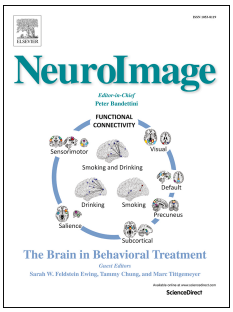
Journal Pre-proof



A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study

Solale Tabarestani, Maryamossadat Aghili, Mohammad Eslami, Mercedes Cabrerizo, Armando Barreto, Naphtali Rishe, Rosie E. Curiel, David Loewenstein, Ranjan Duara, Malek Adjouadi

PII: S1053-8119(19)30908-5

DOI: https://doi.org/10.1016/j.neuroimage.2019.116317

Reference: YNIMG 116317

To appear in: *NeuroImage*

Received Date: 27 April 2019

Revised Date: 24 October 2019

Accepted Date: 26 October 2019

Please cite this article as: Tabarestani, S., Aghili, M., Eslami, M., Cabrerizo, M., Barreto, A., Rishe, N., Curiel, R.E., Loewenstein, D., Duara, R., Adjouadi, M., A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study, *NeuroImage* (2019), doi: https://doi.org/10.1016/j.neuroimage.2019.116317.

A Distributed Multitask Multimodal Approach for the Prediction of Alzheimer's Disease in a Longitudinal Study

Solale Tabarestani[1], Maryamossadat Aghili[1], Mohammad Eslami[2] , Mercedes Cabrerizo[1], Armando Barreto[1], Naphtali Rishe[3], Rosie E. Curiel[4,5], David Loewenstein[4,5,6], Ranjan Duara[5,6], and Malek Adjouadi[1,5]

1 Center for Advanced Technology and Education (CATE), Florida International University, 10555 W Flagler St., Miami, Florida, USA
2 RWTH University Hospital, Pauwelsstrasse 30, 52074 Aachen, Germany
3 School of Computer and Information Sciences, Florida International University, 10555 W Flagler St., Miami, Florida, USA
4 Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, Florida, USA
5 1Florida Alzheimer's Disease Research Center (ADRC), University of Florida, Gainesville, Florida, USA
6 Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, Florida, USA.

ARTICLEINFO

Keywords:
Alzheimer's Disease
Multitask learning
Multimodal regression
Longitudinal study
Missing values
Progression
Gradient boosting
Fused sparse group Lasso

ABSTRACT

Predicting the progression of Alzheimer's Disease (AD) has been held back for decades due to the lack of sufficient longitudinal data required for the development of novel machine learning algorithms. This study proposes a novel machine learning algorithm for predicting the progression of Alzheimer's disease using a distributed multimodal, multitask learning method. More specifically, each individual task is defined as a regression model, which predicts cognitive scores at a single time point. Since the prediction tasks for multiple intervals are related to each other in chronological order, multitask regression models have been developed to track the relationship between subsequent tasks. Furthermore, since subjects have various combinations of recording modalities together with other genetic, neuropsychological and demographic risk factors, special attention is given to the fact that each modality may experience a specific sparsity pattern. The model is hence generalized by exploiting multiple individual multitask regression coefficient matrices for each modality. The outcome for each independent modality-specific learner is then integrated with complementary information, known as risk factor parameters, revealing the most prevalent trends of the multimodal data. This new feature space is then used as input to the gradient boosting kernel in search for a more accurate prediction. This proposed model not only captures the complex relationships between the different feature representations, but it also ignores any unrelated information which might skew the regression coefficients. Comparative assessments are made between the performance of the proposed method with several other well-established methods using different multimodal platforms. The results indicate that by capturing the interrelatedness between the different modalities and extracting only relevant information in the data, even in an incomplete longitudinal dataset, will yield minimized prediction errors.

1. Introduction

According to a March 2018 report from the Alzheimer's Association (AA), nearly 5.7 million US citizens, mostly elderly people, are affected by AD, a statistic that is predicted to reach 13.8 million by 2050. This AA report also indicates that an approximated amount of 277 billion dollars was invested in 2018 in caretaking services for patients with AD and dementia (Alzheimer Association, 2016).

Alzheimer's Disease is a progressive and irreversible brain disorder where subtle brain changes may have started decades prior to any detectable symptoms. In its early stages, AD symptoms begin with mild cognitive decline, which can then progressively lead to more severe physical and functional impairments. Key indicators are associated with severe brain atrophy, beta-amyloid deposition, and evidence of widespread limbic and cortical neurofibrillary

degeneration. In the study by (Jedynak et al., 2012), an interesting computational neurodegenerative disease progression score is proposed on the basis of the dynamics of the different biomarkers in AD.

Alzheimer's Disease progression is generally assessed using clinical measures, but it can also be accomplished using biomarkers involving structural magnetic resonance imaging (MRI), 18-Fluoro-DeoxyGlucose PET imaging (FDG-PET), cognitive examination, cerebrospinal fluid (CSF) and electroencephalography (EEG) (Nimmy John et al., 2018; Poil et al., 2013; Loewenstein et al. 2018). Commonly used MRI biomarkers for detecting the progression of AD include cortical thickness and regional brain volume (Stonnington et al., 2010; Lao et al., 2004; Magnin et al., 2009; Sørensen et al., 2016), whereas the most significant biomarkers of FDG-PET include glucose hypometabolism in neocortical brain regions (Azmi et al., 2017; Alexander et al., 2002; Landau et al., 2012; Cohen and Klunk, 2015). It has also been revealed that an increase in CSF t-tau or Phospho-Tau is a potential biomarker of disease progression (Trushina et al., 2013; Colijn and Grossberg, 2015; Shaw et al., 2009).

Along with neuroimaging modalities, there are other unconventional measurements, known as risk factors, which are associated with Alzheimer's, such as age, genetic information, years of education and ethnicity (Michaelson, 2014; Rogers et al., 2012). As expected, this complementary information shows that age plays a significant role in the onset of AD (Chen et al., 2000; Mungas et al., 2001; Duara et al., 2019). It is also well acknowledged that the most prominent genetic risk factor is the Apolipoprotein E (APOE) gene. This gene and its major alleles (E2, E3, and E4) are known to increase the risk of developing AD in individuals as young as 40 years of age (Farrer et al., 1997; Corder et al., 2008).

While many studies in the literature mainly focus on disease prediction, typically relying on a single modality (Bi et al., 2018; Frisoni et al., 2007; Duchesne et al., 2009; Li et al., 2012; Buerger et al., 2002; Jack et al., 2018), recent studies have shown that incorporating biomarkers from different modalities may lead to a more accurate diagnosis (De Leon et al., 2006; Tong et al., 2017; Ritter et al., 2015; Westman et al., 2012; Zhang et al, 2011). New research directions have come to rely on multimodal neuroimaging data with the inclusion of other biomarkers such as cerebral spinal fluid (CSF), genetics and neuropsychological testing. The main objectives of these research endeavors are either to discriminate patients' status via classification methods or to predict different variables using regression models. Cross-sectional and longitudinal data have been used to explore correlations between clinical neuroimaging tests, neurological exams and biochemical measurements to monitor changes in these important biomarkers. Yet, despite much ongoing research, predicting the progression of AD, especially for enabling early intervention, has remained challenging (Mendez, 2017; Pierce et al., 2017; Lawlor et al., 1994; Wolfe, 2016; Doody et al., 2010; Van Der Flier and Scheltens, 2009; Moradi et al., 2015; Curiel et al., 2018; Lizarraga et al., 2018; C. Li et al., 2017; Loewenstein et al., 2017; Sargolzaei et al., 2015; Duara et al., 2015; Minhas et al., 2017).

In order to study the relative temporal changes in AD, there is need to track pathophysiological changes in a large number of observations using Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Cognitive assessment tests (COG) and Cerebrospinal Fluid (CSF) tests. However, acquiring all these tests within a large population is costly, time-consuming and often difficult to maintain high protocol adherence given the dropout rate and missed follow-up visits given the patients' advanced age and severity and extent of disease progression. Consequently, there are two kinds of challenges in studying longitudinal dynamics and related patterns in medical data. The first one is due to size irregularity because of missing measurements from a specific modality. The second is due to patients missing on follow-up visits or dropping out from the study. Among the many verified assessments that can diagnose the presence of AD and scale the severity of the progression, the Mini-Mental State Examination (MMSE) and the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) are the most common tests used in regression-based models (Zhang, Daoqiuang; Shen, 2013; Wang et al., 2011). One of the earliest work in this domain was done by Tierney et al. in 1996, who used logistic regression to predict the possibility of AD progression over a period of two years (Tierney et al., 1996). The study in (Zhang and Shen, 2012) proposed a sparse linear regression model in conjunction with a group regularization technique. The model was applied across different brain regions to select the most informative longitudinal features. Their model predicts future cognitive clinical scores among MCI subjects over a period of 24-months. Similarly, Izquierdo et al (Izquierdo et al., 2017) predicted cognitive scores using stochastic gradient boosting of decision trees among 1,141 individuals for whom longitudinal clinical and imaging studies were available in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. In another study (Tabarestani et al., 2019), two different variations of recurrent neural networks (RNN), namely Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) have been applied using 1458 multimodal records of subjects from the ADNI database to predict AD progression. By leveraging the patients' historical records from the previous three time points, their model could track the disease progression trends of patients at three other subsequent time points with an accuracy that outperformed methods that relied solely on the baseline records.

Multitask learning, first proposed in 1997, is shown to improve performance by extracting the relationships between multiple similar tasks through the development of a statistical model (Caruana, 1997). It has since attracted a lot of attention in a variety of machine learning algorithms with application domains ranging from finance to bioinformatics (Dong et al., 2015; Greenlaw et al., 2017). This new research trend has delivered promising performance improvement in different categories, including, but not limited to multitask learning using kernel-methods (Evgeniou et al., 2005), interpreting task relationship (Zhang and Yeung, 2012; Widmer et al., 2012), developing probabilistic and statistical models (Bi et al., 2008; Xue et al., 2007), selecting features (Yang et al., 2010; Zhu et al., 2017), learning features (Zhang and Yeung, 2011; Y. Li et al., 2017), feature hashing (Weinberger et al., 2009), and task grouping (Kumar and Daume, 2012; Bakker and Heskes, 2003).

In recent years, multitask learning has been successfully applied to longitudinal clinical data to predict the progression of neurodegenerative diseases (Zhang, Daoqiuang; Shen, 2013b; Emrani et al., 2017b; Nie et al., 2017; Zhou et al., 2012b; Suk et al., 2017). Compared to single-task learning, multitask learning uses a regression model for predicting the future status of patients at multiple time points. The basic assumption in these models is that an inherent correlation exists among multiple records of information, which are derived from the same subjects. These studies demonstrated that capturing this inherent relatedness could improve the generalization of the final prediction model. For example, Zhou et al. in (Zhou et al., 2012b) developed convex and nonconvex fused group Lasso formulation as the regularization term of the multitask learning kernel. Their model could choose the most important sets of biomarkers from different time points to model the progression of AD. Similarly, Emrani et al. employed multitask learning to predict the progression of Parkinson's disease over a period of 4.5 years (Emrani et al., 2017a), and Jie et al. in (Jie et al., 2015) reported that using manifold regularized multitask feature learning could yield better classification performance and could identify disease-related regions in the brain deemed important for disease diagnosis. A Sparse Group Lasso with shared Subspace Multitask learning (SGLS-MTL) has been proposed by Cao et al. (Cao et al., 2017). Their framework uses $\ell_{2,1}$ penalty, group $\ell_{2,1}$ penalty and subspace structure to capture the correlation between the tasks, the sparse feature representation and the shared subspaces. They have applied their SGLS multitask learning method to predict cognitive scores and to detect potential predictive MRI biomarkers. Wang et al. in (Wang et al., 2012.), proposed a high-order multitask feature learning algorithm to model the longitudinal trajectories of the cognitive measures of AD subjects based on neuroimaging biomarkers. They employed non-smooth structured sparsity-inducing norm to utilize the correlation between the adjacent tasks (prediction of cognitive measures at two subsequent time points) and the interrelations that exist between the cognitive measurements. To capture the nonlinearity in the relationship between MRI neuroimaging features and cognitive scores, Cao et al. in (Cao et al., 2018) used the $\ell_{2,1} - l_1$ norm. By combining a joint sparsity regularization term with multitask learning, their proposed model produced more accurate results. Jie et al. in (Jie et al., 2017), introduced a group regularization term to the sparse linear regression model. They have also added two smoothness regularization terms to the objective function to ensure that the model keeps the differences between the weight vectors belonging to adjacent time-points to be small. Their proposed model leveraged the prediction performance of the MMSE and ADAS-Cog scores from other existing sparse learning based models.

The neuropathological symptoms of AD in its different stages are complex and combining different modalities in an effective way does augment the prospects for a more accurate diagnosis. Although there are many studies dealing with multimodal datasets, only a few discussed the discrepancy in the different representations of feature domains (Yang et al., 2010; Cheng et al., 2015). On the other hand, missing a screening test on a given visit or dropping out of an entire follow-up visit results in data scarcity in the multimodal database, a drawback experienced in most longitudinal studies. Therefore, to make a reliable prediction of MMSE changes over time, a distributed multimodal multitask framework is proposed in this study to overcome these types of data scarcity problems. In multitask learning, the regularizing term presumes that an equivalent degree of importance exists in the feature space. Therefore, if a positive correlation between the features from different modalities is not found, or if the features are not linearly correlated, the process may fail to identify relevant patterns. In this case, constructing a unified multitask learning model over the concatenated information may not be the optimal approach. To address this problem, a multitask modality-specific regression framework is proposed to predict future MMSE scores for up to 48 months while relying on measurements provided at baseline. Separate multitask regression matrices are trained for each modality to ensure that the coefficient matrices select the leading features extracted from the same modality between consecutive tasks.

