the total amount of relevant information. An approach is thus proposed based on multivariate mutual information to measure the relevance and redundancy of the features.

To determine the relevance of a feature, univariate filter-based feature selection measures can be used. With such measures, the relationship between each feature and the target variable is evaluated individually. One of the most common criteria for this task is the Pearson correlation coefficient which is a number between [–1, 1], with +1, –1, and 0 representing maximum linear correlation, maximum inverse linear correlation, and no linear correlation between the two variables, respectively. Other univariate criteria include mutual information, ANOVA test, and Chi-squared test, whose performance may vary depending on the type of the input and output variables (continuous or categorical variable). Mutual information (MI) is a powerful statistical metric that measures common information between random variables and is relatively robust to the data type. Unlike the correlation measure, MI can also detect nonlinear relationships between variables. Moreover, it can be extended to more than two variables to determine the redundancy of multiple variables [34]. In this study, a

Uncorrected Author Proof

methodology is proposed to rank features based on pairwise redundancy and complementarity of features using MMI.

MI between two discrete random variables is defined as:

$$I(x;y) = \sum_x \sum_y p(x,y).log\frac{p(x,y)}{p(x)\,p(y)} \quad (1)$$

where x and y are random variables and p(.) is the probability of a random variable. MI is zero when x and y are independent and is positive when there is common information between them.

At first, MI was calculated between each feature and its target variable. This determines the relevance of each feature. Next, to incorporate the interaction of features, MI was calculated between a subset of features and a target variable as *I(S;y)*, where S is a subset of features and y is the target. For the case of a subset of two features ($S=\{x_1,x_2\}$), the relationship between MI of S and y ($I(x_1,x_2;y)$) and MI of each feature and y ($I(x_1;y)$, $I(x_2;y)$) is defined as follows:

$$I(x_1,x_2;y) = I(x_1;y) + I(x_2;y) - I(x_1;x_2;y) \quad (2)$$

where the three terms on the right side can be calculated using (1). Based on (2), the amount of information that (x_1,x_2) have about y can be defined as the sum of the common information of x_1 and y ($I(x_1;y)$) plus that of x_2 and y ($I(x_2;y)$) minus the intersection of the first two terms, which is the common information of all three variables x_1, x_2 and y ($I(x_1;x_2;y)$). The last term is known as the MMI, which determines the shared information between multiple variables and is defined as follows:

$$I(x_1;x_2;y) = \sum_{x_1}\sum_{x_2}\sum_y p(x_1,x_2,y).\,log\frac{p(x_1,x_2,y)}{p(x_1)\,p(x_2)\,p(y)} \quad (3)$$

When MMI is positive, there is redundancy between x_1 and x_2, and the information of a subset of them is less than the sum of their individual information. On the other hand, when MMI is negative, x_1 and x_2 carry complementary information about y, and the information of x_1 and x_2 combined is more than the sum of their individual information. Therefore, in (2), the interaction of features is considered through the MMI term, which can be treated as a measure of redundancy and complementarity.

To rank the features, a metric is defined for each feature based on the MI between that feature and the target variable and the redundancy or complementarity of that feature with every other feature. This new metric is as defined as follows:

$$FS_i = I(x_i;y) - \alpha \sum_{\substack{j \\ j\neq i}} I\left(x_i;x_j;y\right) \quad (4)$$

where FS_i is the score of the i^{th} feature, with α being a constant. The first term is the MI of the i^{th} feature and the target variable, and the second term represents the pairwise interaction (redundancy/complementarity) of the i^{th} feature and all other features, which can consist of positive and negative elements. When α is zero, the interaction term is ignored, and the feature scores only depend on the individual scores. As α increases, a larger weight is assigned to the redundancy term so that the overall score of redundant features decreases while that of complementary ones increases. To select the value of coefficient α, the classification experiment was conducted using different values of α, and the optimal value was determined as the one associated with the highest classification score. The feature score (FS) was then calculated for all features, and the top features were determined accordingly. To evaluate different scenarios, first, the top features were detected for each individual modality to find the prominent regions based on each biomarker. Then, the process was repeated for the multimodal data so that the top regions in terms of all modalities combined were identified. Also, the importance of specific regions and biomarkers at various stages of the disease was evaluated. In the next step, to prove the effectiveness of the new metric for feature selection, multiple classification scenarios were implemented.

Classification

In recent years, artificial intelligence has proved to be a promising tool for diagnosing and predicting the trajectory of the disease. In this study, machine learning architectures were used for AD diagnosis at different stages using single-modality and multimodality data. It is worth noting that before implementing the classification task, the feature space was scaled in the range between zero and one. The scaling estimator was built solely based on the training data (to avoid data leakage from the test set) and was applied to each feature individually in both training and test sets so that each feature is in the [0–1] interval. The models used for the classification task include support vector classifier (SVC), RF of decision trees, and XGB. SVC is a classifier that attempts to categorize data points

Uncorrected Author Proof

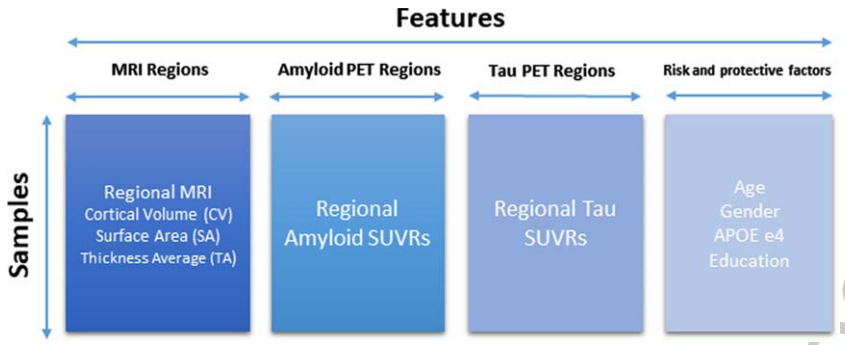

Fig. 2. Structure of the used data for the classification process.

based on their classes in a high-dimensional space by a hyperplane. By mapping the data points onto a higher-dimensional space, SVC can classify nonlinearly separable data using nonlinear kernels like polynomial and radial basis function. To alter the bias and variance of the model, the regularization parameters C and gamma of the SVC can be adjusted. The parameters control the trade-off between the training accuracy and model generalization ability for the testing stage. As the next model, the RF algorithm relies on the key concept of decision trees and leverages the ensembling and voting mechanisms to enhance the classification and prediction accuracy while preventing overfitting. The model parameters include the number of trees, sample size, maximum depth of each tree, and the maximum number of features used for each split. XGB, on the other hand, is a learning technique that consists of an ensemble of weak learners, such as decision trees, that operate in a sequence where each subsequent learner attempts to correct the errors of the previous learner. The number of trees, the maximum depth of a tree, and the sample size for each step are among the XGB control parameters. To evaluate the models and also to optimize the models parameters, k-fold cross-validation was used. In order to prevent data leakage between these two tasks, the nested cross-validation technique was implemented. An inner 5-fold cross-validation was performed for hyperparameter optimization, while an outer 6-fold cross-validation was used for validation and reporting the model scores. The structure of the data for the classification task is shown in Fig. 2. Multiple single modality and multimodality experiments were performed for binary and multiclass classification. A similar set of experiments were then implemented after applying the proposed feature selection approach. Finally, to include the risk and protective factors in the analysis, covariates including age, *APOE* ε4, gender, and education level were integrated into the feature set, and the classification process was repeated.

Interconnection between AD neuropathology and cognitive stage

In this study, MRI and PET scans have been used for automatic classification and prediction of the cognitive stage. However, the classification task remains challenging due to the heterogeneity of the disease. A critical factor that can degrade the model performance is the lack of sufficient biomarkers that are informative enough to perfectly determine the cognitive stage. We tried to explore the available biomarkers to investigate the performance limitation imposed by the dataset.

Due to biomarker insufficiency, cognitive symptoms are not perfectly linked to AD neuropathological changes measured by available biomarkers. Simply put, symptoms are not specific to AD, nor do abnormal AD biomarkers guarantee the existence of symptoms. Neuropathologic changes in AD are determined by postmortem inspections and measured *in vivo* through biomarkers. Clinical AD, on the other hand, is defined based on the cognitive stage and is measured through the symptoms' manifestation. A percentage of individuals with clinical AD do not have postmortem evidence of AD pathology.