The objective function of each regression model uses the correlation and sparsity pattern that exists between all tasks within each modality to improve the longitudinal prediction accuracy. In the second stage of the algorithm, a gradient boosting method is implemented to take a concatenated series of temporal predictions from different modalities and improve the overall performance of the model by predicting a final score. This segregation of modalities in multitask modality-specific regression offers the following advantages:

- Resolves issues related to nonlinear or negative correlations between different feature spaces, which could hinder the performance of multitask learning.
- Provides an error propagation-free framework through a combination of modality-specific multitask learning and gradient boosting. This approach assumes that potential errors might exist in the measurements of a specific modality that originated from capturing, processing or extracting data. Concatenating data from different modalities will thus increase the risk of spreading this error to the fused feature space. Hence, by training separate models and performing a majority vote for the distributed models, the source of error can be detected and consequently prevented from propagating into the fused feature space.
- Overcomes the missing data challenge by projecting a highly dimensional and highly sparse input feature space into multiple low-dimensional and less-sparse spaces. This ensures that the independent coefficient matrices can collectively determine and order the most important biomarkers in the whole dataset.

It is worth noting that the motivation of the model as envisioned is to predict the trajectories of cognitive decline for subjects without any preliminary diagnosis and without regard to the historical records. Thus, the applicability of the proposed framework in terms of providing prediction from baseline information makes it different from methods that need at least a few historical records to be available. For example, Zhu et al in (Zhu et al., 2016) proposed a method for early diagnosis of AD by analyzing longitudinal MRI records and constructing a new feature space from the mean and the difference between the first and last visits measurements. While involving historical records from patients into the training phase may improve the prediction accuracy, it limits the applicability of the model to only those patients with available medical records.

The rest of the paper is organized as follows: Section 2 presents a brief mathematical background of single task regression, multitask regression, and the gradient boosting method. The methodology and implementation steps of the proposed model are described in Section 3. The proposed model is formally introduced with the mathematical formulations that guided this study and with a step-by-step implementation process which are described in subsections 3.1 through 3.4. Section 4 begins with a discussion on the data considered in this study and provides a comprehensive assessment of the experiments conducted. Concluding remarks and a retrospective on the results obtained are provided in Section 5.

2. Background

2.1 Problem Description

The development of Alzheimer's Disease takes place along a trajectory spanning several years with transitions phases that vary from one patient to another. Therefore, in longitudinal AD studies, individuals repeat medical screening tests at multiple follow-up visits and their MMSE scores are recorded and analyzed at each visit. MMSE, with a range of 0 to 30, is the screening test most commonly used for memory and cognitive evaluation. While it is not intended to replace neurological diagnostic labels, it is used to validate the reliability of medical examinations or to evaluate temporal cognitive decline in people suffering from AD. Early intervention plans are effective only if the earliest manifestations of AD are identified at the onset of the disease. Therefore, predicting future trajectories of MMSE scores enables doctors to identify future pathological levels of memory and cognitive impairment. Consequently, the initial objective of this paper is to predict the MMSE scores ($\boldsymbol{b}$) of subjects, by finding the best model g, such that $g: \boldsymbol{b} = \boldsymbol{A}\boldsymbol{w}$, where $\mathbf{w}$ is the regression coefficient and $\boldsymbol{A}$ is the baseline information of the subjects. In support of the proposed approach introduced in Section 3, the required mathematical background is introduced in sub-sections 2.2 through 2.4.

2.2 Single Task Regression

Let $\boldsymbol{A} \in \mathbb{R}^{N\times P}$ be a matrix consisting of N subjects with P features describing each subject, with $\boldsymbol{b}^t \in \mathbb{R}^{N\times 1}$, $t = 1,2,..,T$ defining the clinical scores of those N subjects at the t^{th} time point. The problem of predicting the clinical scores at multiple future time points could be formulated as solving T different regression models as $g^t: \boldsymbol{A} \in \mathbb{R}^{N\times P} \rightarrow \boldsymbol{b}^t \in \mathbb{R}^{N\times 1}, t = 1,2,..T$.

In the simplest form, these T regression problems can be solved using the following *Ridge* regression formula:

$$\widehat{\boldsymbol{w}}^t = \arg\min_{\ddot{\boldsymbol{w}}} \|\boldsymbol{s}\odot(\boldsymbol{b}^t - \mathbf{A}\ddot{\boldsymbol{w}})\|_2^2 + \theta\|\ddot{\boldsymbol{w}}\|_2^2 \quad (1)$$

where $\widehat{\boldsymbol{w}}^t \in \mathbb{R}^{P\times 1}$; $t = 1, 2, \dots, T$ are T independent coefficient vectors calculated by solving the minimization problem in Eq. (1). The $\ddot{w}$ is used as a variable under the arg min function to avoid any confusion with $\boldsymbol{w}$ (the perfect target) and $\widehat{\boldsymbol{w}}$ (the estimated target). In other words, at the last iteration, $\ddot{w}$ that minimizes the arg min function is set as the best estimate $\widehat{\boldsymbol{w}}$ (i.e., $\widehat{\boldsymbol{w}} \leftarrow \ddot{\boldsymbol{w}}$). Symbol $\odot$ defines the component-wise multiplier and vector $\boldsymbol{s} \in \mathbb{R}^{N\times 1}$ defines the missing target values; meaning that $s_n = 0$ if the target value of the n^{th} patient is missing at the t^{th} time point, and $s_n = 1$ if the target value of the n^{th} patient is available at that same time point. In Eq. (1), the $\|\ddot{\boldsymbol{w}}\|_2^2$ is the squared ℓ_2 norm of the coefficient vector $\ddot{\boldsymbol{w}}$, which is controlled by tuning parameter θ. Recall that the p norm of a vector $\boldsymbol{x} \in \mathbb{R}^{K\times 1}$ with $\boldsymbol{x} = [x_1, x_2, \dots, x_K]'$ is defined as:

$$\ell_p = \|\boldsymbol{x}\|_p = \left(\sum_k^K |x_k|^p\right)^{1/p} \quad (2)$$

The penalty term $\theta\|\ddot{\boldsymbol{w}}\|_2^2$, controls the amount of coefficient shrinkage and forces the variance to be close to zero in order to reduce the mean-squared error. Another solution in finding g is to employ the *Lasso* regression formulated as a constrained minimization problem as follows:

$$\widehat{\boldsymbol{w}}^t = \arg\min_{\ddot{\boldsymbol{w}}} \|\boldsymbol{s}\odot(\boldsymbol{b}^t - \mathbf{A}\ddot{\boldsymbol{w}})\|_2^2 + \theta\|\ddot{\boldsymbol{w}}\|_1 \quad (3)$$

In this formula, increasing θ forces the majority of coefficients in $\ddot{\boldsymbol{w}}$, which are associated with features deemed not to be important, to be close to zero and shrink the non-zero coefficients simultaneously. The only difference between these two regression models is in squaring the ℓ_2 norm in *Ridge* regression and using ℓ_1 as the penalty terms in *Lasso* regression, which increases the sparsity of the coefficients.

2.3 Multitask Regression

Another way to tackle the problem of predicting cognitive scores at multiple time points is to employ multitask learning. In the single-task approach, each task is defined as predicting MMSE scores at a single time point and several independent regression models are trained separately to perform prediction for each time point. On the other hand, the multitask approach utilizes the similarities between different tasks to find a more accurate regression model that can carry out multiple prediction tasks. This means that in multi-task learning all the MMSE scores belonging to the T time points will be calculated simultaneously.

Multitask learning can be mathematically formulated as a predictor $G: \boldsymbol{A} \in \mathbb{R}^{N\times P} \rightarrow \boldsymbol{B} \in \mathbb{R}^{N\times T}$ where $\boldsymbol{B} = [\boldsymbol{b}^1, \boldsymbol{b}^2, \dots, \boldsymbol{b}^T]$ is the target values of N subjects at T time points. This multitask predictor G can be modeled using a weight matrix $\boldsymbol{W} = [\boldsymbol{w}^1, \boldsymbol{w}^2, \dots, \boldsymbol{w}^T]$ where $\boldsymbol{W} \in \mathbb{R}^{P\times T}$. In computing the $\boldsymbol{W}$ matrix, one approach is to solve the convex optimization problem as expressed in Eq. (4), also known as the convex fused sparse group Lasso (cFSGL) (Zhou et al., 2012).

$$\widehat{\mathbf{W}} = \arg\min_{\ddot{\mathbf{W}}} \left\|\boldsymbol{S}\odot(\mathbf{B} - \mathbf{A}\ddot{\mathbf{W}})\right\|_F^2 + \theta\left\|\ddot{\mathbf{W}}\right\|_1 + \lambda\left\|\ddot{\mathbf{W}}\right\|_{2,1} + \eta\left\|\boldsymbol{R}\ddot{\mathbf{W}}'\right\|_1 \quad (4)$$

where $\odot$, as defined earlier, is the component-wise multiplier and matrix $\boldsymbol{S} \in \mathbb{R}^{N\times T}$ specifies the missing target values, in which $S_{n,t} = 0$ if the target value of the n^{th} patient is missing at the t^{th} time point, and $S_{n,t} = 1$ if the target value is available. $\widehat{\mathbf{W}}$ is the estimation of the $\mathbf{W}$ achieved by solving the minimization problem. Terms θ, λ, and η are the hyperparameters that control the effect of each regularization term in the cost function and are optimized during the training phase to improve the performance of the algorithm. $\|\boldsymbol{W}\|_1$ is the Lasso penalty term and $\|\boldsymbol{W}\|_F^2$ is the squared Frobenius norm and the $\left\|\ddot{\mathbf{W}}\right\|_{2,1}$ is known as the Group Lasso penalty. Moreover, $\|\boldsymbol{R}\boldsymbol{W}'\|_1$ is the Fused Group Lasso penalty, and $\boldsymbol{R}$ is $(T-1)\times T$ sparse matrix interpreted as a descriptor of the relatedness between different tasks. Assuming each task as a node in a graph, a relationship between every two tasks is represented by a connection between their corresponding nodes. This penalty term controls the transition between neighboring tasks and forces the transition within successive tasks to remain small (a process also known as temporal smoothness). In other words, $R_{i,j} = 0$ indicates that the task assigned to node i is not related to the task assigned to node j, while $R_{i,j} = \alpha$ indicates that task i and task j are associated with each other with a degree of α. In the proposed model, this parameter restrains the variation of predicted cognitive scores in neighboring time steps, meaning that trajectories of MMSE scores at two consecutive time points cannot have spikes. In order to solve Eq.

(4), the accelerated gradient method (AGM) was used, which is available in the MALSAR package (Zhou et al., 2012).

Another approach for finding the weight matrix $\boldsymbol{W}$ is to use the non-Convex Fused Sparse Group Lasso (nFSGL1) as formulated in (Zhou et al., 2012):

$$\widehat{\mathbf{W}} = \arg\min_{\ddot{\mathbf{W}}} \left\| \boldsymbol{S} \odot (\mathbf{B} - \mathbf{A}\ddot{\mathbf{W}}) \right\|_F^2 + \eta \left\| \boldsymbol{R}\ddot{\mathbf{W}}' \right\|_1 + \theta \sum_{i=1}^{P} \sqrt{\|\ddot{\mathbf{w}}_i\|_1} \tag{5}$$

where $\ddot{\boldsymbol{w}}_i$ is the i^{th} row of $\ddot{\mathbf{W}}$. The convex and non-convex Fused Group Lasso formulas allow for joint feature selection across all tasks while selecting distinct feature sets for each task.

The joint selection of the coefficients in **W** could also be penalized in the form of $\ell_{2,1}$-norm with least square loss. Thus, the finding of the optimal **W** can be formulated as:

$$\widehat{\mathbf{W}} = \arg\min_{\ddot{\mathbf{W}}} \frac{1}{2} \left\| \boldsymbol{S} \odot (\mathbf{B} - \mathbf{A}\ddot{\mathbf{W}}) \right\|_F^2 + \lambda_1 \left\| \ddot{\mathbf{W}} \right\|_{2,1} + \lambda_2 \left\| \ddot{\mathbf{W}} \right\|_F^2 \tag{6}$$

To incorporate global and local information in the feature set with a sparse regression method, Zhu et al in (Zhu et al., 2016) reformulated the objective function in equation (6) as follows:

$$\widehat{\mathbf{W}} = \arg\min_{\ddot{\mathbf{W}}} \frac{1}{2} \left\| \boldsymbol{S} \odot (\mathbf{B} - \mathbf{A}\ddot{\mathbf{W}}) \right\|_F^2 + \lambda_1 tr(\ddot{\mathbf{W}}' \boldsymbol{A}' \boldsymbol{L} \mathbf{A}\ddot{\mathbf{W}}) + \lambda_2 \left\| \ddot{\mathbf{W}} \right\|_{2,1} \tag{7}$$

where λ_1 and λ_2 are the regularization parameters and *tr* (.) denotes the trace operator. Here, with **R** being the adjacency matrix, the Laplacian matrix L can be defined as:

$$\mathbf{L} = \mathbf{D} - \mathbf{R} \tag{8}$$

where **D** is the symmetric diagonal matrix in which the diagonal elements D_{ii} = 1 and all the other non-diagonal entries are 0. Zhu et al. in (Zhu et al., 2018.) proposed an iterative method for finding the solution of multitask problem, i.e. $\boldsymbol{W}$, to reduce the number of hyperparameters that must be learned in the multitask learning problem. The objective function in this proposed approach is to find the $\boldsymbol{w}^t$values through the following formulation:

$$\widehat{\mathbf{W}} = \arg\min_{\ddot{\boldsymbol{w}}^t,\ \bar{w}} \sum_t^T \alpha^t \left(\left\| \boldsymbol{s} \odot (\boldsymbol{b}^t - \mathbf{A}\ddot{\boldsymbol{w}}^t) \right\|_2^2 + \left\| \ddot{\boldsymbol{w}}^t - \bar{\boldsymbol{w}} \right\|_{2,1} \right) + \lambda_2 \left\| \ddot{\mathbf{W}} \right\|_1 \tag{9}$$

where $\bar{\boldsymbol{w}}$ is the mean vector of $\ddot{\boldsymbol{w}}^t (t = 1, 2, \ldots T) \in \boldsymbol{W}$. For each task *t*, the weights of each task denoted as α^t are calculated automatically with the following equation:

$$\alpha^t = \frac{1}{2\sqrt{\left\| \boldsymbol{s} \odot (\boldsymbol{b}^t - \boldsymbol{A}\ddot{\boldsymbol{w}}^t) \right\|_2^2 + \left\| \ddot{\boldsymbol{w}}^t - \bar{\boldsymbol{w}} \right\|_{2,1}}} \tag{10}$$

Employing the centralized regularization in the objective function of (9) balances the variances of the coefficients in $\boldsymbol{w}^t$ by penalizing them separately using α^t.

2.4 Gradient Boosting

Ensemble models have been shown to be effective in various prediction tasks by grouping a set of weak learners to construct a more powerful learner. Bagging and boosting are the two mainstream techniques in ensemble learning methods. The former creates independent and uncorrelated learners on subsets of data and generates the final result by voting or averaging the outcomes of independent learners. On the contrary, the latter generates a collection of weak learners, in which the predictors are trained sequentially rather than separately. In boosting methods, the goal is to utilize the error of the previous learners to develop a more efficient model for the next learner. With training the learners sequentially, subsets of data do not have the chance to concurrently affect all the learners. The algorithm invests a larger weight on the samples that were classified inaccurately, forcing the hypothesis of the next weak learners to precisely analyze those tough samples and eventually improve the performance of the model.

An extension of the boosting methods is gradient boosting, which is a supervised machine learning technique based on regression, classification, and ranking. It uses the gradient descent optimization technique to find the global or local minima of the cost function. Using a sequence of weak learners, Gradient Boosting (GB) trains a machine to fit a model on the input feature space such that each learner improves the prediction accuracy of the previous ones. Through multiple iterations, gradient boosting develops a single strong learner by combining multiple weak learners (Friedman, 2001; Ogutu et al., 2011). In the proposed method, GB constructs the final stage of the framework to

improve the prediction accuracy by successively fitting a more accurate model on the residuals of the previous step. This procedure will continue until it achieves a highly accurate model. Sub-sections 3.3 and 3.4 provides more details on the role of GB in the context of the proposed framework.

3. Method

3.1 Notations and parameters

Through the rest of the paper, matrices are denoted as bold uppercase letters and vectors are denoted as italic bold letters. Matrices $\boldsymbol{X}_m^t \subseteq \mathbf{X}$ and $\Omega_m^t \subseteq \Omega$ are the feature space and patients' roster ID associated with the subjects who have been examined at time point t with modality test m. For these subjects, $\boldsymbol{y}^t$ with $t = 1, 2, \dots, \mathrm{T}$ are their respective cognitive scores (independent from the source of the modality). Similarly, $\boldsymbol{\mathcal{F}}$ is the risk factor matrix consisting of age, gender, years of education and APOE4 factors for all patients. It is noted that the (') notation denotes transposition and should not be confused with t = 1, 2, …, T which define the different time points in the longitudinal study, where T denotes the 48$^{\text{th}}$ month.

3.2 Method Overview

Tracking future MMSE scores reveals a subtle but progressive decline in cognitive levels of individuals through the different stages of AD and informs on the nature of the transition phases of the disease. However, prognostication of AD progression, regardless of the label associated with the subject at baseline, remains challenging, especially in a multimodal platform. Certain modalities have shown a relatively higher impact on the asymptomatic or symptomatic phases of AD. This promoted the use of multimodal biomarkers to improve the accuracy of identifying neurobiological and clinical symptoms of the disease. However, the interactions and correlations between the biomarkers from complementary modalities remain intricate. Furthermore, longitudinal datasets continue to suffer from the missing data challenge.

Considering the data scarcity and the discrepancy in the correlation matrix associated with the heterogeneous multimodal longitudinal dataset, we propose to utilize the modality-specific multitask coefficient matrix. These unique multitask coefficient matrices are trained over different sets of biomarkers extracted from each modality to model the temporal interaction between the baseline features and the transitions of the cognitive scores at successive time points.

The strength and capability of different modalities in tracking the progression of AD are still inconclusive. Therefore, granting equal contribution (or equal weight) to the predictive biomarkers from different modalities increases the chance of achieving better prediction accuracy. This is accomplished by capturing the complex yet effective correlation between important modality-exclusive features and eliminating the effect of all other extraneous ones. Next, the initial outcomes of these cooperative multitask learners are fused with risk factors, which are assumed as time-invariant information. Finally, a gradient boosting kernel is trained over this new collective data representation to leverage the prediction accuracy through ensemble learning and looking into sparse and interpretable solutions. In the next section, we will go through the setup of our multimodal-multitask model.

3.3 Method formulation

Suppose that $\boldsymbol{X} \in \mathbb{R}^{N \times P}$ is the multimodal feature space and $\boldsymbol{Y} = [\boldsymbol{y}^1, \boldsymbol{y}^2, \dots, \boldsymbol{y}^T]$ is representing the cognitive trajectories of these N subjects through T time steps. For each interval t, $\boldsymbol{X}^t \subseteq \boldsymbol{X}$ is the set of subjects who are chosen based on Ω^t, the roster ID of population $\boldsymbol{y}^{\mathrm{t}}$. It is worth noting that some subjects may have not returned for the follow-up visit at t^{th} time point and therefore $\Omega^t < \Omega$ is possible. Considering M as the total number of modality sources, $\boldsymbol{X}^t$ and $\boldsymbol{y}^{\mathrm{t}}$ are decomposed into M subgroups, thus constructing $T \times M$ pairs of $\{(\boldsymbol{X}_m^t, \boldsymbol{y}_m^t), m = 1, 2, \dots, M,\ t = 1, 2, \dots,\ T\}$, where each pair of $(\boldsymbol{X}_m^t, \boldsymbol{y}_m^t)$ are the m^{th} single-modality measurements associated with the t^{th} time point.

The single task regression method will be extended to the $T \times M$ optimization problems to calculate $\boldsymbol{w}_m^t$ by solving equations (11) and (12).

$$\hat{\boldsymbol{w}}_m^t = \arg\min_{\ddot{\boldsymbol{w}}} \|(\boldsymbol{y}_m^t - \mathbf{X}_m^t \ddot{\boldsymbol{w}})\|_2^2 + \theta \|\ddot{\boldsymbol{w}}\|_2^2 \tag{11}$$

$$\hat{\boldsymbol{w}}_m^t = \arg\min_{\ddot{\boldsymbol{w}}} \|(\boldsymbol{y}_m^t - \mathbf{X}_m^t \ddot{\boldsymbol{w}})\|_2^2 + \theta \|\ddot{\boldsymbol{w}}\|_1 \tag{12}$$

where $\widehat{\boldsymbol{w}}_m^t \in R^{P_m \times 1}$ is the $\widehat{\boldsymbol{w}}_m$estimate at the t^{th} time point.

In the multitask learning approach, the objective function will be extended to G_m: $\boldsymbol{X}_m^t \rightarrow \overline{\overline{\boldsymbol{Y}}}_m$where $\overline{\overline{\boldsymbol{Y}}}_m \in \mathbb{R}^{N \times T}$ is the concatenated matrix $\overline{\overline{\boldsymbol{Y}}}_m = [\bar{\boldsymbol{y}}_m^1, \bar{\boldsymbol{y}}_m^2, \dots, \bar{\boldsymbol{y}}_m^T]$ with $\bar{\boldsymbol{y}}_m^t$ being the extended versions of their corresponding $\boldsymbol{y}_m^t$, in which the unavailable test scores of the patients are represented by zero values. The size discrepancy in $\bar{\boldsymbol{y}}_m^t$, which is a consequence of missing modalities and dropout is illustrated in Fig. 1.

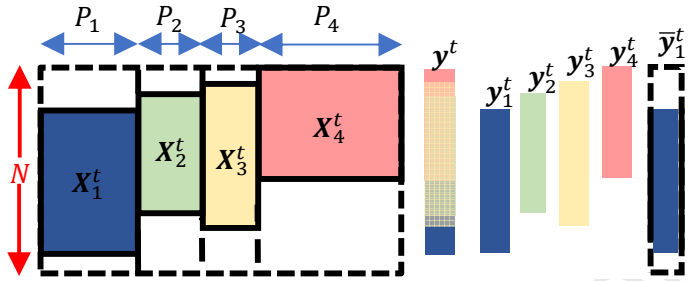


Fig. 1. Illustrative example of size discrepancy in a longitudinal multimodal dataset. Available measurements extracted from each modality are shown with colored boxes and the missing information are displayed in the blank

In this figure, patterns of missing values and arrangements of available information from four modalities are represented over a fixed time period. Using a modality-specific approach, the objective function of multitask learners will be reformulated to calculate M number of $\boldsymbol{W}_m \in R^{P_m \times T}$ where $\boldsymbol{W}_m = [\boldsymbol{w}_m^1, \boldsymbol{w}_m^2, \dots, \boldsymbol{w}_m^T]$. Thus, the cFSGL (convex Fused Sparse Group Lasso) problem can be formulated as follows:

$$\widehat{\boldsymbol{W}}_m = \arg\min_{\ddot{W}} \left\| \boldsymbol{S} \odot (\overline{\overline{\boldsymbol{Y}}}_m - \boldsymbol{X}_m^1 \ddot{\boldsymbol{W}}) \right\|_F^2 + \theta \left\| \ddot{\boldsymbol{W}} \right\|_1 + \lambda \left\| \ddot{\boldsymbol{W}} \right\|_{2,1} + \eta \left\| \boldsymbol{R} \ddot{\boldsymbol{W}}' \right\|_1 \tag{13}$$

And based on nFSGL1 (non-Convex Fused Sparse Group Lasso), the objective function will be formulated as follows:

$$\widehat{\boldsymbol{W}}_m = \arg\min_{\ddot{W}} \left\| \boldsymbol{S} \odot (\overline{\overline{\boldsymbol{Y}}}_m - \boldsymbol{X}_m^1 \ddot{\boldsymbol{W}}) \right\|_F^2 + \eta \left\| \boldsymbol{R} \ddot{\boldsymbol{W}}' \right\|_1 + \theta \sum_{i=1}^{P_m} \sqrt{\|\ddot{\boldsymbol{w}}_i\|_1} \tag{14}$$

Using a similar approach, equations (6), (7), (9) and (10) will be reformulated respectively as follows:

$$\widehat{\mathbf{W}}_m = \arg\min_{\ddot{\mathbf{W}}} \frac{1}{2} \left\| \boldsymbol{S} \odot (\overline{\overline{\boldsymbol{Y}}}_m - \boldsymbol{X}_m^1 \ddot{\boldsymbol{W}}) \right\|_F^2 + \lambda_1 \left\| \ddot{\boldsymbol{W}} \right\|_{2,1} + \lambda_2 \left\| \ddot{\boldsymbol{W}} \right\|_F^2 \tag{15}$$

$$\widehat{\mathbf{W}}_m = \arg\min_{\ddot{\mathbf{W}}} \frac{1}{2} \left\| \boldsymbol{S} \odot (\overline{\overline{\boldsymbol{Y}}}_m - \boldsymbol{X}_m^1 \ddot{\boldsymbol{W}}) \right\|_F^2 + \lambda_1 tr(\ddot{\boldsymbol{W}} \acute{\boldsymbol{X}}_m^1 \boldsymbol{L} \boldsymbol{X}_m^1 \ddot{\boldsymbol{W}}) + \lambda_2 \left\| \ddot{\boldsymbol{W}} \right\|_{2,1} \tag{16}$$

$$\widehat{\mathbf{W}}_m = \arg\min_{\ddot{\boldsymbol{w}}^t,\ \bar{w}} \sum_t^T \alpha^t \left(\left\| \boldsymbol{s} \odot (\bar{\boldsymbol{y}}_m^t - \boldsymbol{X}_m^1 \ddot{\boldsymbol{w}}^t) \right\|_2^2 + \left\| \ddot{\boldsymbol{w}}^t - \bar{w} \right\|_{2,1} \right) + \lambda_2 \left\| \ddot{\boldsymbol{W}} \right\|_1 \tag{17}$$

$$\alpha^t = \frac{1}{2\sqrt{\left\| \boldsymbol{s} \odot (\bar{\boldsymbol{y}}_m^t - \boldsymbol{X}_m^1 \ddot{\boldsymbol{w}}^t) \right\|_2^2 + \left\| \ddot{\boldsymbol{w}}^t - \bar{w} \right\|_{2,1}}} \tag{18}$$

The flowchart of the proposed method in the training stage is illustrated in Fig. 2. In this figure, step 1 represents the training process for the modality-specific regression coefficient matrices $\widehat{\boldsymbol{W}}_m$. The input space is constructed by T stack of modality-specific feature spaces, $\mathbf{X}_m^t$, t = 1, 2, …, T and the targets are their respective cognitive scores characterized as $\hat{\boldsymbol{y}}_m^t$. At the end of the training stage, step 1 generates M modality-specific multitask learning regression coefficient matrices, $\widehat{\boldsymbol{W}}_m \in R^{P_m \times T}$ for $m = 1, 2, \dots, M$, which are comprised of $\widehat{\boldsymbol{w}}_m^t$ for $t = 1, \dots, T$ in the form of $\widehat{\boldsymbol{W}}_m = [\widehat{\boldsymbol{w}}_m^1, \widehat{\boldsymbol{w}}_m^2, \dots, \widehat{\boldsymbol{w}}_m^T]$. Consequently, using $\mathbf{X}_m^t$ as input measurements, the initial prognostications at time point t are established as:

$$\hat{\boldsymbol{y}}_m^t = \mathbf{X}_m^t \times \widehat{\boldsymbol{w}}_m^t \tag{19}$$

for $m = 1, 2, \dots, M$ and $t = 1, 2, \dots, T$.

Modality-wise multitask coefficient matrices capture the mutual relationships between the feature spaces and cognitive score trajectories. This provides a powerful tool in obtaining the inter-modality correlations and examining the predictive power of each modality exclusively. To take advantage of the information provided from each source

of modality, the outcomes of the multitask models along with risk factor parameters are combined together to form the input space for the gradient boosting. It is worth noting that the risk factor parameters, do not carry the unpredictable temporal pattern as in the other biomarkers. In order to reduce unnecessary computational costs, risk factor parameters have not been processed with multitask learning models and have been added to the second stage of the model. Step 2 in Fig. 2 shows the preparation of the data for the second stage of the method.