Similarly, some individuals in the cognitively normal elderly group show signs of AD pathology at autopsy. This may result in false-negative and false-positive outcomes in our classification task. To study this effect, we investigated the available biomarkers and their corresponding cognitive stage based on the AT(N) biomarker profile system introduced in [24]. The AT(N) framework of the National Institute on Aging-Alzheimer's Association is an effort

Uncorrected Author Proof

Table 2
Interaction between clinically diagnosed cognitive stage and AT(N) biomarkers [24]

		Cognitive stage (Clinical diagnosis)		
		Cognitively Normal	Mild Cognitive Impairment	Dementia
Biomarker Profile	A–T–N–	Normal AD biomarkers, and CN	Normal AD biomarkers with MCI	Normal AD biomarkers with dementia
	A+T–N–	AD pathologic change, and CN	AD pathologic change with MCI	AD pathologic change with dementia
	A+T+N– A+T+N+	Preclinical AD with no cognitive impairment	AD biomarkers with MCI	AD biomarkers with dementia
	A+T–N+	AD and concomitant suspected non-AD pathologic change, and CN	AD and concomitant suspected non-AD pathologic change with MCI	AD and concomitant suspected non-AD pathologic change with dementia
	A–T+N– A–T–N+ A–T+N+	non-AD pathologic change, and CN	non-AD pathologic change with MCI	non-AD pathologic change with dementia

CN, cognitively normal; MCI, mild cognitive impairment; AD, Alzheimer's disease; A, Aggregated amyloid-β; T, Aggregated tau; N, Neurodegeneration; +/−, The value of a biomarker summary measure is higher/lower than the cut-point.

toward investigating the interaction between AD neuropathology and cognitive status. In this biomarker grouping system, the biomarkers are classified into three categories based on their underlying pathologic process. The label "A" represents amyloid PET and CSF Aβ as biomarkers of cortical Aβ, "T" denotes tau PET and CSF phosphorylated tau (P-tau) as biomarkers of fibrillar tau, and neurodegeneration is labeled as "(N)" measured by CSF total tau (T-tau), FDG PET, and MRI.

The imaging and CSF biomarkers are expressed in continuous values; however, in certain situations such as research studies and treatment trials, a binary grouping of biomarkers (positive/negative) may be preferred. To achieve such types of positive/negative results, appropriate cut-points are defined for each biomarker. For florbetapir (AV45) SUVR cut-points, we adopted the values reported in [42]. Summary SUVR is defined as the weighted average of florbetapir uptake in lateral temporal and parietal, lateral and medial frontal, anterior, and posterior cingulate normalized by the uptake in the whole cerebellum. Then, a cut-point of 1.11 is applied to this summary SUVR, which is equivalent to the 95th percentile of the biomarker distribution of the young control normal group. For tau PET SUVRs and MRI cortical thickness, the cut-points determined in [43] by Clifford R. Jack Jr. were used. A tau PET summary SUVR is defined based on the volume-weighted average of the SUVR in inferior temporal, middle temporal, entorhinal, amygdala, parahippocampal, and fusiform ROIs normalized to the cerebellar crus grey. For the tau PET summary SUVR, cut-points of 1.19 and 1.32 are defined based on the specificity method (the 95th percentile of the biomarker distribution of the young control normal individuals) and the accuracy of impaired versus age-matched control normal method, respectively. From MRI, the surface-area weighted average is determined for the cortical thickness in entorhinal, inferior temporal, middle temporal, and fusiform regions. Cortical thickness cut-points of 2.69 and 2.57 mm are selected respectively based on specificity and accuracy methods which were also used in the tau PET case.

Based on the defined cut-points, various biomarker profiles can be identified in the AT(N) framework. These biomarker grouping and their relationship with the cognitive stages are shown in Table 2. As seen in the table, the A–T–N– group represents individuals with normal AD biomarkers. Participants with amyloid positive but normal tau pathology and neurodegeneration biomarkers (A+T–N–) are tagged as "Alzheimer's pathologic change." Those with evidence of amyloid deposition along with tau pathology and regardless of neurodegeneration condition (A+T+N+/–) are considered to belong to the "preclinical Alzheimer's disease" group. Amyloid negative individuals with abnormal tau or neurodegeneration biomarkers (A–T–N+, A–T+N–, A–T+N+) are defined as "suspected non-Alzheimer's pathology change". Finally, the A+T–N+ category represents simultaneous "Alzheimer's pathologic change" and "non-AD neurodegeneration". Although the biomarker signature carries some information about the cognition status, each biomarker profile can belong to any cognitive stage.

The AT(N) framework combined with the described cut-points were used to establish the biomarker profile groups for our dataset. We then identified the sub-groups that are more susceptible

Uncorrected Author Proof

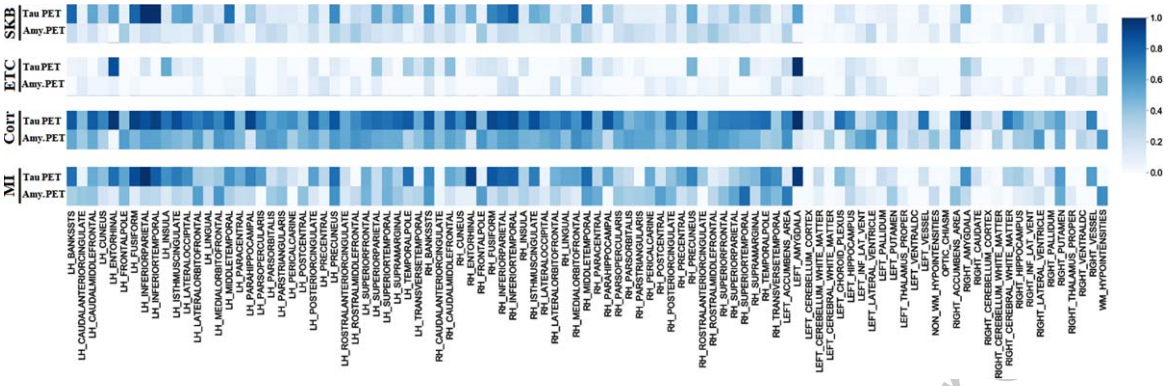

Fig. 3. Regional feature importance scores for amyloid PET SUVRs (AV45) and tau PET SUVRs (AV1451). The feature scores were determined using four filter-based feature selection measures, namely, SelectKBest (SKB), ExtraTreesClassifier (ETC), correlation coefficient (Corr), and mutual information (MI), as shown in the vertical axis. For each region shown in the horizontal axis, one feature is defined for amyloid SUVR and one for tau SUVR. The value of feature scores is normalized between 0 and 1 and is illustrated by the color intensity of their corresponding box in the figure. Features with larger scores are more informative for the classification task. Based on the results, amyloid SUVRs including entorhinal, inferior parietal, inferior temporal, amygdala, and bankssts and tau SUVRs including frontal pole and accumbens are among the top features.

to misclassification and explored their underlying causes. This is done by focusing on those groups in which the biological AD biomarkers cannot be an informative representation of the cognitive stage. For instance, individuals with normal AD biomarkers but clinical AD diagnosis are likely to be classified as non-AD class. Also, subjects with abnormal AD biomarkers but no cognitive impairment might be identified as AD class by the model. The number of subjects in each AT(N) group was calculated for our dataset, and the probability of occurring false positive and false negative outcomes is measured, representing the contribution of biomarker shortage to the classification error.

RESULTS

Feature selection results

Various feature selection approaches were implemented under multiple classification scenarios. At first, conventional univariate criteria and methods, including Correlation coefficient, SelectKBest, ExtraTreesClassifier, and univariate mutual information have been implemented. For the amyloid and tau PET modalities and the three-class classification case (CN/MCI/AD), the heatmap of the feature scores based on the abovementioned metrics is shown in Fig. 3. A total number of 110 features (two features per region for left and right hemispheres) have been included in this analysis. As seen, entorhinal, inferior parietal, inferior temporal, amygdala, and bankssts are among the top features based on tau PET, while

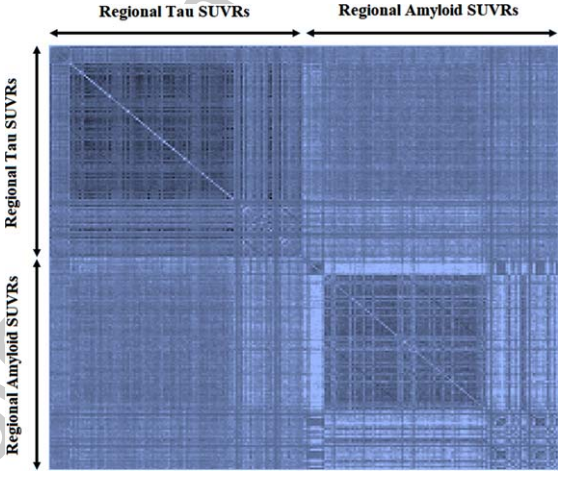

Fig. 4. Heatmap of multivariate mutual information (MMI) between pairwise amyloid and tau SUVR values given the class variable (y), calculated using equation (3). The diagonal elements represent the amount of information that each individual feature carries about the target variable. Brighter colors correspond to a higher amount of information. For non-diagonal elements, a positive MMI value is an indication of redundant information between two features, which corresponds to darker colors in the heatmap. On the other hand, complementary features have a negative MMI represented by brighter colors in the heatmap. As seen, more pairwise redundancy (more dark non-diagonal elements) exists for inside-modality features compared to between-modality features.

regions like frontal pole and accumbens are more prominent based on amyloid PET.