For the dataset used here, it is observed that if the PET measurements are available for a group of subjects, the MRI measurements are also available for that group, but the opposite is not necessarily true. Therefore, five configurations of possible modality combinations are considered in this study: (1) MRI-PET, (2) MRI-PET-CSF, (3) MRI-PET-COG, (4) PET-COG-CSF and (5) MRI-PET-COG-CSF.

The Ω_m^t are the sets of roster IDs from subjects that have participated in test m at the t^{th} time point and ${}_c\Omega^t$ is the intersection between all Ω_m^t with respect to their availability in the c^{th} modality combination. Considering c as an indicator of the modality combination, the GB machines are developed as ${}_cGB^t : {}_c\boldsymbol{Z}^t \longrightarrow \boldsymbol{y}^t$ for $c = 1, \dots, 5$ and $t = 1, \dots, T$ over the set of ${}_c\Omega^t$. In which ${}_c\boldsymbol{Z}^t$ is the new feature space for the c^{th} GB machine and is constructed by concatenating ${}_c\widehat{\boldsymbol{y}}_m^t$ and ${}_c\boldsymbol{\mathcal{F}}^t$, which are the initial predictions and risk factors for the population of ${}_c\Omega^t$. This process has been demonstrated in step 3 of Fig. 2.

For example, if the available modalities are MRI and PET, then c = 1. Meaning that in stage 1, only the modality-specific regression coefficient matrices of $\widehat{\boldsymbol{W}}_1$ and $\widehat{\boldsymbol{W}}_2$ can provide the initial predictions as $\widehat{\boldsymbol{y}}_1^t$ and $\widehat{\boldsymbol{y}}_2^t$. Based on their respective roster IDs, ${}_c\Omega^t$, the input space ${}_1\boldsymbol{Z}^t = [\,{}_1\boldsymbol{\mathcal{F}}^t,\ {}_1\widehat{\boldsymbol{y}}_1^t,\ {}_1\widehat{\boldsymbol{y}}_2^t]$ is constructed in step 2. Then the ${}_1\boldsymbol{Z}^t$ and their corresponding sets of cognitive scores, $\boldsymbol{y}^t$, will be used to train the corresponding ${}_1GB^t$ at step 3.

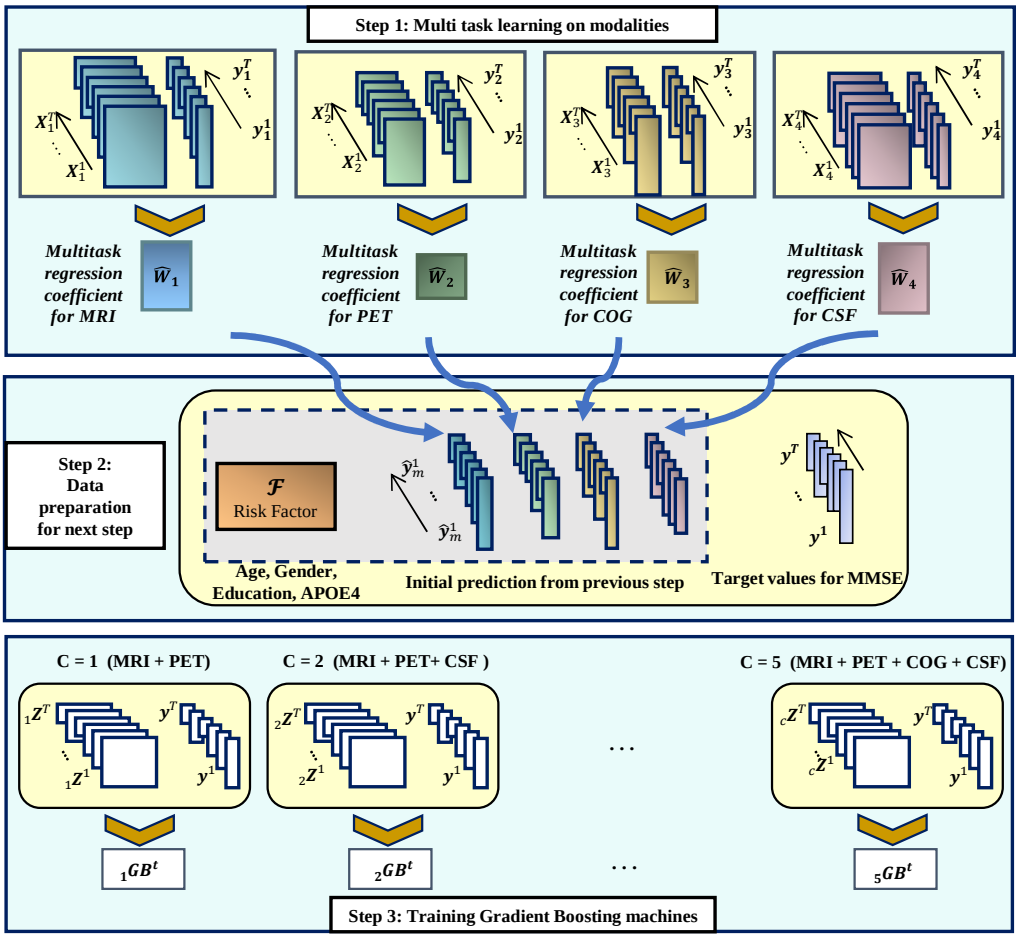
3.4 Test scenario

Suppose that we want to predict the MMSE score at time point t and the patient has completed three modality tests. The available measurements from this patient are thus ($\boldsymbol{x}_1 \in R^{1\times P_1}$) extracted from MRI, ($\boldsymbol{x}_2 \in R^{1\times P_2}$) extracted from PET, ($\boldsymbol{x}_4 \in R^{1\times P_4}$) extracted from CSF test and a vector $\boldsymbol{r}$ containing the risk factor parameters for this patient. In this scenario, the COG modality which is $\boldsymbol{x}_3$ is not available.

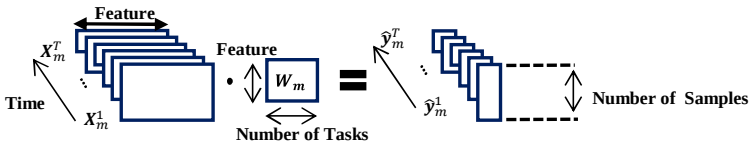
In the first step of the proposed model, modality-wise coefficient matrices will provide the most accurate predictions possible from the measurements of one modality through multitask learning. By feeding $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4$ to their respective modality-wise coefficient matrices, the initial predictions can be calculated as $\hat{y}_1^t = \boldsymbol{x}_1 \times \widehat{\boldsymbol{w}}_1^t$, $\hat{y}_2^t = \boldsymbol{x}_2 \times \widehat{\boldsymbol{w}}_2^t$ and $\hat{y}_4^t = \boldsymbol{x}_4 \times \widehat{\boldsymbol{w}}_4^t$. Next, the initial predictions of $\hat{y}_1^t, \hat{y}_2^t, \hat{y}_4^t$ and risk factors ($\mathcal{F}$) will be concatenated to form the new feature vector ${}_c\boldsymbol{Z}^t = [\boldsymbol{r}, \hat{y}_1^t, \hat{y}_2^t, \hat{y}_4^t]$ where $c = 2$ indicates the mode for modality combination (i.e., MRI-PET-CSF). Then in the second step, gradient boosting employs a boosting approach to ensemble the outcomes from different modalities, determine the correlation among them and reduce their prediction error. The final estimation will be achieved by using the ${}_cGB^t$ machine as $\hat{y}^t = {}_2GB^t({}_2\boldsymbol{Z}^t)$. While incomplete samples with missing intervals are taken care of, through the first step of the algorithm, the second step of the proposed method deals with the missing modalities and the complex relationship between them. The gradient boosting incorporates the predictive power of salient biomarkers from each modality, models the intra-correlation between them, and adjusts the prediction error to improve the final accuracy.

3.5 Data and code availability statement

The clinical data used in conducting this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, and all the details pertaining to the different image processing pipelines can be found in (adni.loni.usc.edu). The code generated for this study can be made available upon request to the corresponding author of this manuscript.



(a)



(b)

Fig. 2. (a) Flowchart of the proposed approach in the training phase, (b) Defining the dimensions in multitask formulation for step 1

Table 1
Demographic characteristic of the studied subjects. valued are specified as mean±standard deviation

Category	Subjects (f/m)	Age	Education(year)	APOE (0/1/2)	MMSE
CN	206/209	74.77±5.74	16.27±2.73	300/103/11	29.07±1.12
MCI	354/510	73.03±7.60	15.91±2.85	427/340/94	27.59±1.81
AD	150/186	74.92±7.81	15.17±2.99	113/156/65	23.18±2.05
MCI to AD	2/3	78.50±2.59	16.40±2.61	1/4/0	26.00±1.58

4. Results and Discussion

4.1 Data

The clinical data used in the preparation of this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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 magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org. ADNI established the following Mini-Mental Exam (MMSE) and Clinical Dementia Rating (CDR) cut off scores to interpret the AD spectrum:

- MMSE of 30 and CDR of 0 is described as cognitively no dementia,
- MMSE of 29-26 and CDR of 0.5 is associated with questionable dementia,
- MMSE of 25-21 and CDR of 1.0 is associated with mild dementia,
- MMSE of 20-11 and CDR of 2.0 is associated with moderate dementia,
- MMSE of 10-0 and CDR of 3.0 is determined as severe dementia.

The experiments in this study used multimodal longitudinal data from 1620 subjects who were enrolled for up to 6 visits in a 4-year time span. This population consists of a total of 1620 subjects with 864 participants with mild cognitive impairment (MCI), 415 cognitively normal subjects (CN), 336 individuals with dementia (AD) and 5 participants whose status changed from mild cognitive impairment to dementia at baseline (MCI to AD conversion). All samples used in this analysis are in the range of 54.4 to 90.3 years old, with 44% female and 56 % male. The majority of the 93.24 % of the population were identified as white, 3.95% as black and the rest were recognized either as Asian, Indian/Alaskan or belonging to more than one ethnicity. 76% reported their marital status as married, 12.61% as widowed, and the rest of the participants were represented as either never married or their status of marriage was recorded as unknown. Table 1 summarizes the demographic characteristics of the ADNI cohort used in this study based on the category of the disease. For the APOE column, the (0, 1, 2) values refer to the number of ε4 alleles in the APOE genotype.

4 .2 Importance of Data Modality and Structure of the Experimental Set-Up

In preparing the data, subjects were partitioned into four categories: individuals who had completed the MRI scanning, individuals with PET scans, individuals with CSF analysis, and individuals with cognitive screening tests. The features extracted from each screening test, and the number of subjects in different time periods, are summarized in Table 2. In relation to time t, t=1 means time point at baseline or T1, t=2 refers to time point at the 6th month or T6, t=3 refers to time point at the 12^{th} month or T12, t=4 refers to the time point at 24^{th} month or T24, t=5 for the time point at 36^{th} month or T36 and finally for t=T, for the last time point at the 48^{th} month or T48. The importance of each data modality in the proposed multitask multimodal approach is reflected in the features that were selected for each modality as shown in Table 2. Observe the decreasing number of observations made at subsequent time points in this ADNI longitudinal study, which highlights the missing data challenge. For this study, through the MRI imaging modality, the main features considered as the most important MRI biomarkers are extracted from seven brain regions to include Ventricular volume, Hippocampus volume, Whole Brain volume, Entorhinal Cortical thickness, Fusiform, Middle temporal gyrus and intracranial volume (ICV). Fig. 3 illustrates these brain regions in the brain template. The PET features are single measurements of the Pittsburgh compound B (PIB), the Florbetapir (AV-45), and the fluorodeoxyglucose (FDG) for cerebral glucose metabolism, all used as agents to image and gauge the extent of amyloid plaques at the different stages of the disease. As we are constrained

to the multimodal features presented in Table 2 for this longitudinal study, future studies could involve the use of PET regional standardized uptake value ratio (SUVRs) as quantitative measures of the radiotracer uptake in regions of interest with respect to a reference region to assess how such measures, especially in disease-prone areas, relate to the MMSE score as used for prediction purposes in this study. In the features listed in Table 2, in accordance with the ADNI multisite study, FDG is the average FDG-PET of angular, temporal, and posterior cingulate, PIB is the average PIB SUVR of frontal cortex, anterior cingulate, precuneus cortex, and parietal cortex and AV45 is the average AV45 SUVR of frontal, anterior cingulate, precuneus, and parietal cortex relative to the cerebellum.

In terms of the cerebrospinal fluid (CSF) biomarkers (Anoop et al., 2010; Hanger et al., 2009; Noble et al., 2013), this study considers Amyloid Beta (ABETA), phosphorylated tau protein (PTAU), and Total tau protein (TAU) as means to assess the extent of amyloid plaques in between neurons and the neurofibrillary tangles made up of tau protein within the neurons themselves, both considered to contribute to the degradation of neurons in Alzheimer's disease and other tauopathies. The other risk factors considered in this study include age, gender, level of education and Apolipoprotein E (APOE) gene. As indicated earlier, APOE with the E4 allele apolipoprotein is considered a major genetic risk factor for AD (Bussy et al., 2019). As for age and gender, it is common knowledge that age is a major risk factor in AD (since only about 5% develop symptoms of AD before the age of 65) and it is estimated that two-thirds of the 5.5 million Americans living with AD are women. Although women tend to live longer than men, we still could not conclude with certainty that this discrepancy in the larger number of women with AD is only due to longevity and experts remain uncertain on other factors that could explain this difference. As for the level of education, there is an understanding and some studies confirm that the higher is the level of education the lower is the risk for dementia, and that cognitive reserve serves as a strength to overcome some the symptoms of AD (Stern, 2012; Buckner, 2004).