Next, the proposed MMI-based feature selection method was implemented. Using equation (3), pairwise MMI was calculated for all features, and the results are presented as a heatmap in Fig. 4. Again,

Uncorrected Proof

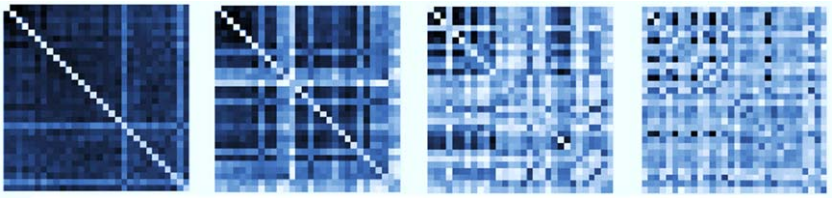

Fig. 5. Heatmap of top 30 features based on the FS-scores for different values of parameter α. For α = 0, the redundancy term is ignored, and the features are selected solely based on their relevance. In this case, dark non-diagonal elements of the heatmap represent more pairwise redundancy between features. For higher values of α, feature redundancy is decreased, and bright non-diagonal elements show less pairwise feature redundancy and more complementarity.

Table 3
Top features (amyloid-β and tau SUVRs) based on the proposed feature ranking method. The SUVR values were ranked using the calculated feature scores, and the top amyloid-β and tau SUVR features are presented. Top features are more informative for the AD diagnosis classification task

Tau PET	Left entorhinal	Left vessel	Third ventricle
	Left amygdala	Left inferior temporal	Right entorhinal
	Left middle temporal	Right amygdala	Right inferior temporal
Amyloid-β PET	Left medial orbitofrontal	Left rostral anterior cingulate	Right medial orbitofrontal
	Left accumbens area	Left hippocampus	CC anterior
	Left frontal pole	Right accumbens area	CC mid anterior
	Left lateral ventricle	Right lateral ventricle	CC posterior
	Left inf lat vent	Right frontal pole	

the CN/MCI/AD case based on the amyloid and tau PET modalities is considered here. In the heatmap, the diagonal elements show the amount of information that each feature has about the target variable. The brighter the color of a square, the more relevant is that particular feature. The non-diagonal elements show the degree of redundancy or complementarity of feature pairs concerning the target variable. The darker the color, the higher is the redundancy, and the lower is the complementarity.

To select the most relevant and informative features, both the individual scores (diagonal) and the mutual scores (non-diagonal) should be considered as described in the Methods section. The FS were calculated using equation (4). As indicated earlier, for each feature, the summation of the second term of the equation represents the interaction of that feature with every other feature. The summation terms are equivalent to each row or column of the heatmap of Fig. 4. The heatmap of the top 30 features based on the proposed FS-score is illustrated in Fig. 5 for different values of α. For α = 0, the score of a given feature solely depends on the feature's relevance. As seen in Fig. 5, in this case, top features include highly relevant (brighter diagonal) but possibly redundant features (darker non-diagonal) at the same time. For higher values of α, the redundancy term comes into play so that more redundant features are removed from the list of the top features. This results in selecting features with brighter non-diagonal elements (less redundant), as shown in Fig. 5 for higher values of α. This is a trade-off between feature relevance and redundancy, which is controlled by adjusting parameter α. It is worthwhile to add that too large values of α should be avoided since, in this situation, valuable features might be dropped only because they have some dependency on other features. For the specific case of α = 0.005, top features (amyloid-β and tau SUVRs) are listed in Table 3. Finally, the resulting scaled feature scores for the amyloid and tau SUVRs for different stages of the disease are represented in Fig. 6.

Classification results

After data preprocessing, exploratory data analysis, and feature selection, classification models (SVC, RF, and XGB) were implemented for MCI, and AD diagnosis and their performance were compared. Since the data is unbalanced, various evaluation metrics, including precision, recall, and F1-score, are reported besides accuracy. Experiments were conducted using different modalities, both separately and combined. Amyloid PET, tau PET, and MRI

Uncorrected Author Proof

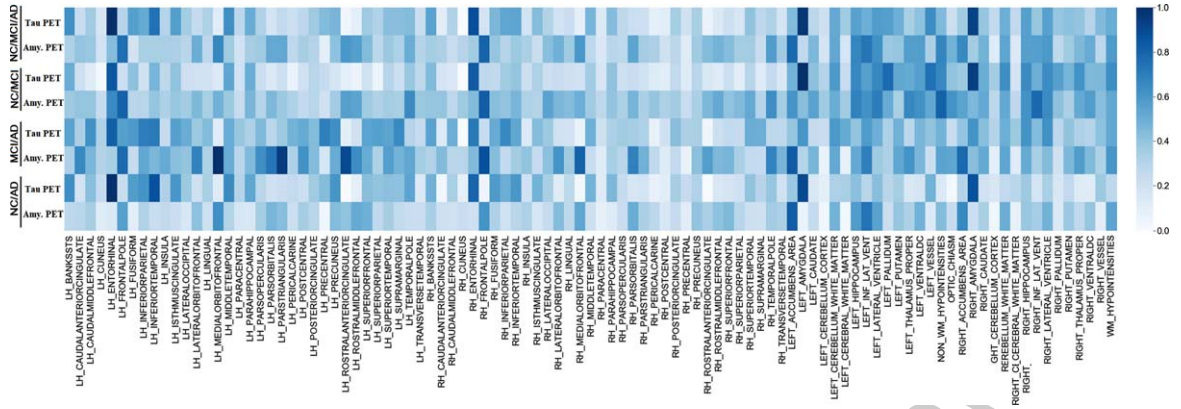

Fig. 6. Regional feature importance scores for amyloid PET SUVR (AV45) and tau PET SUVR (AV1451) based on the proposed feature selection method. As a supervised approach, the features scoring procedure was performed for four different classification tasks, including CN/MCI/AD, CN/MCI, MCI/AD, and CN/MCI/AD as shown in the vertical axis. For each region shown in the horizontal axis, one feature is defined for amyloid SUVR and one for tau SUVR. The value of feature scores is normalized between 0 and 1 and is illustrated by the color intensity of their corresponding box in the figure. Features with larger scores are more informative for the classification task. For tau SUVRs, entorhinal and amygdala were among the top features for all classification tasks, while pallidum and hippocampus were more informative for the CN/MCI case, and inferior parietal, inferior temporal, precuneus, and precentral for the MCI/AD case. On the other hand, for amyloid SUVRs, top features include frontal pole for all classification tasks, inferior lateral ventricle for the CN/MCI, and medial orbitofrontal, pars triangularis, and rostral anterior cingulate for the MCI/AD.

Table 4
Classification results **before feature selection** for three single-modality scenarios including amyloid PET SUVRs (tracer: AV45), tau PET SUVRs (tracer: AV1451), and MRI (cortical thickness) and two multimodality scenarios including "amyloid PET SUVRs & tau PET SUVRs" and "amyloid PET SUVRs & tau PET SUVRs & MRI cortical thickness". Three machine learning models, including SVC, RF, and XGB were used, and four scores, including accuracy, precision, recall, and F1-score are reported

		CN/MCI/AD				CN/MCI				MCI/AD				CN/AD			
Modality	Classifier	ACC	PRE	REC	F1	ACC	PRE	REC	F1	ACC	PRE	REC	F1	ACC	PRE	REC	F1
amyloid-β PET	SVC	60.2	52.6	49.7	50.4	68.9	65.2	61.6	61.9	74.9	66.2	64	64.8	88.6	78.8	76.3	77.4
	RF	58.6	46.4	44.5	44.5	66.9	62.4	60.1	60.3	75.9	67.6	64	65.2	89.6	81.3	76.9	78.8
	XGB	63.5	54.2	50.8	51.4	67.2	62.8	60.4	60.7	75.4	66.7	63	64.1	88.3	78.4	74.4	76.1
tau PET	SVC	64.7	57.8	48.5	49.9	69.4	65.9	62.1	62.5	75.4	66.4	60.9	62	90.9	86.6	75.9	79.9
	RF	62.9	55.3	48.9	50.4	68.2	64.1	61	61.3	79.7	74.4	67.2	69.2	90.6	85.4	75.7	79.4
	XGB	63.1	55.8	49.3	50.9	69.2	65.5	62.8	63.2	77.5	70.4	69.2	69.7	90.6	85.4	75.7	79.4
MRI	SVC	59.5	52.5	50.2	51.1	69.7	67.4	62.1	62.3	75.4	65.1	63.1	63.9	91.6	85.3	79.9	82.3
	RF	63.3	58.7	50.5	52.1	69	66.4	61.4	61.5	77.5	68.1	62.5	63.9	92.5	88.2	80.5	83.7
	XGB	62.6	57.5	50.2	52.2	65.5	61.1	58.8	58.8	78.2	69.5	64	65.5	90.8	83.5	77.6	80.1
Amyloid-β PET & tau PET	SVC	64.2	56.2	49.9	51.3	67.8	63.7	61.3	61.7	76.5	67.8	62.4	63.7	89.9	82.7	74.9	78
	RF	64.9	56.5	50.7	52	71.8	69.5	64.6	65.2	78.6	71.9	65.2	67	91.5	87.3	77.6	81.3
	XGB	64.9	64.4	53.5	56.5	67	62.8	61.1	61.4	80.7	75.5	68.8	70.8	91.2	84.8	79.1	81.6
Amyloid-β PET & tau PET & MRI	SVC	69.3	63	55.3	57.8	73.8	70.6	66.2	67.2	81	74.9	65.4	67.7	91.4	83.5	74.4	77.9
	RF	69	61.8	51.7	54.1	78	76.1	71.7	73	78.9	69.9	65.1	66.6	92.6	87.7	76.4	80.6
	XGB	68.8	62.9	54.6	57	78.3	78.1	70.5	72.2	78.2	68.5	61.5	63	91	81.4	75.5	78

CN, cognitively normal; MCI, mild cognitive impairment; AD, Alzheimer's disease; ACC, accuracy; PRE, precision; REC, recall; F1, F1-score; Amyloid-β PET, SUVR values with AV45 tracer; Tau PET, SUVR values with AV1451 tracer; MRI, Cortical thickness.

as single modalities, and combinations of {amyloid PET & tau PET}, and combinations of {amyloid PET & tau PET & MRI}, as multimodal scenarios were investigated, and the results are presented in Table 4. In terms of machine learning models, generally, SVC yields slightly less accurate scores compared to the other two models. The F1-scores of the three models for various scenarios can be seen in Fig. 7. Among single modality cases, tau PET has slightly higher scores for CN/MCI classification (early stages), and tau PET and MRI have improved results for MCI/AD and CN/AD cases. Multimodal scenarios resulted in enhanced performance in the three-class CN/MCI/AD and CN/MCI cases while not in the MCI/AD case. This is due to the fact that the feature selection has not yet been applied, and

Uncorrected Author Proof

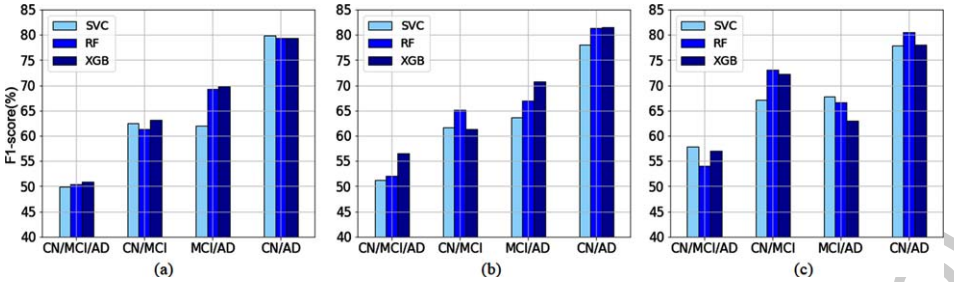

Fig. 7. Classification F1-score before feature selection for the three machine learning models, SVC, RF, and XGB, for different classification scenarios including CN/MCI/AD, CN/MCI, MCI/AD, and CN/AD; (a) Single modality; tau PET, (b) Multimodality; tau and amyloid PET, (c) Multimodality; tau and amyloid PET and MRI.

Table 5
Classification results **after feature selection** for three single-modality scenarios including amyloid PET SUVRs (tracer: AV45), tau PET SUVRs (tracer: AV1451), and MRI (cortical thickness) and two multimodality scenarios including "amyloid PET SUVRs & tau PET SUVRs" and "amyloid PET SUVRs & tau PET SUVRs & MRI cortical thickness". Three machine learning models, including SVC, RF, and XGB were used, and four scores, including accuracy, precision, recall, and F1-score are reported

		CN/MCI/AD				CN/MCI				MCI/AD				CN/AD			
Modality	Classifier	ACC	PRE	REC	F1	ACC	PRE	REC	F1	ACC	PRE	REC	F1	ACC	PRE	REC	F1
amyloid-β PET	SVC	62.4	57.9	52.1	53.9	69.7	66.2	62.8	63.3	78.1	71	68.2	69.3	90.9	84.6	78.5	81.1
	RF	61.3	52.4	49.6	50.1	68.7	64.8	61.4	61.7	78.1	71.4	65.5	67.1	89.6	81.6	76	78.4
	XGB	61.1	52.5	50.7	51.2	65.7	61	59.4	59.7	75.9	67.7	64.7	65.7	89	80.5	73.9	76.6
tau PET	SVC	65.3	55.9	53	53.9	71.9	69.6	64.7	65.4	77.5	70.2	66.5	67.8	89	80.9	73.1	76.1
	RF	64.9	57.9	50.4	52.1	68.7	64.8	61.9	62.3	79.1	72.9	68.2	69.8	92.2	89.2	79.2	83.1
	XGB	64.2	57	52	53.2	68.4	64.5	62.2	62.6	75.9	68.1	66.8	67.3	89.9	83.1	75.4	78.4
MRI	SVC	59.5	52.5	50.2	51.1	68.2	64.8	62.8	63.2	76.4	66.8	64.8	65.6	92.1	86.3	80.8	83.2
	RF	63.3	56.8	49	50.5	69.2	66.4	62.2	62.4	80.3	73.3	67.4	69.3	92.7	87.6	82.4	84.7
	XGB	62	56.8	49.2	51	68.7	65.4	62.8	63.2	79.2	71.2	67.7	69	91	84.6	77.1	80.2
Amyloid-β PET & tau PET	SVC	67.1	61.7	54.8	56.5	73.8	73.8	65.6	66.4	77	68.8	64.9	66.1	92.5	89.4	79.9	83.6
	RF	64.9	59.1	51.6	53.6	72.3	70.2	65.1	65.9	77	68.8	63.4	64.8	91.2	87.7	75.6	80
	XGB	64.2	56.4	51.5	52.7	70	66.7	63.7	64.3	75.9	67.1	64.1	65.1	90.6	84.4	76.1	79.4
Amyloid-β PET & tau PET & MRI	SVC	71.5	66.5	58.5	61.2	75.9	73.6	68.7	70.0	82.4	76.6	69.5	71.8	93.3	88.9	79.4	83.2
	RF	70.7	64.3	51.2	53.6	77.7	76.6	70.3	71.8	81.7	76.9	65.9	68.4	90.6	80.5	74	76.7
	XGB	69.9	62.9	55	57.3	75.6	73.1	68.5	69.7	80.3	73	65	67	91.8	86.4	73.3	77.9

CN, cognitively normal; MCI, mild cognitive impairment; AD, Alzheimer's disease; ACC, accuracy; PRE, precision; REC, recall; F1, F1-score; Amyloid-β PET, SUVR values with AV45 tracer; Tau PET, SUVR values with AV1451 tracer; MRI, cortical thickness.

thus, in multimodal cases, the feature space is of high dimensionality, and the model could not handle it effectively. This issue is reinvestigated in the next section, where the feature selection is applied before fitting the models.