Table 2- Summary of ADNI dataset, the number of observations in each follow-up visit and the features extracted from each modality

Source*	**Number of observations**						**Features**
	T1	T6	T12	T24	T36	T48	
MRI	1465	1333	1191	987	617	451	Ventricular volume, Hippocampus volume, Whole Brain volume, Entorhinal Cortical thickness, Fusiform, Middle temporal gyrus, and intracranial volume (ICV)
PET	1127	1009	892	714	429	335	FDG, Pittsburgh Compound-B (PIB), AV45
Cognitive Test**	1525	1357	1207	997	627	456	Rey Auditory Verbal Learning Test (RAVLT Immediate, RAVLT Learning, RAVLT Forgetting, RAVLT Perc Forgetting), Functional Activities Questionnaires (FAQ), Everyday Cognition (Ecog) scales: (EcogPtMem, EcogPtLang, EcogPtVisspat, EcogPtPlan, EcogPtOrgan, EcogPtDivatt, EcogPtTotal,EcogSPMem, EcogSPLang, EcogSPVisspat, EcogSPPlan, EcogSPOrgan, EcogSPDivatt, and EcogSPTotal)
CSF	1014	914	806	662	404	305	Amyloid Beta (ABETA), Phosphorylated Tau Protein (PTAU), and Total Tau Protein (TAU)
Risk factors	1737						Age, gender, years of education, and APOE4

* In this table MRI refers to Magnetic Resonance Imaging, PET refers to Positron Emission Tomography, COG refers to Cognitive assessment tests and CSF refers to Cerebrospinal Fluid test.

**The Mini-Mental State Examination (MMSE) and Clinical Dementia Rating Sum of Boxes (CDRSB) scores (*since initially used for labelling subjects*) and Alzheimer's Disease Assessment Score (ADAS11, ADAS13) and the Montreal Cognitive Assessment (MoCA) (*since highly correlated with MMSE*) were excluded from the feature set in the training and testing phases of the proposed prediction model.

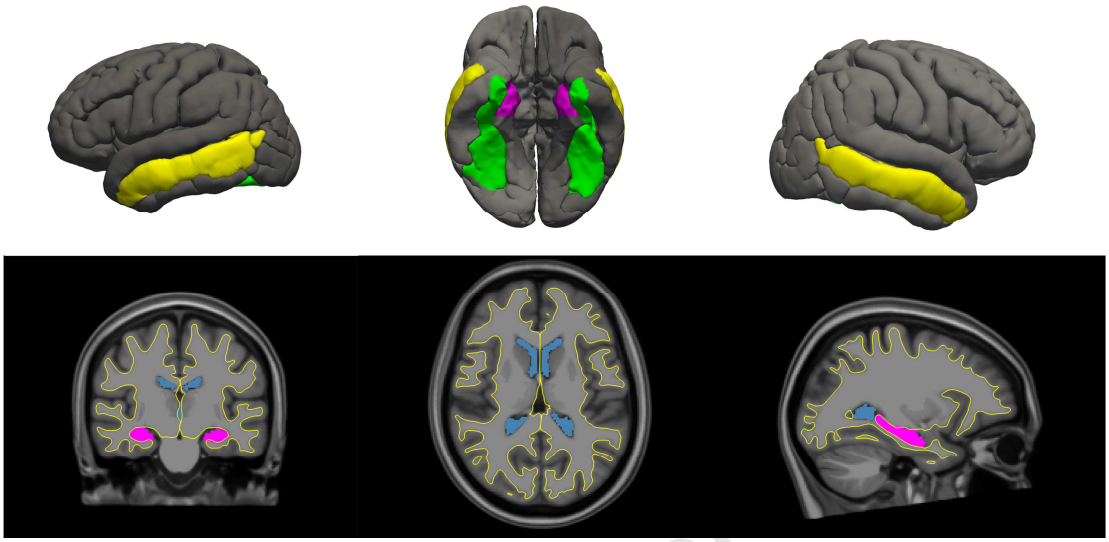


Fig. 3. Selected MRI brain regions for tracking the progression of Alzheimer's disease. 3D mesh surface map, with purple, green, and yellow areas representing Entorhinal, fusiform, and middle temporal regions, respectively (Top). The volumetric segmentation, in which the yellow line depicts the interface between grey and white-matter, and the purple and blue regions representing the hippocampus and ventricles, respectively (Bottom).

In the preprocessing step, ADAS11, ADAS13, MoCA, the Diagnosis labels (DX) and CDR were removed from the feature set since it is known that they have a high correlation with the MMSE score. We further excluded non-stable CN participants (CN to MCI or CN to AD) and subjects who are facing a reverse-phase in the progression stage (MCI to CN, AD to MCI).

Given the number of subjects considered for this study (1620), to compensate for the small sample size, nested cross-validation has been applied to our data set. From the whole dataset, 70% were randomly selected as the training set and 30% were set aside as the testing set. This process of randomly splitting the data has been repeated 10 times to avoid any bias in the evaluation of data. For hyperparameter selection, in each of those data splits, 5-fold inner cross-validation along with exhaustive search is used to select the optimal hyperparameters for each method. For regression methods, the regularization parameters were selected in a range of $\{10^{-3}$ to $10^{3}\}$. As for the XGBoost method, the number of estimators is searched between {1 and 500}, learning rate has been searched between $\{10^{-3}$ and 1}, the number of columns used by each tree (colsample_bytree) has been searched between {0.1 to 1} and max depth has been searched between {1 and 15}.

Through the rest of the paper, reported values are the mean and standard deviation of the experiments in these10 different random train and test split. It is important to mention that, feature space from every observation in both the training set and the testing set were normalized separately using the Z- score (i.e., dividing the difference between each value and the mean by the standard deviation).

4.3 Selecting modality-specific multitask models

The first stage of the model is focused on developing modality-specific multitask coefficient matrices. The motivation is to not confuse the multitask regression coefficients with modeling the relationship between different modalities and to preserve the maximum learning capacity to be devoted to learning the trajectories of cognitive decline. The following state-of-the-art algorithms are selected as the competing methods in the investigation of predicting clinical decline at multiple time points.

- Ridge regression
- Elastic Lasso
- Temporal Group Lasso (TGL)
- Convex Fused Sparse Group Lasso (cFSGL)
- Non-convex Fused Sparse Group Lasso (nFSGL)
- Subspace Regularized Sparse multitask learning (Zhu et al., 2016)

- Parameter-free least Lasso (Zhu et al., 2018)

For single task learners, six separate regression models have been trained to predict cognitive scores for each time point. However, in multitask learning, the regression coefficients for all time points are trained together. This approach improves the efficiency of the final model by identifying and capturing the correlation between the transitions of cognitive scores at successive time points. To benchmark the performance of different methods, Root Mean Square Error (RMSE) and R correlation coefficient (denoted as *Corr* in Tables and figures that follow) are selected as the main evaluation metrics through this study. Fig. 4 demonstrates the comparison of prediction accuracy of regression models using different sets of biomarkers. Several important empirical observations can be made from analyzing the results given in Fig. 4.

First, single-task models yield a competitive performance at earlier time points but multitask learners significantly surpassed them at subsequent time points. This analysis found clear evidence for the superiority of multitask learners over single task learners.

Second, the sparsity and temporal sample size of each modality-specific feature space differ from each other. For each modality, the regression model which yields the highest winning rate is selected as the best predictor. The winning rate is defined here as the number of times a specific method achieves the best performance in term of lowest error across all intervals and highest correlation in comparison to all the other methods. It is important to emphasize that the winning models are selected during the training phase without seeing the test data. It can be observed that cFSGL proved to be the best method for PET and CSF, just as the method in (Zhu et al., 2018) yielded the best overall performance results for COG measurements, and the coefficient matrix in (Zhu et al., 2016) achieved the best prediction accuracy for MRI measurements. The ℓ_2 norm regularization penalty term in cFSGL results in non-zero values in $\boldsymbol{W}$. Since the feature spaces for PET and CSF are low dimensional and less sparse, using ℓ_2 norm will help determine and keep the best predictive biomarkers. The COG modality was found to have a higher dimensionality and the pattern of features is highly sparse, which enabled the coefficient matrix in (Zhu et al., 2018) to achieve better generalization than other methods.

Third, the cognitive modality achieved the smallest error in comparison to all other modalities in predicting the cognitive decline. However, it must be pointed out that ADAS11, ADAS13, MoCA, CDR, and diagnosis labels were removed from the cognitive feature space to ensure that variables with a strong correlation with the MMSE label are not biasing the prediction. The scatter plot for cognitive assessment modality is shown in Fig. 5.

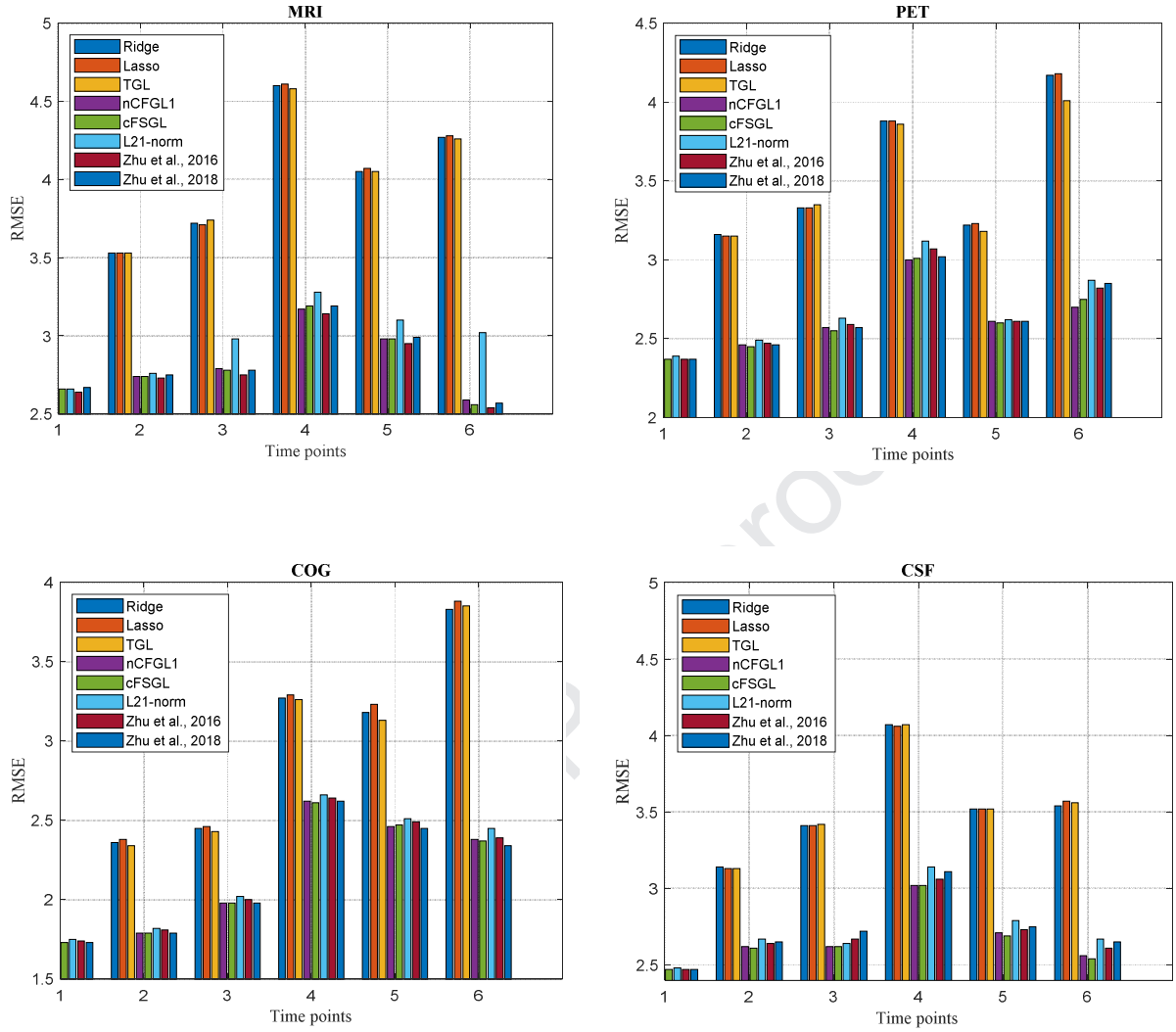


Fig. 4. Performance comparison of different regression methods on longitudinal prediction of MMSE using different modalities.

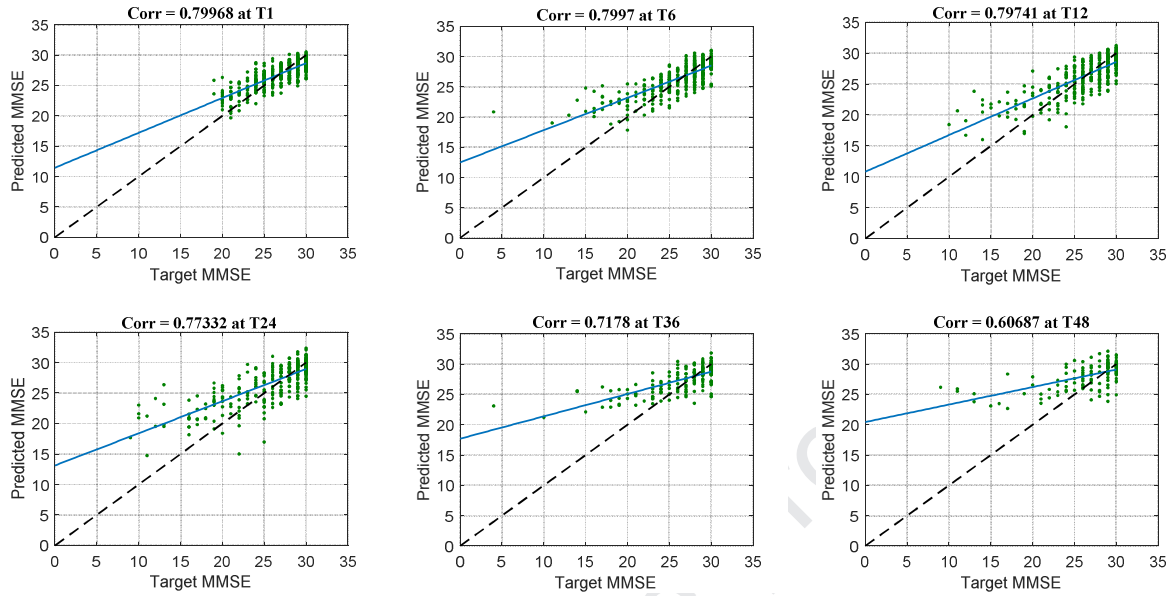


Fig. 5. Scatter plot of predicted MMSE scores versus actual values in six time points using the cognitive assessment modality. The green line is the regression line achieved by the winning coefficient matrix and the black dashed line is the reference for perfect correlation.