The classification scores with feature selection are shown in Table 5. The SVC results have improved in most cases, while the RF and XGB results have not changed significantly since these two algorithms have an embedded feature selection process and are not affected substantially by external feature selection. Figure 8 shows the feature selection effect on SVC and XGB F1-scores for three scenarios. In most cases, SVC with feature selection yields the highest scores, which proves the effectiveness of the proposed feature selection approach. Next, Fig. 9 compares the individual modality and multimodality results. In the single modality classification, tau PET has higher scores, specifically in the CN versus MCI case. This proves the effectiveness of tau PET compared to amyloid PET and MRI in mild cognitive impairment diagnosis, which conforms with previous studies [21]. Generally, multimodal data enhances the scores, which is more notable when feature selection is applied.

To investigate the effect of age, gender, *APOE* ε4, and education on the classification performance, we added them to the model variables and repeated the experiments using the best-performing model and top regional features. Figure 10 presents the classification scores with and without the covariates age, gender, *APOE4*, and education. In most cases, the

Uncorrected Author Proof

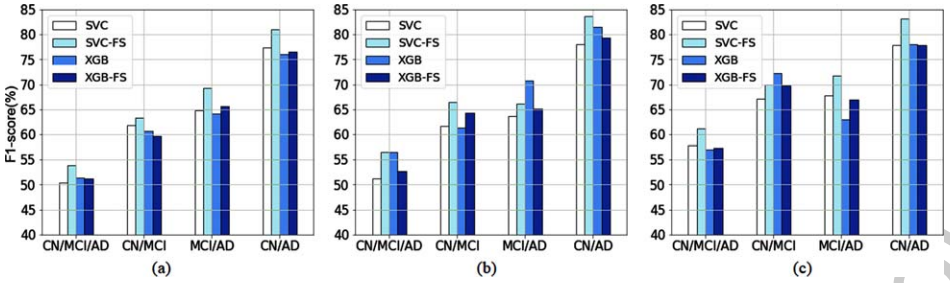

Fig. 8. Classification F1-score before and after feature selection (FS) using two machine learning models, SVC and XGB, for different classification scenarios including CN/MCI/AD, CN/MCI, MCI/AD, and CN/AD; (a) Single modality; amyloid PET, (b) Multimodality; tau and amyloid PET, (c) Multimodality; tau and amyloid PET and MRI.

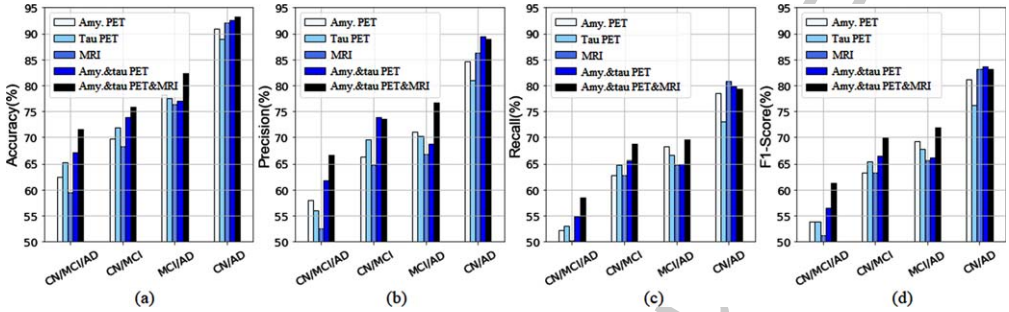

Fig. 9. Classification scores for single-modal and multimodal scenarios after feature selection; (a) Accuracy, (b) Precision, (c) Recall, (d) F1-score.

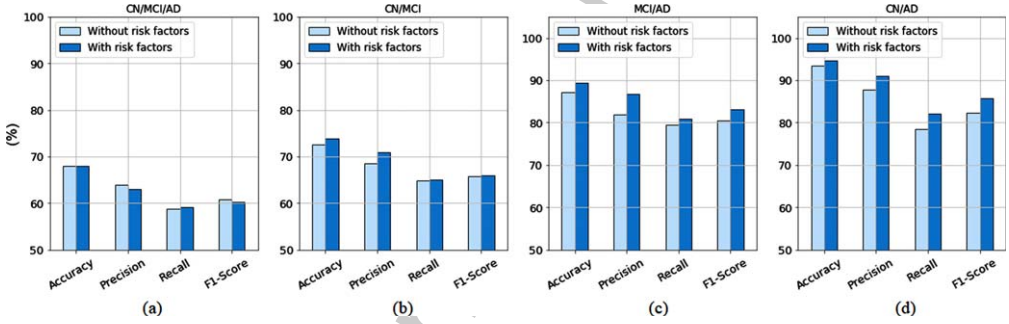

Fig. 10. Classification scores with and without the covariates age, gender, APOE4, and education using the SVC model and top selected features, for classification tasks (a) CN/MCI/AD, (b) CN/MCI, (c) MCI/AD, (d) CN/AD.

classification scores increased. The binary classification cases, MCI/AD and CN/AD, experienced the highest performance improvement which can be due to the higher interclass variance of covariates such as age for these classes. On the other hand, the scores for the three-class classification case, CN/MCI/AD, remained almost unchanged, which can be due to the lower interclass variance of age between the CN and MCI classes and also the more complex nature of the multiclass classification task.

Biomarker profile grouping

The merit of using the National Institute on Aging-Alzheimer's Association AT(N) framework was examined to address the challenge in ascertaining discrepancies between cognitive stage (determined clinically) and biological AD (determined by the classification model using biomarkers). Biomarker profiles were thus defined based on amyloid/tau/neurodegeneration (A/T/N) positivity and negativity,

Table 6
Grouping the study participants into AT(N) biomarkers categories and their corresponding clinically diagnosed cognitive stage (CN, MCI, and AD). The AT(N) groups are defined using two different cut-points for each biomarker. Confident cut-points {1.11, 1.32, 2.57} and conservative cut-points {1.11, 1.19, 2.69} were used for amyloid SUVRs, tau SUVRs, and MRI cortical thickness, respectively. The distribution of subjects shows that in each biomarker profile specifically for the preclinical AD group (A+T+N– and A+T+N+), subjects can belong to any of the three cognitive stages, which is due to the heterogeneity of the disease. This results in a more challenging classification of the cognitive stage. For the confident cut-points, more subjects are categorized in the A–T–N– and A+T–N– groups, while for the conservative cut-points, groups with more positive biomarkers include a larger number of subjects. This is expected as the confident cut-point case has a larger threshold for tau SUVR and a smaller threshold for cortical thickness compared to the conservative cut-point case

	Clinically diagnosed cognitive stage					
	Confident cut-points			Conservative cut-points		
	CN	MCI	AD	CN	MCI	AD
A–T–N–	82	38	2	56	23	1
A+T–N–	41	15	5	23	9	2
A+T+N– A+T+N+	9	14	12	22	21	16
A+T–N+	2	3	2	7	2	1
A–T+N– A–T–N+ A–T+N+	4	9	0	30	24	1

CN, Cognitively normal; MCI, Mild cognitive impairment; AD, Alzheimer's disease; A, Aggregated amyloid-β; T, Aggregated tau; N, Neurodegeneration.

Table 7
Grouping the study participants into AT(N) biomarkers categories and their corresponding clinical and predicted cognitive stage (CN, MCI, and AD). The AT(N) groups are defined using confident cut-points {1.11, 1.32, 2.57} for amyloid SUVRs, tau SUVRs, and MRI cortical thickness, respectively. For the normal biomarker profile (A–T–N–), more subjects were predicted as the CN class (compared to the clinical diagnosis) due to the dominance of CN subjects in this specific AT(N) group. The Alzheimer's pathological change group (A+T–N–) experienced a similar but less severe situation than the previous group. In the preclinical AD group (A+T+N– and A+T+N+), all three cognitive classes include a significant portion of subjects for both clinical and predicted cases

	Clinical cognitive stage			Predicted cognitive stage		
	CN	MCI	AD	CN	MCI	AD
A–T–N–	137	52	4	160	33	0
A+T–N–	66	20	8	71	17	6
A+T+N– A+T+N+	13	24	18	15	25	15
A+T–N+	2	4	3	3	2	4
A–T+N– A–T–N+ A–T+N+	5	9	0	6	8	0

CN, Cognitively normal; MCI, Mild cognitive impairment; AD, Alzheimer's disease; A, Aggregated amyloid-β; T, Aggregated tau; N, Neurodegeneration.

as summarized in Table 2. The study participants were categorized according to their biomarker signature and cognitive stage. The total number of subjects falling under each category is reported in Table 6. The numbers are reported for two sets of cut-points: {1.11, 1.32, 2.57} and {1.11, 1.19, 2.69} for {amyloid SUVRs, tau SUVRs, and MRI cortical thickness}, respectively. The former set has a larger cut-point for tau and a smaller cut-point for MRI (confident scenario, resulting in less positive cases) compared to the second set (conservative scenario, with more positive cases). Based on this table, the inconsistencies between the neuropathologic biomarkers and clinical diagnosis can be investigated specifically in challenging categories such as normal AD biomarkers with a dementia diagnosis and preclinical AD with cognitively unimpaired diagnosis. In the studied cohort, the "normal AD biomarker (A–T–N–) with an AD diagnosis" group includes 2 and 1 individuals based on the confident and conservative cut-points, respectively. Although this inconsistency between the biomarkers and clinical diagnosis might be partially caused by inaccurate binary biomarker grouping, it can potentially be one of the contributors to misclassification. Another controversial case is related to individuals with "preclinical Alzheimer's disease biomarkers" (A+T+N– and A+T+N+). As seen in Table 6, this group has a considerable number of subjects in all three cognitive stages making the classification task even more challenging.