Table 3
Hyper parameters used for tuning of Gradient Boosting

Modality Combination (C)	max_depth	Learning rate	Colsample_bytree	n_estimators
MRI_PET	2	0.07	0.98	90
MRI_PET_CSF	3	0.05	1.00	120
MRI_PET_COG	3	0.07	1.00	90
PET_COG_CSF	3	0.07	0.98	80
MRI_PET_COG_CSF	3	0.10	0.50	50

4.4 Final results and discussion

In order to model the complex relationship between different modalities, the outcomes of the winning predictors from Fig. 4 are combined with the risk factor measurements, as non-temporal biomarkers. These new sets of features have been utilized as the input for the gradient boosting (GB) machines. The GB machines have been trained over five combinations of modalities. Grid search has been adopted to estimate the hyperparameters of gradient boosting for different combinations of modalities. The optimal hyperparameter values for each modality have been reported in Table 3. The experimental results, in terms of RMSE, are shown in Table 4.

For all methods reported in Table 4, the training and testing sets are identical, except for the fact that the competing methods are using the conventional approach in which all features from different modalities are concatenated together. For the statistical test, the correlation coefficient between the observed and predicted values is calculated on 100 bootstrapped samples, generated from the original sample size. By testing the null hypothesis of no correlation, the significance of the correlation, p-value, is calculated for each time point.

The proposed model achieved a correlation coefficient of 0.82 (p = 6.20e-47) at T1, 0.86 (p = 4.18e-62) at T6, 0.80 (p = 1.18e-41) at T12, 0.81 (p = 1.82e-38) at T24, 0.79 (p = 6.11e-20) at T36 and 0.76 (p = 1.44e-15) at T48 on the test data. Coefficient of determination is another statistical metric to evaluate the accuracy of regression models. This parameter presents the percentage of the variation in the dependent variable (predicted value) that can be described by the independent variable (target value). The coefficient of determination for the proposed model is 0.67 at T1, 0.73 at T6, 0.64 at T12, 0.66 at T24, 0.62 at T36 and 0.58 at T48. Fig. 6 shows the scatter plots of predicted MMSE scores versus the actual scores with correlation value reported within each scatter plot. Colors are representing groups of subjects belonging to different stages of AD. The progressive nature of AD results in a

steady, though uncertain slope in terms of cognitive decline. Patients who are diagnosed with late stages of AD at baseline have a higher chance to encounter a steep descent to severe cognitive decline within the following 48 months. Therefore, at the time points with an unbalanced population, in terms of the cognitive score distribution, individuals with a severely low MMSE score are detected as outliers. For example, according to Fig. 6, there are very few subjects with a cognitive score of less than ten, which makes it difficult for the system to keep track of all values. It should be pointed out that considering a weighting scheme of the distributions at the different stages of the disease and at different time points could help in improving the prediction accuracy of the trajectories in cognitive decline (Sugiyama et al, 2007).

Table 4: Comparison of the results from our proposed method with other existing methods on longitudinal multi modal data. The error has been reported using RMSE metric in six different future time points.

Method	Modality	Time Points					
		T1	**T6**	**T12**	**T24**	**T36**	**T48**
Ridge	MRI, PET, COG, CSF	1.90±0.47	2.33±0.68	2.43±0.74	3.17±0.73	3.20±0.83	4.05±0.90
Lasso	MRI, PET, COG, CSF	1.83±0.37	2.34±0.64	2.45±0.53	3.11±0.70	3.15±0.74	4.00±0.76
TGL	MRI, PET, COG, CSF	1.93±0.43	2.32±0.45	2.42±0.55	3.22±0.67	3.10±0.82	3.87±0.93
nCFGL1	MRI, PET, COG, CSF	1.81±0.55	2.31±0.58	2.41±0.67	3.28±0.46	3.49±0.59	4.06±0.70
cFSGL	MRI, PET, COG, CSF	1.88±0.85	2.33±0.64	2.40±0.73	3.20±0.68	3.03±0.86	3.61±0.78
$\ell_{2,1}$-norm	MRI, PET, COG, CSF	1.89±0.75	2.34±0.52	2.38±0.76	3.24±0.59	3.08±0.67	3.64±0.69
Zhu et al., 2016	MRI, PET, COG, CSF	1.87±0.52	2.31±0.66	2.32±0.50	3.27±0.62	2.98±0.96	3.56±0.87
Zhu et al., 2018	MRI, PET, COG, CSF	1.86±0.53	2.27±0.61	2.38±0.64	3.23±0.57	3.02±0.84	3.42±0.64
Proposed	MRI, PET	2.02±0.26	2.30±0.32	2.88±0.36	3.06±0.35	2.51±0.31	2.60±0.32
	MRI, PET, CSF	1.95±0.39	2.22±0.31	2.81±0.30	2.92±0.33	2.51±0.37	2.72±0.30
	MRI, PET, COG	1.60±0.27	1.79±0.23	2.30±0.23	2.41±0.35	2.53±0.32	2.20±0.30
	PET, COG, CSF	1.63±0.20	1.80±0.28	2.25±0.20	2.38±0.25	2.41±0.26	2.38±0.29
	MRI, PET, COG, CSF	1.62±0.24	1.78±0.22	2.24±0.24	2.38±0.21	2.28±0.22	2.19±0.15

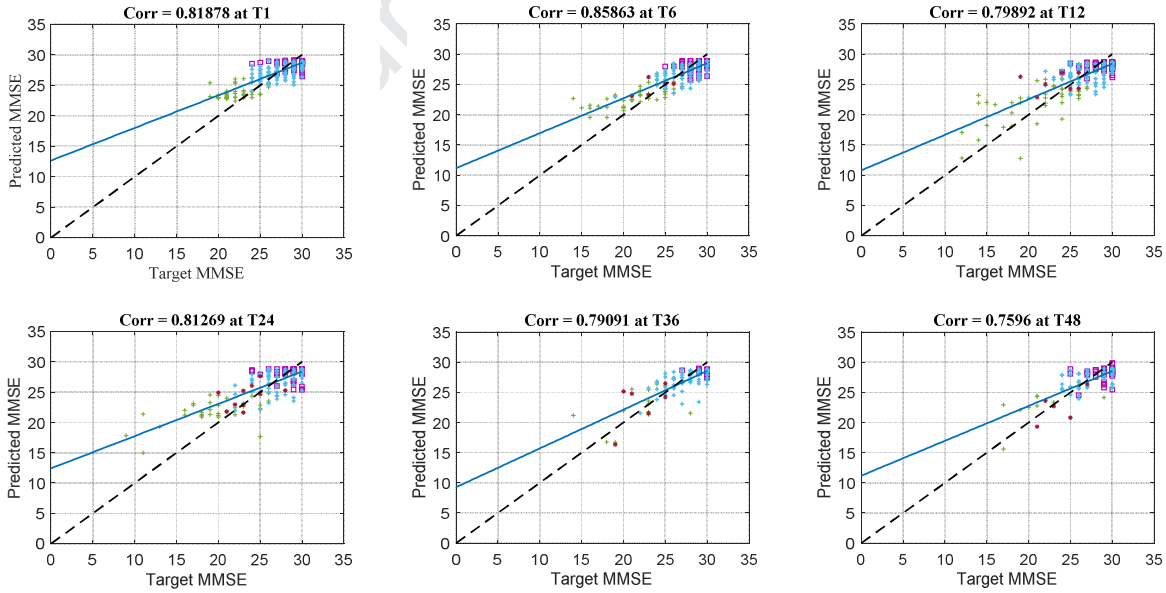


Fig. 6. Scatter plot of predicted MMSE scores versus actual values at six different time points. The blue line is the fitted regression line achieved by the proposed model and the dashed black line is the perfect correlation. Red squares (□) are the CN group, blue plus signs (+) are the MCI group, red asterisks (*) are the MCI converter group and green plus signs (+) are the AD group.

Since the focus of this paper is in predicting the trajectories of MMSE scores, the longitudinal distributions of predicted versus actual target MMSE scores for each group are provided in Fig. 7.

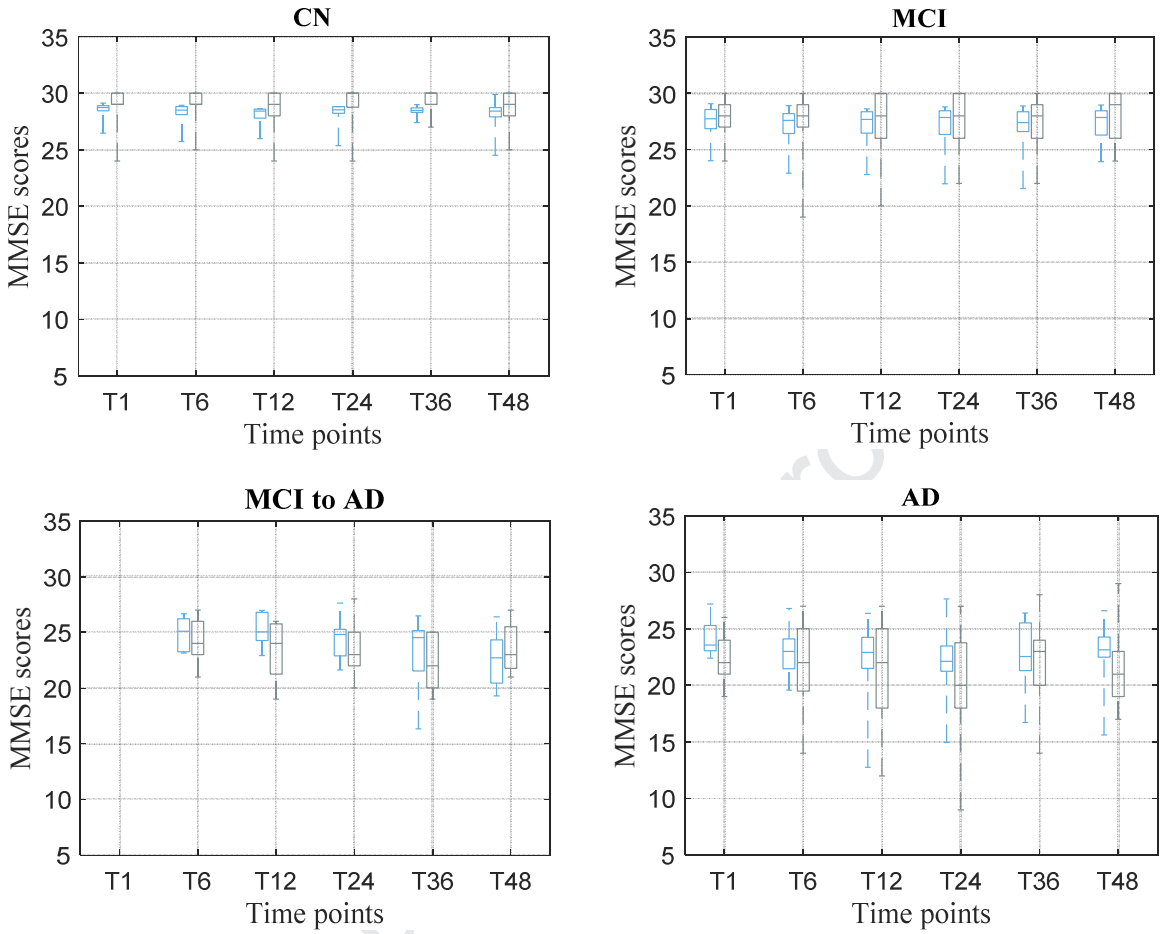


Fig. 7. Longitudinal trajectories of MMSE scores through 6 time points for each category of disease. In each figure, boxplots in blue are used for the distribution of predicted MMSE scores and black boxplots are used for the distribution of target MMSE scores.

To further evaluate the superiority of the proposed model, following the approach described in (Jie et al., 2017), paired *t*-test has been performed on the residuals of the proposed method and each of the competing method. The results summarized in Table 5 show that except for the baseline, the proposed method for all other five future time points demonstrates statistical significance, with all p-values less than 0.05, proving its effectiveness.

Table 5

Comparison of p-values obtained from residuals of the proposed method and the competing methods using the combination of modalities of MRI, PET, COG, CSF

	Ridge	**Lasso**	**TGL**	**nCFGL1**	**cFSGL**	$\ell_{2,1}$norm	**Zhu et al., 2016**	**Zhu et al., 2018**
T1	0.063	0.083	0.386	0.386	0.501	0.086	0.029	0.032
T6	0.007	0.011	0.003	0.001	< 0.001	< 0.001	0.024	0.013
T12	0.004	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.002	< 0.001
T24	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
T36	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.010	0.012
T48	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.013	0.010

Since independent models are separately trained over each feature space, our model brings the following advantages: (1) feature scarcity from one particular modality would not be an issue for the other regression models; (2) any error within the data of one modality could be prevented from propagating through other modalities; (3) the model could be easily extended to include other modality sources with little adjustments and to consider sparsity patterns of the measurements; (4) the proposed model is applicable to a wide variety of subjects with any combination of modality sources, without being restricted to their baseline diagnosis or to their historical records; and (5) the robustness and flexibility of the presented framework in handling missing data preserves enough information to monitor and predict MMSE trajectories with a relatively high accuracy.

5. Conclusion

Predicting MMSE over time, through multimodal longitudinal data, could augment our prospects for analyzing the interplay between the different multimodal features used in the input space in relation to the predicted MMSE scores. Such a prediction model could also be used to ascertain the effectiveness of treatment or therapeutic protocol by comparing actually taken MMSE tests against predicted scores by the model, allowing at the same time to observe the conversion rate in the different stages of individuals who are at risk of developing AD. A novel distributed multitask multimodal framework is introduced for predicting cognitive measures in the progression of Alzheimer's disease even when burdened with the missing data challenge. The model is capable of handling size discrepancy between the number of observations belonging to different time points and assuming different recording modalities. The proposed approach also has the potential to directly consider the inherent temporal sparsity patterns of different modalities and their relative correlation strength. This provides flexibility in utilizing complementary information from multimodal data. Furthermore, the model also terminates the propagation of potential error from one modality to another which may have originated from corrupted data.

It is important to emphasize that in designing the proposed prediction model, the Mini-Mental State Examination (MMSE) and Clinical Dementia Rating Sum of Boxes (CDRSB) scores (*since initially used for labelling subjects*) and Alzheimer's Disease Assessment Score (ADAS11, ADAS13) and the Montreal Cognitive Assessment (MoCA) (*since highly correlated with MMSE*) were excluded from the feature set or input space in the training and testing phases of the proposed prediction model. The longitudinal MMSE scores were instead used as labels to be predicted by the model on the basis of the multimodal feature set considered for the different time points as listed in Table 2. The experimental results proved that this method can effectively predict the progression of Alzheimer's disease over a period of four years in terms of the predicted MMSE scores on the basis of neuroimaging features (MRI and PET), cognitive tests not used initially for labelling the subjects or found to be highly correlated with MMSE to avoid any bias, cerebrospinal fluid (CSF) and other risk factors associated with age, gender, years of education, and the APOE gene. While the proposed approach mitigates the consequence of the negative correlation between various modalities, there could still be unrelated information between different tasks within a single modality. Future studies using longitudinal data may be able to improve the performance of these prediction algorithms. The general approach described for predicting progression used in this study, as expressed in Fig. 2, could be extended not only to other longitudinal studies involving other neurological disorders, but could also be used for the prediction of other cognitive scores such as ADAS11 and RAVLT to assess the singular merits of such cognitive scores and how related and correlated they may be to the MMSE test.

Acknowledgments

This work was supported by National Science Foundation (NSF) under NSF grants (CNS- 1920182, CNS-1532061, CNS-1551221, CNS-1338922), the 1Florida Alzheimer's Disease Research Center (ADRC) (NIA 1P50AG047266-01A1) and the Ware Foundation. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson

Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.;Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California.

REFERENCES

Alexander, G.E., Chen, K., Pietrini, P., Rapoport, S.I., Reiman, E.M., 2002. Longitudinal PET evaluation of cerebral metabolic decline in dementia: A potential outcome measure in Alzheimer's disease treatment studies. American Journal of Psychiatry 159, 738–745. https://doi.org/10.1176/appi.ajp.159.5.738

Alzheimer Association, 2016. 2016 Alzheimer's Disease Facts and Figures. Alzheimer's & Dementia 2016 12, 1–80. https://doi.org/10.1016/j.jalz.2016.03.001

Anoop, A., Singh, P.K., Jacob, R.S., Maji, S.K., 2010. CSF Biomarkers for Alzheimer's Disease Diagnosis. International Journal of Alzheimer's Disease 2010, 1–12. https://doi.org/10.4061/2010/606802

Azmi, M.H., Saripan, M.I., Nordin, A.J., Ahmad Saad, F.F., Abdul Aziz, S.A., Wan Adnan, W.A., 2017. 18F-FDG PET brain images as features for Alzheimer classification. Radiation Physics and Chemistry 137, 135–143. https://doi.org/10.1016/j.radphyschem.2016.08.028

Bakker, B., Heskes, T., 2003. Task Clustering and Gating for BayesianMultitask Learning. Journal of Machine Learning Research 1, 83–99. https://doi.org/10.1162/153244304322765658

Bi, J., Bi, J., Xiong, T., Xiong, T., Yu, S., Yu, S., Dundar, M., Dundar, M., Rao, R.B., Rao, R.B., 2008. An Improved Multi-task Learning Approach with Applications in Medical Diagnosis. Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databases-Part I 117–132.

Bi, X.-A., Shu, Q., Sun, Q., Xu, Q., 2018. Random support vector machine cluster analysis of resting-state fMRI in Alzheimer's disease. PLoS ONE 13, 1–17. https://doi.org/10.1371/journal.pone.0194479

Buckner, R.L., 2004. Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. Neuron 44, 195– 208.

Buerger, K., Teipel, S.J., Zinkowski, R., Blennow, K., Arai, H., Engel, R., Hofmann-Kiefer, K., McCulloch, C., Ptok, U., Heun, R., Andreasen, N., DeBernardis, J., Kerkman, D., Moeller, H.-J., Davies, P., Hampel, H., 2002. CSF tau protein phosphorylated at threonine-231 correlates with cognitive decline in MCI subjects. Neurology 59, 627–629.

Bussy, A., Snider, B.J., Coble, D., Xiong, C., Fagan, A.M., Cruchaga, C., Benzinger, T.L.S., Gordon, B.A., Hassenstab, J., Bateman, R.J., Morris, J.C., 2019. Effect of apolipoprotein E4 on clinical, neuroimaging, and biomarker measures in noncarrier participants in the Dominantly Inherited Alzheimer Network. Neurobiology of Aging 75, 42–50. https://doi.org/10.1016/j.neurobiolaging.2018.10.011

Cao, P., Liu, X., Yang, J., Zhao, D., Huang, M., Zaiane, O., 2018. $\ell 2,1-\ell 1$ multi-task representation learning based cognitive performance prediction of Alzheimer's disease. Pattern Recognition 79, 195–215. https://doi.org/10.1016/j.patcog.2018.01.028

Cao, P., Shan, X., Zhao, D., Huang, M., Zaiane, O., 2017. Sparse shared structure based multi-task learning for MRI based cognitive performance prediction of Alzheimer's disease. Pattern Recognition 72, 219–235. https://doi.org/10.1016/j.patcog.2017.07.018

Caruana, R., 1997. Multitask Learning. Machine Learning 28, 41–75. https://doi.org/10.1023/A:1007379606734

Chen, G., Chen, K.S., Knox, J.H., Inglis, J., Bernard, A., Martin, S.J., Justice, A., McConlogue, L., Games, D., Freedman, S.B., Morris, R.G.M., 2000. A learning deficit related to age and b-amyloid plaques in a mouse model of Alzheimer's disease. Nature 408, 975–979. https://doi.org/10.1038/35050103

Cheng, B., Liu, M., Suk, H. Il, Shen, D., Zhang, D., Munsell, B.C., Yang, Q., 2015. Multimodal manifold-regularized transfer learning for MCI conversion prediction. Brain Imaging and Behavior 9, 1805–1817. https://doi.org/10.1007/s11682-015-9356-x

Colijn, M.A., Grossberg, G.T., 2015. Amyloid and Tau Biomarkers in Subjective Cognitive Impairment. Journal of Alzheimer's Disease 47, 1–8. https://doi.org/10.3233/JAD-150180

Corder, A.E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, W., Roses, A.D., Haines, J.L., 2008. Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer ' s Disease in Late Onset Families Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/2882127 261, 921–923.

Curiel, R.E., Loewenstein, D.A., Rosselli, M., Penate, A., Greig-Custo, M.T., Bauer, R.M., Guinjoan, S.M., Hanson, K.S., Li, C., Lizarraga, G., Barker, W.W., Torres, V., DeKosky, S., Adjouadi, M., Duara, R., 2018. Semantic Intrusions and Failure to Recover From Semantic Interference in Mild Cognitive Impairment: Relationship to Amyloid and Cortical Thickness. Current Alzheimer Research 15, 848–855. https://doi.org/10.2174/1567205015666180427122746

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

De Leon, M.J., DeSanti, S., Zinkowski, R., Mehta, P.D., Pratico, D., Segal, S., Rusinek, H., Li, J., Tsui, W., Saint Louis, L.A., Clark, C.M., Tarshish, C., Li, Y., Lair, L., Javier, E., Rich, K., Lesbre, P., Mosconi, L., Reisberg, B., Sadowski, M., DeBernadis, J.F., Kerkman, D.J., Hampel, H., Wahlund, L.O., Davies, P., 2006. Longitudinal CSF and MRI biomarkers improve the diagnosis of mild cognitive impairment. Neurobiology of Aging 27, 394–401. https://doi.org/10.1016/j.neurobiolaging.2005.07.003

Dong, D., Wu, H., He, W., Yu, D., Wang, H., 2015. Multi-Task Learning for Multiple Language Translation. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) 1723–1732. https://doi.org/10.3115/v1/P15-1166

Doody, R.S., Pavlik, V., Massman, P., Rountree, S., Darby, E., Chan, W., 2010. Predicting progression of Alzheimer ' s disease. Alzheimer's research & therapy 77030. https://doi.org/10.1186/alzrt38

Duara, R., Barker, W., Loewenstein, D., Greig, M.T., Rodriguez, R., Goryawala, M., Zhou, Q., Adjouadi, M., 2015. Insights into cognitive aging and Alzheimer's disease using amyloid PET and structural MRI scans. Clinical and Translational Imaging. https://doi.org/10.1007/s40336-015-0110-6

Duara, R., Loewenstein, D., Lizarraga, G., Adjouadi, M., Barker, W.W., Greig-Custo, M.T., Rosselli, M., Penate, A., Shea, Y.F., Behar, R., Ollarves, A., Robayo, C., Hanson, K., Marsiske, M., Burke, S., Ertekin-Taner, N., Vaillancourt, D., De Santi, S., Golde, T., "Effect of age, ethnicity, sex, cognitive status and APOE genotype on amyloid load and the threshold for amyloid positivity", Neuroimage Clin. 2019 Mar 27;22:101800. doi: 10.1016/j.nicl.2019.101800. [Epub ahead of print] PMID: 30991618.

Duchesne, S., Caroli, A., Geroldi, C., Collins, D.L., Frisoni, G.B., 2009. Relating one-year cognitive change in mild cognitive impairment to baseline MRI features. NeuroImage 47, 1363–1370. https://doi.org/10.1016/j.neuroimage.2009.04.023

Emrani, S., Carolina, N., Mcguirk, A., Carolina, N., Carolina, N., 2017a. Prognosis and Diagnosis of Parkinson ' s Disease Using Multi-Task Learning 1457–1466. https://doi.org/10.1145/3097983.3098065

Emrani, S., McGuirk, A., Xiao, W., 2017b. Prognosis and Diagnosis of Parkinson's Disease Using Multi-Task Learning. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17 1457–1466. https://doi.org/10.1145/3097983.3098065

Evgeniou, T., Micchelli, C., Pontil, M., 2005. Learning multiple tasks with kernel methods. Jmlr 6, 615–637.

Farrer, L. a, Cupples, L.A., Haines, J.L., Hyman, B., Kukull, W. a, Mayeux, R., Myers, R.H., Pericak-Vance, M. a, Risch, N., van Duijn, C.M., 1997. Effects of Age, Sex, and Ethnicity on the Association Between Apolipoprotein E Genotype and Alzheimer Disease. The Journal of the American Medical Association 278, 1349–1356. https://doi.org/10.1001/jama.1997.03550160069041

Friedman, J., 2001. Greedy Function Approximation : A Gradient Boosting Machine Author (s): Jerome H . Friedman Source : The Annals of Statistics , Vol . 29 , No . 5 (Oct ., 2001), pp . 1189-1232 Published by : Institute of Mathematical Statistics Stable URL : http://www. The Annals of Statistics 29, 1189–1232. https://doi.org/10.1214/009053606000000795

Frisoni, G.B., Pievani, M., Testa, C., Sabattoli, F., Bresciani, L., Bonetti, M., Beltramello, A., Hayashi, K.M., Toga, A.W., Thompson, P.M., 2007. The topography of grey matter involvement in early and late onset Alzheimer's disease. Brain 130, 720–730. https://doi.org/10.1093/brain/awl377

Greenlaw, K., Szefer, E., Graham, J., Lesperance, M., Nathoo, F.S., 2017. A Bayesian group sparse multi-task regression model for imaging genetics. Bioinformatics (Oxford, England) 33, 2513–2522. https://doi.org/10.1093/bioinformatics/btx215

Hanger, D.P., Anderton, B.H., Noble, W., 2009. Tau phosphorylation: the therapeutic challenge for neurodegenerative disease. Trends in Molecular Medicine 15, 112–119. https://doi.org/10.1016/j.molmed.2009.01.003

Izquierdo, W., Martin, H., Cabrerizo, M., Barker, W.W., Loewenstein, D.A., Duara, R., Adjouadi, M., 2017. Predicting Cognitive Test Scores In Alzheimer's Patients Using Multimodal Longitudinal Data. Alzheimer's & Dementia: The Journal of the Alzheimer's Association 13, P796–P797. https://doi.org/10.1016/j.jalz.2017.06.1078

Jack, C.R., Wiste, H.J., Schwarz, C.G., Lowe, V.J., Senjem, M.L., Vemuri, P., Weigand, S.D., Therneau, T.M., Knopman, D.S., Gunter, J.L., Jones, D.T., Graff-Radford, J., Kantarci, K., Roberts, R.O., Mielke, M.M., Machulda, M.M., Petersen, R.C., 2018. Longitudinal tau PET in ageing and Alzheimer's disease. Brain. https://doi.org/10.1093/brain/awy059

Jedynak, B.M., Lang, A., Liu, B., Katz, E., Zhang, Y., Wyman, B.T., Raunig, D., Jedynak, C.P., Caffo, B., Prince, J.L., 2012. A computational neurodegenerative disease progression score: Method and results with the Alzheimer's disease neuroimaging initiative cohort. NeuroImage 63, 1478–1486. https://doi.org/10.1016/j.neuroimage.2012.07.059

Jie, B., Liu, M., Liu, J., Zhang, D., Shen, D., 2017. Temporally Constrained Group Sparse Learning for Longitudinal Data Analysis in Alzheimer's Disease. IEEE Transactions on Biomedical Engineering 64, 238–249. https://doi.org/10.1109/TBME.2016.2553663

Jie, B., Zhang, D., Cheng, B., Shen, D., 2015. Manifold regularized multitask feature learning for multimodality disease classification. Human Brain Mapping 36, 489–507. https://doi.org/10.1002/hbm.22642

Kumar, A., Daume, H., 2012. Learning Task Grouping and Overlap in Multi-task Learning.