To further investigate this scenario, we reconstructed the AT(N) biomarker-cognition table for the predicted cognitive stage aside from the clinically diagnosed cognitive stage. Table 7 represents the results for the clinical and predicted diagnosis side by side. It should be noted that here we used a different case study than Table 6. As can be seen from the results, for the normal biomarker group (A–T–N–), all dementia subjects and some of the MCI subjects were misclassified as the CN group (false negative). A less severe outcome is seen for the AD pathological change group (A+T–N–), where some AD and MCI subjects were misclassified as CN. As for the challenging preclinical AD group (A+T+N– and A+T+N+), a clear conclusion cannot be drawn solely from Table 7. Thus, a classification confusion matrix was constructed for the specific case of preclinical AD, as shown in Table 8. From this table, it is clear that many CN subjects were misclassified as MCI,

Uncorrected Author Proof

Table 8
Classification confusion matrix for the AT(N) preclinical AD group (biomarker profiles A+T+N– and A+T+N+). For the CN class (true label), a significant portion of subjects (6 out of 13) was classified (predicted label) as MCI and AD, which can be related to those preclinical AD individuals that have not yet advanced to AD. On the other hand, a considerable number of AD subjects (true label) were classified (predicted label) as MCI and CN, which could belong to those AD subtypes with a different pattern and less severe biomarker levels. Overall, the classification scores for this preclinical AD category are: accuracy = 56.4%, precision = 57.3%, recall = 56.4%, f1-score = 55.5%

True/Pred	CN	MCI	AD
CN	**7**	4	2
MCI	7	**14**	3
AD	1	7	**10**

and a large number of AD subjects were misclassified as MCI.

DISCUSSION

The objective of this research was to determine the cognitive stage using neuroimaging biomarkers and analyze the dependencies between biomarker profiles and the cognitive stage. For the model variables, including amyloid and tau PET SUVR values and cortical thickness, a trade-off was made between variables relevance and redundancy using an information theory-based metric. The advantage of the proposed approach is to incorporate the effect of features complementarity and redundancy to maximize the total amount of information in the feature set. It is important to note that the redundancy part should not be overweighted since highly relevant features can also be partially redundant. This situation is seen in Fig. 5 for larger values of the coefficient α, where feature relevance is sacrificed for even a minor redundancy. By incorporating a moderate redundancy coefficient into the equations, for tau SUVRs, entorhinal and amygdala were among the top regions for all stages of AD, with amygdala being most informative for the CN/MCI case. Abnormal tau deposition in these regions is known as a biomarker for preclinical AD by previous studies [18, 20, 44]. It is reported in the literature that amygdala shows early atrophy independent of amyloid deposition, and it might be related to neurofibrillary tangles instead [45, 46]. Other prominent regions include pallidum and hippocampus based on tau PET for CN/MCI case, and inferior parietal, inferior temporal, precuneus, and precentral for the MCI/AD case. It is stated in [47–49] that tau burden in these specific ROIs is correlated with cognitive decline. On the other hand, for amyloid PET SUVRs, frontal pole for all stages, and inferior lateral ventricle for the CN/MCI case, and medial orbitofrontal, pars triangularis, and rostral anterior cingulate for the MCI/AD case are among the more prominent variables. These findings are consistent with previous studies [50–52].

By incorporating the effect of redundancy and synergy, some features experienced a score change. For instance, the score of frontal pole amyloid SUVR (but not tau SUVR) for the early stage increased significantly, so that this region is considered a complementary variable for the classification task. This is in agreement with the literature [45, 53], where it is reported that the frontal pole shows early amyloid deposition while atrophy and tau deposition are later events. Some amyloid and tau SUVR values that experienced a boost in their score include the hippocampus, inferior lateral ventricle, and lateral ventricle, which are known to be critical for AD diagnosis in previous studies. On the other hand, a score drop was seen in some of the tau SUVRs, including fusiform, inferior parietal, inferior temporal, isthmus cingulate, orbitofrontal, middle temporal, precuneus, and bankssts. A lower score does not necessarily disqualify a feature. Instead, the model tries to replace the most redundant features with a possibly less relevant but complementary one so that additional information is added to the analysis.

In the classification part, tau PET modality produced more accurate results than amyloid PET and MRI modalities, specifically in CN/MCI classification (early stage). On the other hand, multimodal scenarios have achieved the highest F1-scores in most cases, especially in the early stages of the disease. Feature selection was most effective in the SVC case, making SVC achieve higher scores compared to RF and XGB in many cases. This was expected as RF and XGB have internal feature selection, with less room for improvement. In retrospect, these findings suggest that the classification of high-dimensional multimodal datasets would be most accurate when feature selection is carried out most effectively, with the relevance of each feature quantified through a ranking score metric as proposed in this study. When such measures are taken, reducing the dimensionality of the feature space can be accomplished while still maintaining high accuracy in the classification results. More specifically, Fig. 9d shows that the F1-score of the multimodal case with feature selection is up to 5% higher than other scenarios.

One of the major challenges in the AD diagnosis is the heterogeneity of the disease related

Uncorrected Author Proof

to the AD subtypes (hippocampal-sparing, limbic-predominant, typical AD). It is shown that the AD risk factors and protective factors have a meaningful variance among the AD subtypes [54]. As seen in the Result section, the inclusion of these covariates into the model variables could improve the classification scores. This can be explained through the characteristics of different subtypes and the variation of risk factors among them. Typical AD subtype cases experience more severe pathology compared to other subtypes, while limbic-predominant cases have more typical biomarkers than hippocampal-sparing subjects. Since typical AD is more prevalent than other subtypes, if the classification model only relies on biomarkers, it might be biased toward this group and yields false-negative results for other AD subtypes as they have less severe biomarkers and are less prevalent. Therefore, these other categories of subjects with minimal atrophy and non-typical biomarkers might be misclassified as CN and MCI classes. At this stage, the risk and protective factors can complement the biomarkers and help to correctly classify these subtypes as the AD group and thus alleviate the heterogeneity issue. Concerning the risk factors, subjects with typical and limbic-predominant AD tend to be older than those with hippocampal-sparing AD. On the other hand, the hippocampal-sparing category includes fewer *APOE4* carriers and highly educated individuals compared to other groups. In terms of gender, females are more frequent in the limbic-predominant group.

As described in this study, another challenge in the classification problems is biomarker insufficiency. This may result in a disconnection between biomarkers and clinical diagnosis to some extent. Studies revealed that almost 30% of clinically unimpaired elderly participants have AD in postmortem examinations or have abnormal amyloid deposition [24, 43]. In our study, in one of the scenarios (Table 6), 6.5%–16% (9–22 individuals) of the CN group have preclinical AD with abnormal amyloid and tau pathology for the two cut-point levels, as seen in Table 6. It is anticipated that the classification model classifies some of these individuals as MCI or AD groups since both AD-specific biomarkers (amyloid and tau) are abnormal in this case (false positive). This was confirmed in Table 8, where almost half of the CN subjects were misclassified as MCI and AD. Moreover, for the same preclinical AD group, a large number of AD subjects were misclassified. This can be explained by the heterogeneity of AD, where some AD subjects with less severe biomarkers are predicted by the model as non-AD and vice versa. The results proved the preclinical AD subjects to be one of the most challenging groups for the model, with a classification accuracy of 56%, which is lower than the overall accuracy of 65% for all subjects of the scenario presented in Table 7. These outcomes were expected since the preclinical biomarker profile includes subjects in all three cognitive stages which is due to the heterogeneity of the disease and the lack of sufficient biomarkers required for a more accurate delineation of the classes. Similarly, the "normal AD biomarker" (A–T–N–) and "non-Alzheimer's pathologic change" (A–) groups are also susceptible to misclassification as they have non-AD-specific biomarkers, but some are labeled as MCI (AD prodromal stage) and AD in the ADNI dataset. It has been shown in other studies that 10% to 30% of clinically diagnosed AD cases do not have AD at autopsy or have normal AD biomarkers [24, 43]. In the ADNI cohort used in our study, 10–20% of subjects were detected with the described condition. In the classification process, the normal biomarkers are likely to predict a cognitively normal stage rather than AD (false negative). These results can be explained by the fact that the clinical diagnosis and cognitive labeling practices are generally based on symptoms and are independent of the biomarkers. The outcomes reveal the insufficiency of the available biomarkers in making an accurate prediction of the clinically defined cognitive stage.

Since the biomarkers might not be accessible in many situations, clinical diagnosis is made solely based on symptoms as ascertained through cognitive tests. The AT(N) biomarker framework establishes a biomarker-based definition of AD and emphasizes the independence of the biological and clinical definitions of AD, yet it tries to clarify the interaction between the two. This can be valuable for in-depth research purposes as well as personalized medicine. The AT(N) framework shows that the cognitive stage cannot be entirely determined through the AT(N) biomarkers since any particular biomarker profile can belong to any cognitive stage. The fact that a wide range of biomarker profiles can define a specific cognitive stage is due to the heterogeneity of the disease, which can be explained by the subtypes of AD (hippocampal-sparing, limbic-predominant, typical AD). Different subtypes have similar amyloid loads; however, tau and neurodegeneration pathology and also concomitant non-AD pathologies vary across subtypes. Also, other contributing factors to differentiate between AD subtypes include risk factors (age,

Uncorrected Author Proof

gender, education, and *APOE*) and protective factors (cognitive reserve, brain resilience, and brain resistance). Incorporation of these factors in the context of the AT(N) system can be a step toward a more in-depth analysis of the computer-aided diagnosis of AD and augmenting the research prospects for more effectual personalized medicine.

One of the limiting factors for our analysis was the considerable amount of missing data, specifically for the tau PET modality. This issue is more critical when we are interested in subjects with all modalities available, which is a requirement for having a fair comparison between single modality scenarios. Also, the study could be more valuable if longitudinal data were available so that the effect of biomarker change through time could be considered. Longitudinal tau PET data is very limited in the ADNI dataset since tau PET is a relatively new technology, and its longitudinal data collection and processing is still in progress. Also, the missing data issue is even more severe for the longitudinal data. Moreover, in the data collection process, a time difference may exist between capturing the MRI and PET scans for some participants. This time lag between modalities is inevitable in many situations in practice. While small time-lags might be neglected in some studies, more significant delays can be included in the analysis with appropriate considerations. In our study, we have not integrated this variable in our analysis due to the lack of such information for some of the participants, which would result in additional missing values for the dataset. In this study, we conducted a cross-sectional study and handled the missing values by mean-value imputation and by making use of models that are more robust to missing values. Moreover, using the AT(N) analysis, the intra-class biomarker variance was studied so that the contribution of biomarker shortage on the classification performance was determined.

ACKNOWLEDGMENTS

This research is supported by the National Science Foundation under grants: CNS-1920182, CNS-1532061, CNS-1338922, CNS-2018611, and CNS-1551221, and with the National Institutes of Health through NIA/NIH grants 1R01AG055638-01A1, 5R01AG061106-02, 5R01AG047649-05, and the 1P30AG066506-01 with the 1Florida Alzheimer's Disease Research Center (ADRC).

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/21-0064r3).

REFERENCES

[1] Alzheimer's Association (2019) 2019 Alzheimer's disease facts and figures. *Alzheimers Dement* **15**, 321-387.

[2] Tabarestani S, Aghili M, Eslami M, Cabrerizo M, Barreto A, Rishe N, Curiel R, Loewenstein D, Duara R, Adjouadi M (2020) A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study. *Neuroimage* **206**, 116317.

[3] Liu S, Liu S, Cai W, Che H, Pujol S, Kikinis R, Feng D, Fulham M, ADNI (2015) Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans Biomed Eng* **62**, 1132-1140.

[4] Ewers M, Sperling R, Klunk W, Weiner M, Hampel H (2011) Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. *Trends Neurosci* **34**, 430-442.

[5] Dubois B, Hampel H, Feldman H, Scheltens P, Aisen P, Andrieu S, Bakardjian H, Benali H, Bertram L, Blennow K, Broich K, Cavedo E, Crutch S, Dartigues J, Duyckaerts C, Epelbaum S, Frisoni G, Gauthier S, Genthon R, Gouw A, Habert M, Holtzman D, Kivipelto M, Lista S, Molinuevo J, O'Bryant S, Rabinovici G, Rowe C, Salloway S, Schneider L, Sperling R, Teichmann M, Carrillo M, Cummings J, Jack C (2016) Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimers Dement* **12**, 292-323.

[6] Loewenstein D, Curiel R, DeKosky S, Bauer R, Rosselli M, Guinjoan S, Adjouadi M, Peñate A, Barker W, Goenaga S, Golde T, Greig-Custo M, Hanson K, Li C, Lizarraga G, Marsiske M, Duara R (2018) Utilizing semantic intrusions to identify amyloid positivity in mild cognitive impairment. *Neurology* **91**, e976-e984.

[7] Albert M, DeKosky S, Dickson D, Dubois B, Feldman H, Fox N, Gamst A, Holtzman D, Jagust W, Petersen R, Snyder P, Carrillo M, Thies B, Phelps C (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270-279.

[8] Misra C, Fan Y, Davatzikos C (2009) Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI. *Neuroimage* **44**, 1415-1422.

[9] Selkoe D (2019) Early network dysfunction in Alzheimer's disease. *Science* **365**, 540-541.

[10] Ridgway G, Lehmann M, Barnes J, Rohrer J, Warren J, Crutch S, Fox N (2012) Early-onset Alzheimer disease clinical variants: Multivariate analyses of cortical thickness. *Neurology* **79**, 80-84.

[11] Bergeron M, Landset S, Zhou X, Ding T, Khoshgoftaar T, Zhao F, Du B, Chen X, Wang X, Zhong L, Liu X, Ashford J (2020) Utility of MemTrax and machine learning modeling in classification of mild cognitive impairment. *J Alzheimers Dis* **77**, 1545-1558.

[12] Ezzati A, Zammit A, Harvey D, Habeck C, Hall C, Lipton R (2019) Optimizing machine learning methods to improve predictive models of Alzheimer's disease. *J Alzheimers Dis* **71**, 1027-1036.

[13] Gill S, Mouches P, Hu S, Rajashekar D, MacMaster F, Smith E, Forkert N, Ismail Z (2020) Using machine learning to predict dementia from neuropsychiatric symptom and neuroimaging data. *J Alzheimers Dis* **75**, 277-288.

Uncorrected Author Proof

[14] Vecchio F, Miraglia F, Alù F, Menna M, Judica E, Cotelli M, Rossini P (2020) Classification of Alzheimer's disease with respect to physiological aging with innovative EEG biomarkers in a machine learning implementation. *J Alzheimers Dis* **75**, 1253-1261.

[15] Villemagne V, Fodero-Tavoletti M, Masters C, Rowe C (2015) Tau imaging: Early progress and future directions. *Lancet Neurol* **14**, 114-124.

[16] Johnson K, Schultz A, Betensky R, Becker J, Sepulcre J, Rentz D, Mormino E, Chhatwal J, Amariglio R, Papp K, Marshall G, Albers M, Mauro S, Pepin L, Alverio J, Judge K, Philiossaint M, Shoup T, Yokell D, Dickerson B, Gomez-Isla T, Hyman B, Vasdev N, Sperling R (2015) Tau positron emission tomographic imaging in aging and early Alzheimer disease. *Ann Neurol* **79**, 110-119.

[17] Ittner L, Götz J (2010) Amyloid-β and tau — a toxic pas de deux in Alzheimer's disease. *Nat Rev Neurosci* **12**, 67-72.

[18] Leuzy A, Chiotis K, Lemoine L, Gillberg P, Almkvist O, Rodriguez-Vieitez E, Nordberg A (2019) Tau PET imaging in neurodegenerative tauopathies—still a challenge. *Molr Psychiatry* **24**, 1112-1134.

[19] De Wolf F, Ghanbari M, Licher S, McRae-McKee K, Gras L, Weverling G, Wermeling P, Sedaghat S, Ikram M, Waziry R, Koudstaal W, Klap J, Kostense S, Hofman A, Anderson R, Goudsmit J, Ikram M (2020) Plasma tau, neurofilament light chain and amyloid-β levels and risk of dementia; a population-based cohort study. *Brain* **143**, 1220-1232.

[20] Schöll M, Maass A (2019) Does early cognitive decline require the presence of both tau and amyloid-β? *Brain* **143**, 10-13.

[21] Ossenkoppele R, Smith R, Ohlsson T, Strandberg O, Mattsson N, Insel P, Palmqvist S, Hansson O (2019) Associations between tau, Aβ, and cortical thickness with cognition in Alzheimer disease. *Neurology* **92**, e601-e612.

[22] Jedynak B, Lang A, Liu B, Katz E, Zhang Y, Wyman B, Raunig D, Jedynak C, Caffo B, Prince J (2012) A computational neurodegenerative disease progression score: Method and results with the Alzheimer's disease neuroimaging initiative cohort. *Neuroimage* **63**, 1478-1486.

[23] Jack C, Knopman D, Jagust W, Shaw L, Aisen P, Weiner M, Petersen R, Trojanowski J (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* **9**, 119-128.

[24] Jack C, Bennett D, Blennow K, Carrillo M, Dunn B, Haeberlein S, Holtzman D, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo J, Montine T, Phelps C, Rankin K, Rowe C, Scheltens P, Siemers E, Snyder H, Sperling R, Elliott C, Masliah E, Ryan L, Silverberg N (2018) NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14**, 535-562.

[25] 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.

[26] Shi Y, Suk H, Gao Y, Lee S, Shen D (2020) Leveraging coupled interaction for multimodal Alzheimer's disease diagnosis. *IEEE Trans Neural Netw Learn Syst* **31**, 186-200.

[27] Tabarestani S, Aghili M, Shojaie M, Freytes C, Adjouadi M (2018) Profile-specific regression model for progression prediction of Alzheimer's disease using longitudinal data. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*.

[28] Tabarestani S, Aghili M, Shojaie M, Freytes C, Cabrerizo M, Barreto A, Rishe N, Curiel R, Loewenstein D, Duara R, Adjouadi M (2019) Longitudinal prediction modeling of Alzheimer disease using recurrent neural networks. *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*.

[29] An L, Adeli E, Liu M, Zhang J, Lee S, Shen D (2017) A hierarchical feature and sample selection framework and its application for Alzheimer's disease diagnosis. *Sci Rep* **7**, 45269.

[30] Shi J, Zheng X, Li Y, Zhang Q, Ying S (2018) Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. *IEEE J Biomed Health Inform* **22**, 173-183.

[31] Hao X, Bao Y, Guo Y, Yu M, Zhang D, Risacher S, Saykin A, Yao X, Shen L (2020) Multimodal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease. *Med Image Anal* **60**, 101625.

[32] Zhu X, Suk H, Wang L, Lee S, Shen D (2017) A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. *Med Image Anal* **38**, 205-214.

[33] Tohka J, Moradi E, Huttunen H (2016) Comparison of feature selection techniques in machine learning for anatomical brain MRI in dementia. *Neuroinformatics* **14**, 279-296.

[34] Zhu X, Suk H, Lee S, Shen D (2015) Canonical feature selection for joint regression and multiclass identification in Alzheimer's disease diagnosis. *Brain Imaging Behav* **10**, 818-828.

[35] Vergara J, Estévez P (2013) A review of feature selection methods based on mutual information. *Neural Comput Appl* **24**, 175-186.

[36] Saeys Y, Inza I, Larranaga P (2007) A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**, 2507-2517.

[37] Garali I, Adel M, Bourennane S, Guedj E (2018) Histogram-based features selection and volume of interest ranking for brain PET image classification. *IEEE J Transl Eng Health Med* **6**, 2100212.

[38] Zhu X, Suk H, Lee S, Shen D (2016) Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification. *IEEE Trans Biomed Eng* **63**, 607-618.

[39] Liu M, Zhang D, Shen D (2016) Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment. *IEEE Trans Med Imaging* **35**, 1463-1474.

[40] Xu L, Yao Z, Li J, Lv C, Zhang H, Hu B (2019) Sparse feature learning with label information for Alzheimer's disease classification based on magnetic resonance imaging. *IEEE Access* **7**, 26157-26167.

[41] Jiang P, Wang X, Li Q, Jin L, Li S (2019) Correlation-aware sparse and low-rank constrained multi-task learning for longitudinal analysis of Alzheimer's disease. *IEEE J Biomed Health Inform* **23**, 1450-1456.

[42] Landau S, Mintun M, Joshi A, Koeppe R, Petersen R, Aisen P, Weiner M, Jagust W (2012) Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Ann Neurol* **72**, 578-586.

[43] Jack C, Wiste H, Weigand S, Therneau T, Lowe V, Knopman D, Gunter J, Senjem M, Jones D, Kantarci K, Machulda M, Mielke M, Roberts R, Vemuri P, Reyes D, Petersen R (2016) Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers & Dement* **13**, 205-216.

[44] Mishra S, Gordon B, Su Y, Christensen J, Friedrichsen K, Jackson K, Hornbeck R, Balota D, Cairns N, Morris J, Ances B, Benzinger T (2017) AV-1451 PET imaging of tau pathology in preclinical Alzheimer disease: Defining a summary measure. *Neuroimage* **161**, 171-178.

Uncorrected Author Proof

[45] Jorge L, Martins R, Canário N, Xavier C, Abrunhosa A, Santana I, Castelo-Branco M (2021) Investigating the spatial associations between amyloid-β deposition, grey matter volume, and neuroinflammation in Alzheimer's disease. *J Alzheimers Dis* **80**, 113-132.

[46] Wang W, Yu J, Liu Y, Yin R, Wang H, Wang J, Tan L, Radua J, Tan L (2015) Voxel-based meta-analysis of grey matter changes in Alzheimer's disease. *Transl Neurodegener* **4**, 6.

[47] Jack C, Wiste H, Schwarz C, Lowe V, Senjem M, Vemuri P, Weigand S, Therneau T, Knopman D, Gunter J, Jones D, Graff-Radford J, Kantarci K, Roberts R, Mielke M, Machulda M, Petersen R (2018) Longitudinal tau PET in ageing and Alzheimer's disease. *Brain* **141**, 1517-1528.

[48] Maass A, Landau S, Baker S, Horng A, Lockhart S, La Joie R, Rabinovici G, Jagust W (2017) Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease. *Neuroimage* **157**, 448-463.

[49] Lowe V, Wiste H, Senjem M, Weigand S, Therneau T, Boeve B, Josephs K, Fang P, Pandey M, Murray M, Kantarci K, Jones D, Vemuri P, Graff-Radford J, Schwarz C, Machulda M, Mielke M, Roberts R, Knopman D, Petersen R, Jack C (2017) Widespread brain tau and its association with ageing, Braak stage and Alzheimer's dementia. *Brain* **141**, 271-287.

[50] Liu Y, Yan W, Tan C, Li J, Xu W, Cao X, Tan L, Yu J (2019) Common variant in TREM1 influencing brain amyloid deposition in mild cognitive impairment and Alzheimer's disease. *Neurotox Res* **37**, 661-668.

[51] Chetelat G, Villemagne V, Bourgeat P, Pike K, Jones G, Ames D, Ellis K, Szoeke C, Martins R, O'Keefe G, Salvado O, Masters C, Rowe C (2010) Relationship between atrophy and beta-amyloid deposition in Alzheimer disease. *Ann Neurol* **67**, 317-324.

[52] Crocco E, Loewenstein D, Curiel R, Alperin N, Czaja S, Harvey P, Sun X, Lenchus J, Raffo A, Peñate A, Melo J, Sang L, Valdivia R, Cardenas K (2018) A novel cognitive assessment paradigm to detect Pre-mild cognitive impairment (PreMCI) and the relationship to biological markers of Alzheimer's disease. *J Psychiatr Res* **96**, 33-38.

[53] Schwindt GC, Black SE (2009) Functional imaging studies of episodic memory in Alzheimer's disease: A quantitative meta-analysis. *Neuroimage* **45**, 181-190.

[54] Ferreira D, Nordberg A, Westman E (2020) Biological subtypes of Alzheimer disease. *Neurology* **94**, 436-448.

Uncorrected Author Proof