Landau, S.M., Mintun, M. a., Joshi, A.D., Koeppe, R. a., Petersen, R.C., Aisen, P.S., Weiner, M.W., Jagust, W.J., 2012. Amyloid Deposition,

Hypometabolism, and Longitudinal Cognitive Decline. Ann Neurol 72, 578–586. https://doi.org/10.1002/ana.23650

Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C., 2004. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. NeuroImage 21, 46–57. https://doi.org/10.1016/j.neuroimage.2003.09.027

Lawlor, B.A., Ryan, T.M., Schmeidler, J., Mohs, R.C., Davis, K.L., 1994. Clinical symptoms associated with age at onset in Alzheimer's disease. Am J Psychiatry 151, 1646–1649.

Li, C., Loewenstein, D.A., Duara, R., Cabrerizo, M., Barker, W., Adjouadi, M., 2017. The relationship of brain amyloid load and APOE status to regional cortical thinning and cognition in the ADNI cohort. Journal of Alzheimer's Disease 59, 1269–1282. https://doi.org/10.3233/JAD-170286

Loewenstein DA, Curiel RE, DeKosky S, Bauer RM, Rosselli M, Guinjoan SM, Adjouadi M, Peñate A, Barker WW, Goenaga S, Golde T, Greig-Custo MT, Hanson KS, Li C, Lizarraga G, Marsiske M, Duara R, "Utilizing semantic intrusions to identify amyloid positivity in mild cognitive impairment", Neurology, Vol. 91 (10), pp. E976-E984, September 2018. PMID: 30076274.

Li, Y., Tian, X., Liu, T., Tao, D., 2017. On Better Exploring and Exploiting Task Relationships in Multitask Learning: Joint Model and Feature Learning. IEEE Transactions on Neural Networks and Learning Systems. https://doi.org/10.1109/TNNLS.2017.2690683

Li, Y., Wang, Y., Wu, G., Shi, F., Zhou, L., Lin, W., Shen, D., 2012. Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. Neurobiology of Aging 33, 1–29. https://doi.org/10.1016/j.neurobiolaging.2010.11.008

Lizarraga, G., Li, C., Cabrerizo, M., Barker, W., Loewenstein, D.A., Duara, R., Adjouadi, M., 2018. A neuroimaging web services interface as a cyber physical system for medical imaging and data management in brain research: Design study. Journal of Medical Internet Research 20, 1–17. https://doi.org/10.2196/medinform.9063

Loewenstein, D.A., Curiel, R.E., Wright, C., Sun, X., Alperin, N., Crocco, E., Czaja, S.J., Raffo, A., Penate, A., Melo, J., Capp, K., Gamez, M., Duara, R., 2017. Recovery from Proactive Semantic Interference in Mild Cognitive Impairment and Normal Aging: Relationship to Atrophy in Brain Regions Vulnerable to Alzheimer's Disease. Journal of Alzheimer's Disease 56, 1119–1126. https://doi.org/10.3233/JAD-160881

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehéricy, S., Benali, H., 2009. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51, 73–83. https://doi.org/10.1007/s00234-008-0463-x

Mendez, M.F., 2017. Early-Onset Alzheimer Disease. Neurologic Clinics. https://doi.org/10.1016/j.ncl.2017.01.005

Michaelson, D.M., 2014. APOE ε4: The most prevalent yet understudied risk factor for Alzheimer's disease. Alzheimer's and Dementia 10, 861–868. https://doi.org/10.1016/j.jalz.2014.06.015

Minhas, S., Khanum, A., Riaz, F., Khan, S., Alvi, A., 2017. Predicting Progression from Mild Cognitive Impairment to Alzheimer's Disease using Autoregressive Modelling of Longitudinal and Multimodal Biomarkers. IEEE Journal of Biomedical and Health Informatics 1–1. https://doi.org/10.1109/JBHI.2017.2703918

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage 104, 398–412. https://doi.org/10.1016/j.neuroimage.2014.10.002

Mungas, D., Reed, B.R., Ellis, W.G., Jagust, W.J., 2001. The effects of age on rate of progression of Alzheimer disease and dementia with associated cerebrovascular disease. Archives of neurology 58, 1243–7. https://doi.org/10.1001/archneur.58.8.1243

Nie, L., Zhang, L., Meng, L., Song, X., Chang, X., Li, X., 2017. Modeling Disease Progression via Multisource Multitask Learners: A Case Study with Alzheimer's Disease. IEEE Transactions on Neural Networks and Learning Systems 28, 1508–1519. https://doi.org/10.1109/TNNLS.2016.2520964

Nimmy John, T., D. Puthankattil, S., Menon, R., John, T.N., Puthankattil, S.D., Menon, R., 2018. Analysis of long range dependence in the EEG signals of Alzheimer patients. Cognitive Neurodynamics 12, 183–199. https://doi.org/10.1007/s11571-017-9467-8

Noble, W., Hanger, D.P., Miller, C.C.J., Lovestone, S., 2013. The Importance of Tau Phosphorylation for Neurodegenerative Diseases. Frontiers in Neurology 4. https://doi.org/10.3389/fneur.2013.00083

Ogutu, J.O., Piepho, H.P., Schulz-Streeck, T., 2011. A comparison of random forests, boosting and support vector machines for genomic selection. BMC Proceedings 5, S11. https://doi.org/10.1186/1753-6561-5-S3-S11

Cohen, A.D., Klunk, W.E., 2014. Early detection of Alzheimer's disease using PiB and FDG PET. Neurobiol. Dis. https://doi.org/10.1016/j.nbd.2014.05.001

Pierce, A.L., Bullain, S.S., Kawas, C.H., 2017. Late-Onset Alzheimer Disease. Neurologic Clinics of NA 35, 283–293. https://doi.org/10.1016/j.ncl.2017.01.006

Poil, S.S., de Haan, W., van der Flier, W.M., Mansvelder, H.D., Scheltens, P., Linkenkaer-Hansen, K., 2013. Integrative EEG biomarkers predict progression to Alzheimer's disease at the MCI stage. Frontiers in Aging Neuroscience 5, 1–12. https://doi.org/10.3389/fnagi.2013.00058

Ritter, K., Schumacher, J., Weygandt, M., Buchert, R., Allefeld, C., Haynes, J.D., 2015. Multimodal prediction of conversion to Alzheimer's disease based onincomplete biomarkers. Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring 1, 206–215.

https://doi.org/10.1016/j.dadm.2015.01.006

Rogers, J.A., Polhamus, D., Gillespie, W.R., Ito, K., Romero, K., Qiu, R., Stephenson, D., Gastonguay, M.R., Corrigan, B., 2012. Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: A beta regression meta-analysis. Journal of Pharmacokinetics and Pharmacodynamics 39, 479–498. https://doi.org/10.1007/s10928-012-9263-3

Sargolzaei, S., Sargolzaei, A., Cabrerizo, M., Chen, G., Goryawala, M., Pinzon-Ardila, A., Gonzalez-Arias, S.M., Adjouadi, M., 2015. Estimating Intracranial Volume in Brain Research: An Evaluation of Methods. Neuroinformatics 13, 427–441. https://doi.org/10.1007/s12021-015-9266-5

Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V.M.Y., Trojanowski, J.Q., 2009. Cerebrospinal fluid biomarker signature in alzheimer's disease neuroimaging initiative subjects. Annals of Neurology 65, 403–413. https://doi.org/10.1002/ana.21610

Sørensen, L., Igel, C., Liv Hansen, N., Osler, M., Lauritzen, M., Rostrup, E., Nielsen, M., 2016. Early detection of Alzheimer's disease using MRI hippocampal texture. Human Brain Mapping 37, 1148–1161. https://doi.org/10.1002/hbm.23091

Stern, Y., 2012. Cognitive reserve in ageing and Alzheimer's disease. The Lancet Neurology 11, 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6

Stonnington, C.M., Chu, C., Klöppel, S., Jack, C.R., Ashburner, J., Frackowiak, R.S.J., 2010. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. NeuroImage 51, 1405–1413. https://doi.org/10.1016/j.neuroimage.2010.03.051

Sugiyama, M., Krauledat, M., Müller, K.-R., 2007. Covariate shift adaptation by importance weighted cross validation. J. Mach. Learn. Res. 8, 1027–1061.

Suk, H.-I., Lee, S.-W., Shen, D., 2017. Deep ensemble learning of sparse regression models for brain disease diagnosis. Medical Image Analysis 37, 101–113. https://doi.org/10.1016/j.media.2017.01.008

Tierney, M.C., Szalai, J.P., Snow, W.G., 1996. Prediction of probable Alzheimer ' s disease in memory-impaired patients : Neurology 46, 661–665. https://doi.org/10.1212/WNL.46.3.661

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

Trushina, E., Dutta, T., Persson, X.M.T., Mielke, M.M., Petersen, R.C., 2013. Identification of Altered Metabolic Pathways in Plasma and CSF in Mild Cognitive Impairment and Alzheimer's Disease Using Metabolomics. PLoS ONE 8. https://doi.org/10.1371/journal.pone.0063644

Van Der Flier, W.M., Scheltens, P., 2009. Alzheimer disease: Hippocampal volume loss and Alzheimer disease progression. Nature Reviews Neurology. https://doi.org/10.1038/nrneurol.2009.94

Wang, H., Nie, F., Huang, H., Risacher, S., Ding, C., Saykin, A.J., Shen, L., 2011. Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance, in: Proceedings of the IEEE International Conference on Computer Vision. pp. 557–562. https://doi.org/10.1109/ICCV.2011.6126288

Wang, H., Nie, F., Huang, H., Yan, J., Kim, S., Risacher, S., Saykin, A., Shen, L., 2012. High-order multi-task feature learning to identify longitudinal phenotypic markers for Alzheimer's disease progression prediction. In: Bartlett, P., Pereira, F., Burges, C., Bottou, L., Weinberger, K. (Eds.), Advances in Neural Information Processing Systems, 25, pp. 1286–1294.

Weinberger, K., Dasgupta, A., Attenberg, J., Langford, J., Smola, A., 2009. Feature Hashing for Large Scale Multitask Learning. https://doi.org/10.1145/1553374.1553516

Westman, E., Muehlboeck, J.S., Simmons, A., 2012. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. NeuroImage 62, 229–238. https://doi.org/10.1016/j.neuroimage.2012.04.056

Widmer, C., Org, C.M., Rätsch, G., 2012. Multitask Learning in Computational Biology. Conference Proceedings 27, 207–216.

Wolfe, M.S., 2016. Prospects and Challenges for Alzheimer Therapeutics, in: Developing Therapeutics for Alzheimer's Disease: Progress and Challenges. pp. 605–637. https://doi.org/10.1016/B978-0-12-802173-6.00023-X

Xue, Y., Liao, X., Carin, L., Krishnapuram, B., Com, B.K., 2007. Multi-Task Learning for Classification with Dirichlet Process Priors. Journal of Machine Learning Research 8, 35–63.

Yang, Z., Rong, J., Steven C., H.H., 2010. Exclusive Lasso for Multi-task Feature Selection. Aistats 9, 988–995.

Zhang, D., Shen, D., 2012. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. PLoS ONE 7. https://doi.org/10.1371/journal.pone.0033182

Zhang, Daoqiuang; Shen, D., 2013a. Multi modal multi task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. Neuroimage 59, 895–907. https://doi.org/10.1016/j.neuroimage.2011.09.069

Zhang, Y., Yeung, D.-Y., 2012. A Convex Formulation for Learning Task Relationships in Multi-Task Learning.

Zhang, Y., Yeung, D.Y., 2011. Multi-Task Learning in Heterogeneous Feature Spaces. Aaai 1, 1.

Zhou, J., Chen, J., Ye, J., 2012a. User's Manual MALSAR: Multi-tAsk Learning via StructurAl Regularization. Arizona State University.

Zhou, J., Liu, J., Narayan, V.A., Ye, J., 2012b. Modeling disease progression via fused sparse group lasso. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12 1095.

https://doi.org/10.1145/2339530.2339702

Zhu, X., Li, H., Fan, Y., n.d. Parameter-Free Centralized Multi-Task Learning for Characterizing Developmental Sex Differences in Resting State Functional Connectivity 8.

Zhu, X., Suk, H.-I., Lee, S.-W., Shen, D., 2016. Subspace Regularized Sparse Multitask Learning for Multiclass Neurodegenerative Disease Identification. IEEE Transactions on Biomedical Engineering 63, 607–618. https://doi.org/10.1109/TBME.2015.2466616

Zhu, X., Suk, H. Il, Wang, L., Lee, S.W., Shen, D., 2017. A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. Medical Image Analysis 38, 205–214. https://doi.org/10.1016/j.media.2015.10.008

Zhu, Y., Zhu, X., Kim, M., Shen, D., Wu, G., 2016. Early Diagnosis of Alzheimer's Disease by Joint Feature Selection and Classification on Temporally Structured Support Vector Machine, in: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Springer International Publishing, Cham, pp. 264–272. https://doi.org/10.1007/978-3-319-46720-7_31

Highlights

- A novel machine learning algorithm is proposed for predicting the progression of Alzheimer's disease using a distributed multimodal, multitask learning method.
- A new approach for predicting longitudinal trajectories of cognitive decline up to 48th month.
- The missing data challenge (missing modality, missing follow-up visit and drop out) is handled.
- Ability to capture complex relationships between different modalities while ignoring nonrelevant information.

Ethical Statement for NeuroImage

I testify on behalf of all co-authors that our article submitted to NeuroImage entitled " A Distributed Multitask Multimodal Approach for the Prediction of Alzheimer's Disease in a Longitudinal Study" that:

1) This material has not been published in whole or in part elsewhere;
2) The manuscript is not currently being considered for publication in any other journal;
3) All authors have been personally and actively involved in substantive work leading to the manuscript and are responsible for its content;
4) All authors do not have any potential conflict of interest

Date:
10/24/2019

Corresponding author's signature: