Proceedings Paper:

Spatio-temporal Similarity Measure based Multi-Task Learning for Predicting Alzheimer's Disease Progression using MRI data

Xulong Wang
Department of Computer Science
University of Sheffield
Sheffield, UK
xl.wang@sheffield.ac.uk

Yu Zhang
Department of Computer Science
University of Sheffield
Sheffield, UK
yzhang489@sheffield.ac.uk

Menghui Zhou
Department of Computer Science
University of Sheffield
Sheffield, UK
mzhou47@sheffield.ac.uk

Tong Liu
Department of Computer Science
University of Sheffield
Sheffield, UK
tliu.soton@gmail.com

Jun Qi
Department of Computing
Xi'an JiaoTong-Liverpool University
Suzhou, China
Jun.Qi@xjtlu.edu.cn

Po Yang
Department of Computer Science
University of Sheffield
Sheffield, UK
po.yang@sheffield.ac.uk

Abstract—Identifying and utilising various biomarkers for tracking Alzheimer's disease (AD) progression have received many recent attentions and enable helping clinicians make the prompt decisions. Traditional progression models focus on extracting morphological biomarkers in regions of interest (ROIs) from MRI/PET images, such as regional average cortical thickness and regional volume. They are effective but ignore the relationships between brain ROIs over time, which would lead to synergistic deterioration. For exploring the synergistic deteriorating relationship between these biomarkers, in this paper, we propose a novel spatio-temporal similarity measure based multi-task learning approach for effectively predicting AD progression and sensitively capturing the critical relationships between biomarkers. Specifically, we firstly define a temporal measure for estimating the magnitude and velocity of biomarker change over time, which indicate a changing trend(temporal). Converting this trend into the vector, we then compare this variability between biomarkers in a unified vector space(spatial). The experimental results show that compared with directly ROI based feature learning, our proposed method is more effective in predicting disease progression. Our method also enables performing longitudinal stability selection to identify the changing relationships between biomarkers, which play a key role in disease progression. We prove that the synergistic deteriorating biomarkers between cortical volumes or surface areas have a significant effect on the cognitive prediction.

Index Terms—Alzheimer's disease, brain biomarker correlation, cosine similarity, multi-task learning

I. INTRODUCTION

Alzheimer's disease (AD) is a serious neurodegenerative disease, which is characterized by memory loss and cognitive decline due to the progressive damage of neurons and their connections, which directly leads to death [1]. According to World Health Organization (WHO), it is estimated that there are globally 47.5 million people with dementia in 2016 with 7.7 million new cases every year. Previous research has focused on using biomarkers combined with machine learning algorithms to predict patients' Mini Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale cognitive subscale (ADAS-Cog) scores as the target data to predict whether a patient is an AD patient and find the weight of each biomarker feature at different prediction time points, existing AD disease progression models mainly use machine learning regression algorithms [2], survival models based on statistical probabilities [3], [4], and deep learning methods based on neural networks [5]–[7]. The above-mentioned research focuses on using the data obtained by the patient during the first test (baseline data) to make predictions, which is a method that uses a small number of input features to make predictions. The disadvantage is that it ignores the information contained in the biomarkers in the process of changing over time.

Some previous studies in brain science have studied the differences in the correlation between brain biomarkers for AD, cognitively normal older individuals (NL) and mild cognitive impairment (MCI). [8] proposed a deformation-based framework to jointly model the effects of aging and AD on the evolution of brain morphology, confirming the existence of components that significantly accelerate aging in AD patients, while highlighting the Specific morphological changes can help identify clinical conditions, even in the prodromal phase. [9] evaluated the correlation of MRI and CSF biomarkers with clinical diagnosis and cognitive performance in subjects with NL and aMCI (amnestic mild cognitive impairment) and AD patients. It is concluded that MRI provides stronger cross-sectional grouping and recognition ability and has better correlation with general cognitive and functional status on the cross-section, and although MRI and CSF provide complementary information, MRI better reflected the clinically determined disease stage than CSF biomarkers. Some previous studies
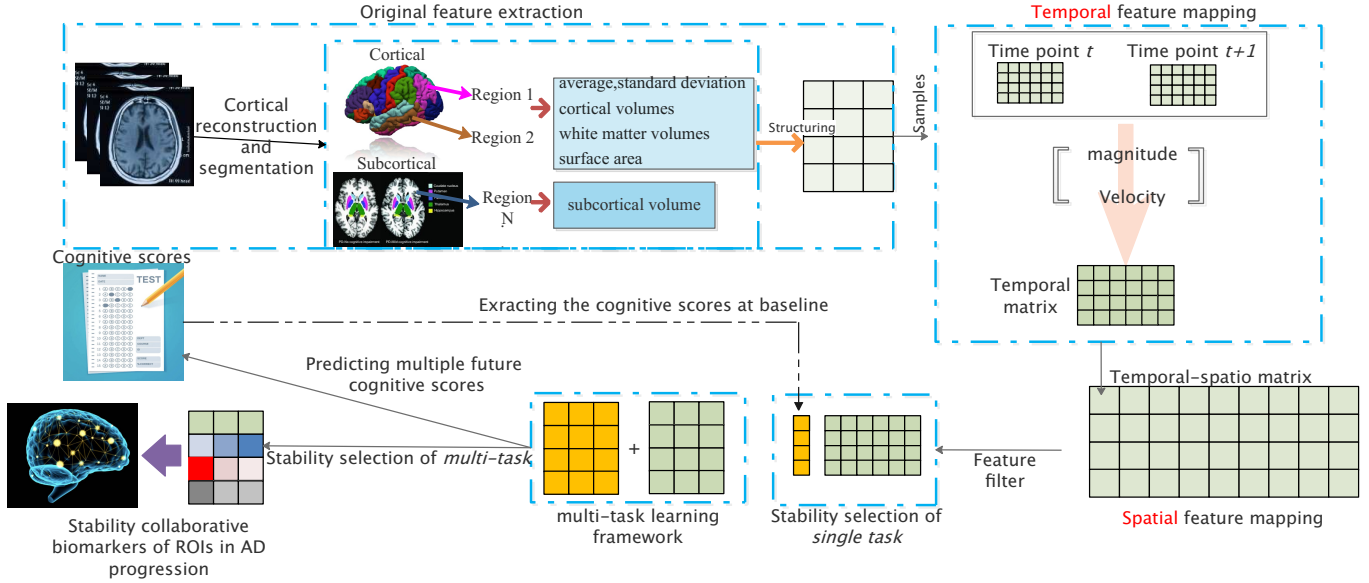
Fig. 1. The workflow of proposed MTL approach using spatio-temporal measure.

focused on the similarity between biomarkers form ROIs, [10] employed the correlation of regional average cortical thickness and multi-kernel support vector machine to integrate relevant information with ROI-based information to improve the classification performance. However, the above-mentioned researches only focused on the use of a single biomarker or the same type of biomarkers and did not focus on the relationships of temporal and spatial changes between different types of biomarkers.

To address the above challenges and uncover the critical relationships between biomarkers, we propose to utilise the temporal and spatial information of brain changes to model the disease process of AD. Additionally, to reinforce temporal relationships between follow-up time points, a multi-task learning method [11] based on temporal smoothness is introduced for interpretably modelling disease progression.

In this paper, we propose to utilise the spatio-temporal similarity between biomarkers changes to predict clinical scores of patients. Specifically, we firstly define a temporal measure for estimating the magnitude and velocity of biomarker change over time, which indicate a changing trend(Fig.1:temporal feature mapping). Converting this trend into the vector, we then compare this variability between biomarkers in a unified vector space(Fig.1:spatial feature mapping). The computation of spatial similarity results in an increase in data dimension by an order of magnitude of square. Faced with the scarcity of samples and a large number of feature dimensions, we introduce multiple loss terms with $\mathcal{L}_1$ [12] and its variant norm [11] to overcome the Curse of dimensionality and interpretably capture the key relationships. The contributions of this work are summarized as follows:

- A novel spatio-temporal similarity measure approach is

proposed of analysing and extracting reliable features from MRI. This similarity measure will effectively quantify the synergistic deterioration between these biomarkers over time;
- A multi-task feature learning algorithm with spatio-temporal embedding is designed for effectively predicting AD progression, visualising brain biomarkers related to this progression;
- A comprehensive experimental analysis is carried out by accessing impact of AD progression on brain function synergistic deteriorating biomarkers.

II. RELATED WORK

In traditional machine learning paradigm, an accurate learner is usually treated as one single learning task (e.g., classification, regression) and learnt by a large number of training samples. For instance, deep learning model [5], [13] can train an accurate AD prediction model of neural network with hundreds of layers contacting a great amount of parameters via massive labelled biomarkers at baseline from ADNI. But one key challenge here is that sufficient and well-labelled longitudinal AD data at multiple time points are hardly collected from AD patients. The problem of missing, sparse and insufficient data strongly impacts on learning a fine model. Differing with traditional ML approaches, Multi-Task Learning (MTL) [14] considers the prediction of AD progression as multiple learning tasks each of which can be a general prediction task art certain time point. Among these prediction tasks, all of them are assumed to be related to each other in time domain with relevant temporal features (e.g., biomarkers in MRI). We demonstrate a typical pipeline of leveraging MTL algorithms for predicting cognitive functionality of AD patients from their

brain imaging scans [15], where the predictive information is shared and transferred among related models to reinforce their generalization performance. The data sources employed are Freesufer (Extracted features from MRI like Volume of Hippocampus) and cognitive functional scores (AD cognitive scores like MMSE [16] or ADAS-cog [17]) from selected AD patients repeatedly by multiple time points. By considering the prediction of cognitive scores at a single time point (like 6, 12 or 18 months) as a regression task. The prediction of clinical scores at multiple future time points as a multi-task regression problem. MTL model matrix is trained and optimized through processing pre-extracted features from MRI and baseline cognitive scores.

Two important issues affect the progress of applying MTL in AD modelling problems. First, it is important to obtain good quality of baselines from AD raw data, where Magnetic resonance imaging (MRI) reflects changes in brain structure, such as the cerebral cortex and ventricle; cognitive score directly shows cognitive functions of AD patients. Sparse representation [18] is a popular method in MTL for capturing key biomarkers in AD, which uses sparseness as a regularization condition, image blocks with key characteristics. Cognitive measure can be achieved by using worldwide standard AD cognitive assessment, such as Mini Mental State Exam score (MMSE) [16], Alzheimer's Disease Assessment Scale cognitive total score (ADAS-cog) [17] and Rey Auditory Verbal Learning Test (RAVLT) [19], [20]. As the second issue, utilizing and improving advanced regression models [21] in MTL are highly critical, where they could better explore the relationship and correlations between MRI features and cognitive measures. Here, structural regularization [11] is a common approach in MTL for minimize the penalized empirical loss and bundling the correlations between tasks in the assumption. In the field of MTL in AD, there are many prior work that model relationships among tasks using novel regularizations [22], [23]. The addition of kernel method problems allows the algorithm to fit non-linear relationships [24]. The benchmark of this paradigm is derived from [25] and subsequent achievements are mostly aimed at theoretical structure, relevance, and fusing the multi-modality data applications. So far to our best knowledge, above regularized MTL approaches deliver promising performance in many AD prediction applications.

III. PROPOSED APPROACH

A. Spatio-temporal change similarity calculation of MRI biomarkers

Two consecutive MRI scans are used to calculate the temporal and spatial changes of brain biomarkers. For instance, we utilise BL and M06 MRI to calculate the magnitude and velocity for biomarkers, let x be the detection value of brain biomarkers and t be the MRI test dates, the magnitude is $\frac{x_{M06}-x_{BL}}{x_{BL}}$, the velocity is $\frac{x_{M06}-x_{BL}}{t_{M06}-t_{BL}}$ per month. Use the magnitude and velocity to compose a vector that represents the changing trend of the brain biomarker.

Cosine similarity is used to calculate the similarity between two vectors to express the similarity of the temporal and spatial changes of two MRI biomarkers. Cosine similarity uses the cosine value of the angle between two vectors in the vector space as a measure of the difference between two individuals. As the values of different types of biomarkers are different in MRI dataset, while the cosine similarity measures the difference in trend rather than the value. The temporal and spatial relationships of brain biomarkers of AD, NL and MCI displayed by Cosine similarity, Euclidean distance and Mahalanobis distance. Compared with Mahalanobis distance, the difference in the distribution of Cosine similarity is greater between AD, NL, and MCI.

B. Regression model via structural regularization

Regression analysis has been widely used in statistical, medical and industrial applications. It is a mathematical and statistical analysis of dependent influences (independent variables) and predictors (dependent variables). Its strength lies in its strong interpretation. By fitting the data, the parameter values corresponding to the independent variable indicate its effect on the dependent variable.

Considering the problem of prediction as a linear model. In order to obtain models with generalizability, loss functions with empirical structural loss risk minimization $L(Y, X, W)$ and the regularization.

The regularization term is considered as the addition of a prior, and common paradigms are Ridge regression and Lasso, which respectively add the $\mathcal{L}_2$ and $\mathcal{L}1$ norm. Statistical theory can prove that Ridge regression specifies a prior that the model obeys a Gaussian distribution and Lasso specifies a prior that the model obeys a Laplace distribution. This regularization term can be expressed as:

$$\min_w L(Y, X, W) + \lambda ||W||_1 \tag{1}$$

$$\min_w L(Y, X, W) + \lambda ||W||_2 \tag{2}$$

Ridge regression constrains variables to a smaller range for reducing some factors with little impacts on model's prediction. Unfortunately, this reduction means that these variables are still considered. To solve this problem, Lasso was proposed as a new sparse representation linear algorithm, which simultaneously performs feature selection and regression. Some variables are set to zero directly to achieve sparsity and dimensionality reduction.

C. Experiment protocol

Firstly, we verified that MTFL is superior in following AD progression and combined with randomization techniques to locate stable and sensitive cortical biomarkers identified by MTEN. Our empirical protocol design is based on a pipeline shown in Fig.1. The complete experimental process mainly includes 7 steps:

1) Original feature extraction, which performed the operation of transform the MRI to structural data that machine learning algorithm can be recognized.

2) Spatial feature mapping, which enumerates all possible associations expressed synergistically between ROIs.
3) Spatial feature mapping, which considers the variations of spatial associations from time t to time $t+1$(or $t+n$).
4) feature selection. Through this stage, the dimensionality of the features is greatly reduced, and the core information is retained to the greatest extent possible. Noting the cognitive scores at baseline period were used to complete feature selection.
5) Utilising the MTL algorithm to fit data.
6) Using trained models to predict cognitive scores.
7) Embedding MTL methods in the general stability selection to excavate stability collaborative biomarkers of ROIs in AD progression.

Secondly, the evaluation metric of cross-validation is employed to evaluate the performance of AD progression model. When a metric is set in the cross-validation experiment process, a set of hyper-parameters can be obtained. By comparing the pros and cons of the results, the suitable metric for the model is finally determined. The regression performance metric often employed in MTL is normalized mean square error (nMSE) and root mean square error (rMSE) is employed to measure the performance of each specific regression task. In particular, nMSE has been normalized to each task before evaluation, so it is widely used in MTL methods based on regression tasks. Also, weighted correlation coefficient (wR) as employed in the medical literature addressing AD progression problems [25]–[27]. nMSE, rMSE and wR are defined as follows:

$$\text{nMSE}(Y, \hat{Y}) = \frac{\sum_{i=1}^{t} \left\| Y_i -, \hat{Y}_i \right\|_2^2 / \sigma\left(Y_i\right)}{\sum_{i=1}^{t} n_i} \quad (3)$$

$$\text{rMSE}(y, \hat{y}) = \sqrt{\frac{\|y - \hat{y}\|_2^2}{n}} \quad (4)$$

$$\text{wR}(Y, \hat{Y}) = \frac{\sum_{i=1}^{t} \text{Corr}\left(Y_i, \hat{Y}_i\right) n_i}{\sum_{i=1}^{t} n_i} \quad (5)$$

Finally, as for repeated experimental times, one evaluation consensus in MTL models for AD study is that one experiment result is usually accidental and unreliable. To reduce experiment accidental errors, repeated experiments are required. So, we evaluate the performance of four selected regularized MTL models under different repeated experimental times and lastly evaluate typical factors like data size and number of tasks affecting MTL models.

D. Multi-task learning

Consider a multi-task learning of k tasks with n training samples of d features. Let $x_1, x_2, ..., x_n$ be the input data for the samples, and $y_1, y_2, ..., y_n$ be the predicted value for each sample, where each $x_i \in \mathbb{R}^d$ represents the feature data of an AD patient, and $y_i \in \mathbb{R}$ is the predicted value of cognitive score of different types of scales. Specifically, $x_i^j = [\text{m}, \text{v}]$ denotes spatio-temporal ROIs similarity on the j_{th} feature of the i_{th} sample, m, v represent the magnitude and velocity of two specific biomarkers (j_{th} and $(j+k)_{th}$; $j, (j+k) \in (0, d]$) over time separately.

Then, let $X = [x_1, ..., x_n]^T \in \mathbb{R}^{n \times d}$ be the data matrix, $Y = [y_1, ..., y_n]^T \in \mathbb{R}^{n \times k}$ be the predicted matrix, and $W = [w_1, ..., w_k]^T \in \mathbb{R}^{d \times k}$ be the weight matrix. The process of establishing a MTL model is to estimate the value of W, which is the parameter to be estimated from the training samples.

In order to solve above problem, many prior works in MTL that model relationships among tasks using regularization methods. Normally, they assume the empirical loss to be square loss and common regularization terms are $\mathcal{L}_1$ and $\mathcal{L}_2$ norms, separately named as Lasso regression and ridge regression models as shown in Eq. 6 and 7. Ridge regression constrains variables to a smaller range for reducing some factors with little impacts on model's prediction. Unfortunately, this reduction means that these variables are still considered. To solve this problem, Lasso was proposed as a new sparse representation linear algorithm, which simultaneously performs feature selection and regression. Some variables are set to zero directly to achieve sparsity and dimensionality reduction.

$$\min_w L(Y, X, W) + \lambda ||W||_1 \quad (6)$$

$$\min_w L(Y, X, W) + \lambda ||W||_2 \quad (7)$$

In AD study, the task of predicting AD patient's cognitive score at certain time point is strongly associated with other tasks at adjacent time points. Thus, many recent studies have focused on designing novel structural regularization methods to improve their performance in AD study.

In this paper, we concentrate on two AD progression prediction models : Temporal Group Lasso (TGL) [15] and Convex Fused Sparse Group Lasso (cFSGL) [28]. Specifically, TGL contains a time smoothing term and a group Lasso term as constraints, which ensures that all regression models at different time points share a common set of features. The TGL formulation solves the following convex optimization problem:

$$\min_w ||XW - Y||_F^2 + \theta_1 ||W||_F^2 + \theta_2 ||WH||_F^2 + \delta ||W||_{2,1} \quad (8)$$

where the first term measures the empirical error on the training data, $||W||_F$ is the Frobenius norm, $||WH||_F^2$ is the temporal smoothness term, which ensures a small deviation between two regression models at successive time points, and $||W||_{2,1}$ is the group lasso penalty, which ensures that a small subset of features will be selected for the regression models at all-time points.

cFSGL involves sparsity between tasks, where it considers both common features at different points in time and unique features to each task. This feature is helpful to improve the overall performance of the model. cFSGL formulation solves the following convex optimization problem:

$$\min_w ||XW - Y||_F^2 + \theta_1 ||W||_1 + \theta_2 ||RW^T||_1 + \delta ||W||_{2,1} \quad (9)$$

where the first term measures the empirical error on the training data, $||W||_1$ is the lasso penalty, $||RW^T||_1$ is the fused lasso penalty, and $||W||_{2,1}$ is the group lasso penalty.

Lasso and group lasso combined employ is called sparse group lasso, which allows simultaneous selection of a common feature for all time points and internally generates sparse solutions in response to different time points. Fused lasso penalty having a given temporal smoothness, which makes selected features at nearby time points similar to each other. In addition, notice that cFSGL's formula involves three non-smooth terms. Accelerated gradient descent method is utilised to solve this problem.

E. Stability Selection via structural MTL

In order to improve the interpretability and robustness of the results, stability selection was modified to meet our actual needs. The original strategy of feature selection was included a Lasso algorithm as core feature subsets searches approaches. In this paper, MTL algorithms were utilised to embedded in stability selection.

Let F be the overall set of features and let $f \in F$ be the subset of features by sub-sampling. Let γ denote the iteration number of sub-sampling and $Di = \{X(i), Y(i)\}$ denote one random sub-sample operation of number $i \in (0, \gamma]$. Each operation size account for $\lfloor \frac{n}{2} \rfloor$. Let Λ be the regularization parameter space. For a $\lambda \in \Lambda$, let $\hat{W}^{(i)}$ denote the model coefficient of MTFL that fitted on a subset of $D(i)$. Then, the subset of features generated in task j by the sparse constraints of the MTFL algorithm can be denote as:

$$S_j^\lambda \left(D_{(i)} \right) = \left\{ f : \hat{W}_j^{(i)} \neq 0 \right\}. \tag{10}$$

With stability selection, we do not simply select one model in the parameter space λ. Instead the data are perturbed (e.g. by sub-sampling) γ times at task j and we choose all structures or variables that occur in a large fraction of the resulting selection sets:

$$\hat{\pi}_j^\lambda = \frac{\sum_{i=1}^\gamma I \left(f \in S_j^\lambda \left(D_{ij} \right) \right)}{\gamma}. \tag{11}$$

Where indicator function $I(\bullet)$ denote $I(x) = \begin{cases} 1, & x = 0 \\ 0, & others \end{cases}$ and $\hat{\pi}_j^\lambda \in [0, 1]$ denote the stability probability of task j at MTFL approaches which feature selection is not based on individual operations but on multiple task collaboration constraints.

Repeat the above procedure for all $\lambda \in \Lambda$, we obtain the stability score $S_j(f)$ for each feature f at task j:

$$S_j(f) = \max_{\lambda \in \Lambda} \left(\hat{\pi}_j^\lambda \right). \tag{12}$$

Finally, for a cut-off π_{th} with $0 < \pi_{th} < 1$ and a set of regularization parameters Λ, the set of stable variables is defined as:

$$\hat{S}^{\text{stable}} = \{k : S_j(f) \geq \pi_{th}\} = \left\{ k : \max_{\lambda \in \Lambda} \left(\hat{\pi}_j^\lambda \right) \geq \pi_{th} \right\}. \tag{13}$$

The embedded multi-task approach ensures that the selected features have the following properties:1) Stability. A cortical region of the brain that is closely related to the subject's

TABLE I
SCREENING SUBJECTS

Time Span	Scanning Subjects	MMSE	Baseline Subjects
Baseline to M06	700	429	408
Baseline to M12	670	429	402
Baseline to M24	533	429	373
Baseline to M36	337	429	327

disease progression. 2) Global significance. MTL makes sure that the selected features are important for each task. One technique that arises here is to pick the coefficient value for one of the tasks when doing statistics on the stability of the selected features at equation 9.

IV. EXPERIMENTAL SETTINGS

A. Subjects

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early Alzheimer's disease (AD).

To track the effectiveness of disease progression models, ADNI-1 dataset is explored that contains longitudinal multi-center study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. As shown in the Table I. Subjects are between 55–90 years of age, the male accounts for 52.18%, the degree of suffering from the dementia, the data ratio of AD, MCI and NL are 25%, 50% and 25% respectively.

To explore the impact of the correlation between ROIs on progression with AD, MRI data from two follow-up points in the longitudinal cohort were extracted to facilitate observation of this spatiotemporal variation. At the same time, the cognitive scores (like MMSE or ADAS-cog) of longitudinal cohorts are employed to estimate the patients' cognitive functional decline during the AD progression. During the screening period, all the subject must satisfy the data integrity for verifying the reliable result. In other words, the cohort subjects must complete participation in two follow-up point MRI scans and multiple cognitive scoring assessments. In addition, the enrolled participates went through a rigorous selection process.

In this paper, only ADNI-1 subjects with all corresponding MRI and cognitive scores are evaluated.

B. Data pre-processing

For guarantees high image quality and reliable data handling, the MR images used in the paper were derived from standardized datasets, which provide the intensity normalized and gradient unwrapped TI image volumes. Subsequently, the FreeSurfer [29] was performed to feature extraction of the MR, which execute cortical reconstruction and volumetric segmentations for processing and analysing brain MR images.

For each MRI, cortical regions and subcortical regions are generated after this pre-processing suite. For each cortical

region, the cortical thickness average, standard deviation of thickness, surface area, and cortical volume were calculated as features. For each subcortical region, subcortical volume was calculated as feature. Data cleaning operations are performed as the following steps: 1) removal of individuals who failed cortical reconstruction and failed quality control; 2) removal of features with more than half of the sample missing values; 3) individual subject whose removal of baseline did not screen for MRI; 4) using the average of the features to fill in missing data; and 5) removal of cognitive function tests in individuals with missing follow-up points in longitudinal studies.

C. Feature selection

To discover the impact of the correlation between ROIs on progression with AD, we couple all the regions in pairs, which allows 326 block statistic indicators to combine 52975 features. For a given sample size, the higher the dimensionality, the sparser the distribution of the sample in space. As the dimension increase, the exponential number of samples are required to satisfy model estimate parameters effectively. In addition, the probability of the sample distribution being near the centre becomes lower and lower as the dimensionality increases. If the high-dimensional data is directly applied to the algorithm, it will cause the model to overfit. In order to solve this challenge, some basic methods and one advanced feature selection are employed to extract the representative feature from large number of original feature spaces.

1) Variance Threshold: It removes all features whose variance doesn't meet some threshold. If there are many features, but not every feature can well reflect the degree of discrimination, then such features will not have the value of analysis.
2) Univariate feature selection: Univariate feature selection works by selecting the best features based on univariate statistical tests. It removes all but a user-specified highest scoring percentage of features.
3) Stability Selection: Stability selection is based on sub-sampling in combination with (high dimensional) selection algorithms. The stability ranking score gives the probability that it is naturally interpretable.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. ROIs synchronization represents the progress of AD

We first accomplish three relevance approaches of estimating ROIs relevant criteria: Euclidean Distance (ED), Mahalanobis Distance (MD) and Cosine Similarity (CS). And then, each criterion between vectors composed of the magnitude and the non-absolute value of velocity of the biomarker are used and the feature subset selecting the original feature space to evaluate the subjects cognitive scores. Table II shows the different criterion tracking the AD progression. Note that Table II shows only the averaged results and variance of 30 independent experiments; and the temporal distance from baseline to M06 period. Besides, we also reproduced the model achieved by [15], [25], [28], with only MRI data as features.

TABLE II
DIFFERENT SIMILARITY MEASURE

	Original ROI	Mahalanobis Distance	Euclidean Distance	Cosine Similarity
Target: MMSE				
nMSE	0.827±0.065	0.973±0.076	0.944±0.080	**0.743±0.060**
wR	0.461±0.053	0.273±0.080	0.552±0.043	**0.552±0.043**
BL rMSE	1.750±0.157	1.782±1.509	1.901±0.176	**1.436±0.134**
M06 rMSE	2.326±0.302	2.420±0.195	2.240±0.286	**2.190±0.215**
M12 rMSE	2.599±0.366	2.943±0.291	**2.505±0.415**	2.541±0.415
M24 rMSE	3.516±0.777	3.747±0.627	3.689±0.694	**3.227±0.575**
M36 rMSE	4.169±0.831	4.394±0.803	5.020±0.866	**4.125±0.846**
Target: ADAS-cog				
nMSE	0.693±0.054	0.773±0.087	0.790±0.067	**0.666±0.058**
wR	0.579±0.041	0.514±0.057	0.488±0.055	**0.604±0.043**
BL rMSE	4.093±0.388	3.809±0.371	4.238±0.470	**3.670±0.606**
M06 rMSE	4.540±0.609	4.375±0.497	4.665±0.529	**4.399±0.740**
M12 rMSE	4.932±0.781	4.759±0.627	4969±0.590	**4.693±0.562**
M24 rMSE	**5.466±0.774**	6.234±1.104	6.537±1.023	5.706±0.899
M36 rMSE	**7.661±1.092**	8.943±1.969	8.851±1.352	8.133±1.720

TABLE III
ROIS SYNCHRONIZATION REPRESENTS THE PROGRESS OF AD

	Original ROI	BL to M06	BL to M12	BL to M24	BL to M36
Target: MMSE					
nMSE	0.827±0.065	0.743±0.060	0.726±0.092	**0.693±0.070**	0.724±0.121
wR	0.461±0.053	0.552±0.043	0.581±0.047	**0.595±0.044**	0.582±0.065
BL rMSE	1.750±0.157	1.436±0.134	1.472±0.152	1.408±0.177	**1.335±0.152**
M06 rMSE	2.326±0.302	2.190±0.215	2.265±0.268	2.134±0.194	**1.983±0.332**
M12 rMSE	2.599±0.366	2.541±0.415	2.440±0.335	2.559±0.481	**2.053±0.306**
M24 rMSE	3.516±0.777	3.227±0.575	3.197±0.549	3.244±0.644	**2.710±0.517**
M36 rMSE	4.169±0.831	4.125±0.846	4.157±0.704	3.847±0.829	**3.345±0.701**
Target: ADAS-cog					
nMSE	0.693±0.054	0.666±0.058	0.691±0.087	**0.653±0.075**	0.881±0.056
wR	0.579±0.041	0.604±0.043	0.592±0.051	**0.626±0.049**	0.387±0.053
BL rMSE	4.093±0.388	3.670±0.606	3.522±0.329	**3.648±0.479**	4.084±0.414
M06 rMSE	4.540±0.609	4.399±0.740	4.406±0.514	**4.206±0.530**	4.462±0.609
M12 rMSE	4.932±0.781	4.693±0.562	4.847±0.681	**4.872±0.702**	5.067±0.630
M24 rMSE	**5.466±0.774**	5.706±0.899	5.953±0.929	5.707±1.124	5.530±0.625
M36 rMSE	**7.661±1.092**	8.133±1.720	8.100±1.349	8.255±1.727	7.761±1.518

Overall the cosine similarity representation of our proposed ROIs synchronization approaches outperforms the original ROIs feature. We have the following observations: 1) The collaborative expression of ROIs is better than independent ROI to a certain extent. 2) The expression of Cosine Similarity performs better than that of Cosine Similarity and Mahalanobis Distance. 3) The proposed Cosine Similarity representation witnesses significant improvement for the early time point. This may be due to the data spanning from baseline and M06 period.

B. Temporal Span of MRI Scan

Inspired by the above experiments, we further explored the influence of temporal span on the progress of positioning AD under the collaborative expression of ROIs. In this section, only cosine similarity was utilized to estimate the cognitive functional progression.

There are four temporal span group performed, namely baseline to M06 period, baseline to M12 period, baseline to M24 period and baseline to M36 period. Table III shows that the normalized results of different visited time span and the root mean square error of each sub-task results. We follow the same experimental procedure as above. The experimental results are presented in Table III.

We can observe from the table that as the time span increases, the overall generalization performance of the model improves. When the temporal span growths, we also have the
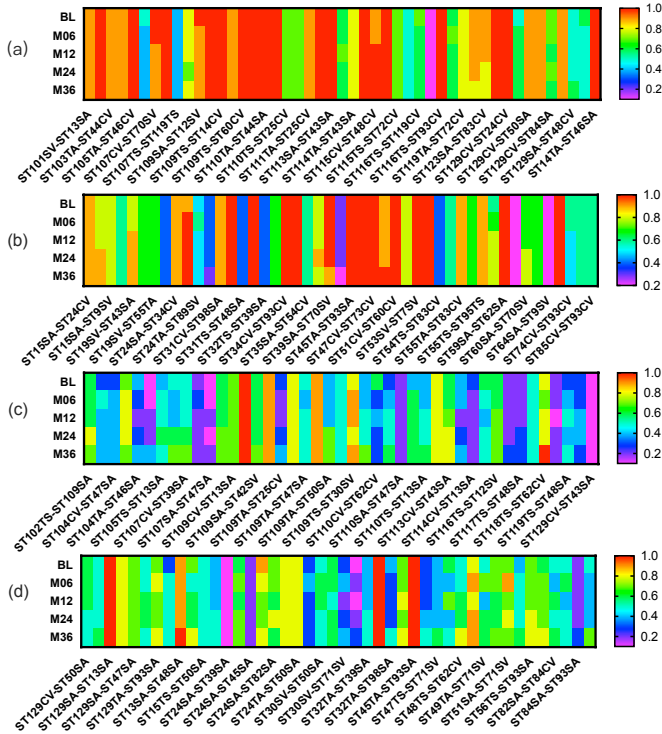
Fig. 2. the vectors of stability temporal collaborative patterns. A total of 94 and 87 stable co-expression pairs respectively. Specifically, (a) and (b) belong to the MMSE-targeted model of AD progress; (c) and (d) belong to ADAS-cog-targeted model of AD progress.

following observations: 1) The performance of the subtasks will gradually improve. 2) The task of the latter point in time has been greatly enhanced. This may be due to the latter MRI scanning support more collaborative expression of ROIs and these results further validate the efficacy of the proposed method for temporal-spatial collaborative expression of ROIs. 3) during the BL to M24, the overall task performance outperforms others. 4) during the BL to M36, Although the performance of the global model has decreased, the performance of each subtask has been greatly improved. On the one hand, it proves once again that the collaborative expression of ROIs can have an important effect. On the other hand, the performance of overall model means that the fundamental premise of multi-task model was weakened: there is no longer a simple linear relationship between ROIs and cognitive function. Yet this problem also appears in the model of Zhou [49], with the help of collaborative expression of ROIs, this defect can be properly compensated, which verify the effect of ROIs synchronization represents the progress of AD.

C. Stability temporal collaborative patterns of MRI biomarkers

Firstly, we use the data from a set of experiments with the best performance in experiment: Temporal Span of MRI Scan, namely the temporal span for baseline to M24 periods, which contains 94 dimensions corresponding a crucial couples of ROIs pairs. Secondly, a set of environmental parameters are clearly indicated: 1) Only half of the overall sample in each sampling subset is randomly selected. 2) A total of 210 combinations of model hyperparameters. 3) during every combination, 10 samplings were executed. Finally, the vectors of stability temporal collaborative patterns are showed in Fig.2. For the MMSE set, the result shows that the synergistic effect of left insula on left entorhinal cortex, left posterior cingulate cortex, right bankssts, left caudal anterior cingulate cortex, left pars triangularis. The synergistic effect of right posterior cingulate gyrus on right isthmus of cingulate cortex, left temporal pole. For the ADAS-cog set, the result shows that the synergistic effect of left insula on left entorhinal cortex, left posterior cingulate cortex, left bankssts, left pars triangularis. The synergistic effect of left entorhinal on left parahippocampal, right cuneus, medial orbitofrontal cortex. The synergistic effect of right posterior cingulate cortex on left pars triangularis, left parahippocampal. The fact that our findings are in line with those of previous studies [30]–[32] demonstrates the validity of our proposed model.

VI. DISCUSSION

In the selection of longitudinal stability, we observed 29 most stable features with MMSE score, which are shown in Fig.2. Among them, the correlation features based on Cortical Volume and Cortical Volume are the majority (6 features), which shows that the similarity of the change trend of the biomarkers based on Cortical Volumes have important effect in AD prediction. Previous studies have also observed a significant improvement in the classification performance of abnormal cortical patterns and the coordinated patterns of cortical morphology are widely altered in AD patients [10]. In addition, the number of correlation features based on the similarity of changes between Surface Area and Surface Area is also relatively large (5 features).

VII. CONCLUSION

The correlation between biomarkers may improve the accuracy of AD progression modeling to help doctors and patients in early intervention of AD to treat patients and improve the quality of life. Therefore, we propose a new method to extract morphological information from MRI, combined with MTL to effectively model and predict AD progress. This paper has three main contributions. First, we use cosine similarity to represent the temporal and spatial relationships between brain biomarkers. Second, we express the disease progression prediction as a MTL problem and combine the cosine similarity to predict the disease progression of AD. Third, we use multitask-based stability selection to analyze the temporal and spatial dynamic patterns between biomarkers. We prove that correlate information can better describe the brain structural changes in patients with NL, MCI and AD. Combining MTL and the correlation between biomarkers can improve the predictive performance of AD disease progression.

REFERENCES

[1] A. Association, "2019 alzheimer's disease facts and figures," *Alzheimer's & dementia*, vol. 15, no. 3, pp. 321–387, 2019.

[2] S. Tabarestani, M. Aghili, M. Eslami, M. Cabrerizo, A. Barreto, N. Rishe, R. E. Curiel, D. Loewenstein, R. Duara, and M. Adjouadi, "A distributed multitask multimodal approach for the prediction of alzheimer's disease in a longitudinal study," *NeuroImage*, vol. 206, p. 116317, 2020.

[3] R. S. Doody, V. Pavlik, P. Massman, S. Rountree, E. Darby, and W. Chan, "Predicting progression of alzheimer's disease," *Alzheimer's research & therapy*, vol. 2, pp. 1–9, 2010.

[4] C. Green, J. Shearer, C. W. Ritchie, and J. P. Zajicek, "Model-based economic evaluation in alzheimer's disease: a review of the methods available to model alzheimer's disease progression," *Value in health*, vol. 14, no. 5, pp. 621–630, 2011.

[5] M. Nguyen, N. Sun, D. C. Alexander, J. Feng, and B. T. Yeo, "Modeling alzheimer's disease progression using deep recurrent neural networks," in *2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*. IEEE, 2018, pp. 1–4.

[6] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of alzheimer's disease with deep learning," in *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*. IEEE, 2014, pp. 1015–1018.

[7] P. Yang, G. Yang, J. Liu, J. Qi, Y. Yang, X. Wang, and T. Wang, "Duapm: An effective dynamic micro-blogging user activity prediction model towards cyber-physical-social systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5317–5326, 2019.

[8] R. Sivera, H. Delingette, M. Lorenzi, X. Pennec, N. Ayache, A. D. N. Initiative *et al.*, "A model of brain morphological changes related to aging and alzheimer's disease from cross-sectional assessments," *NeuroImage*, vol. 198, pp. 255–270, 2019.

[9] P. Vemuri, H. Wiste, S. Weigand, L. Shaw, J. Trojanowski, M. Weiner, D. S. Knopman, R. C. Petersen, C. Jack *et al.*, "Mri and csf biomarkers in normal, mci, and ad subjects: predicting future clinical change," *Neurology*, vol. 73, no. 4, pp. 294–301, 2009.

[10] C.-Y. Wee, P.-T. Yap, D. Shen, and A. D. N. Initiative, "Prediction of alzheimer's disease and mild cognitive impairment using cortical morphological patterns," *Human brain mapping*, vol. 34, no. 12, pp. 3411–3425, 2013.

[11] J. Zhou, J. Chen, and J. Ye, "Malsar: Multi-task learning via structural regularization," *Arizona State University*, vol. 21, pp. 1–50, 2011.

[12] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[13] M. M. Ghazi, M. Nielsen, A. Pai, M. J. Cardoso, M. Modat, S. Ourselin, L. Sørensen, A. D. N. Initiative *et al.*, "Training recurrent neural networks robust to incomplete data: Application to alzheimer's disease progression modeling," *Medical image analysis*, vol. 53, pp. 39–46, 2019.

[14] K.-H. Thung and C.-Y. Wee, "A brief review on multi-task learning," *Multimedia Tools and Applications*, vol. 77, pp. 29 705–29 725, 2018.

[15] J. Zhou, L. Yuan, J. Liu, and J. Ye, "A multi-task learning formulation for predicting disease progression," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, pp. 814–822.

[16] L. Kurlowicz and M. Wallace, "The mini-mental state examination (mmse)," pp. 8–9, 1999.

[17] L. Chu, K. Chiu, S. Hui, G. Yu, W. Tsui, and P. Lee, "The reliability and validity of the alzheimer's disease assessment scale cognitive subscale (adas-cog) among the elderly chinese in hong kong." *Annals of the Academy of Medicine, Singapore*, vol. 29, no. 4, pp. 474–485, 2000.

[18] J. Qi, P. Yang, L. Newcombe, X. Peng, Y. Yang, and Z. Zhao, "An overview of data fusion techniques for internet of things enabled physical activity recognition and measure," *Information Fusion*, vol. 55, pp. 269–280, 2020.

[19] M. Schmidt *et al.*, *Rey auditory verbal learning test: A handbook*. Western Psychological Services Los Angeles, CA, 1996, vol. 17.

[20] E. Vakil and H. Blachstein, "Rey auditory-verbal learning test: structure analysis," *Journal of clinical psychology*, vol. 49, no. 6, pp. 883–890, 1993.

[21] J. Wan, Z. Zhang, B. D. Rao, S. Fang, J. Yan, A. J. Saykin, and L. Shen, "Identifying the neuroanatomical basis of cognitive impairment in alzheimer's disease by correlation-and nonlinearity-aware sparse bayesian learning," *IEEE transactions on medical imaging*, vol. 33, no. 7, pp. 1475–1487, 2014.

[22] P. Cao, X. Shan, D. Zhao, M. Huang, and O. Zaiane, "Sparse shared structure based multi-task learning for mri based cognitive performance prediction of alzheimer's disease," *Pattern Recognition*, vol. 72, pp. 219–235, 2017.

[23] M. Wang, D. Zhang, D. Shen, and M. Liu, "Multi-task exclusive relationship learning for alzheimer's disease progression prediction with longitudinal data," *Medical image analysis*, vol. 53, pp. 111–122, 2019.

[24] J. Peng, X. Zhu, Y. Wang, L. An, and D. Shen, "Structured sparsity regularized multiple kernel learning for alzheimer's disease diagnosis," *Pattern recognition*, vol. 88, pp. 370–382, 2019.

[25] J. Zhou, J. Liu, V. A. Narayan, J. Ye, A. D. N. Initiative *et al.*, "Modeling disease progression via multi-task learning," *NeuroImage*, vol. 78, pp. 233–248, 2013.

[26] K. Ito, B. Corrigan, Q. Zhao, J. French, R. Miller, H. Soares, E. Katz, T. Nicholas, B. Billing, R. Anziano *et al.*, "Disease progression model for cognitive deterioration from alzheimer's disease neuroimaging initiative database," *Alzheimer's & Dementia*, vol. 7, no. 2, pp. 151–160, 2011.

[27] C. M. Stonnington, C. Chu, S. Klöppel, C. R. Jack Jr, J. Ashburner, R. S. Frackowiak, A. D. N. Initiative *et al.*, "Predicting clinical scores from magnetic resonance scans in alzheimer's disease," *Neuroimage*, vol. 51, no. 4, pp. 1405–1413, 2010.

[28] J. Zhou, J. Liu, V. A. Narayan, and J. Ye, "Modeling disease progression via fused sparse group lasso," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 1095–1103.

[29] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[30] Y. Liu, T. Paajanen, Y. Zhang, E. Westman, L.-O. Wahlund, A. Simmons, C. Tunnard, T. Sobow, P. Mecocci, M. Tsolaki *et al.*, "Combination analysis of neuropsychological tests and structural mri measures in differentiating ad, mci and control groups—the addneuromed study," *Neurobiology of Aging*, vol. 32, no. 7, pp. 1198–1206, 2011.

[31] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski, "Prediction of mci to ad conversion, via mri, csf biomarkers, and pattern classification," *Neurobiology of aging*, vol. 32, no. 12, pp. 2322–e19, 2011.

[32] L. Wang, F. C. Goldstein, E. Veledar, A. I. Levey, J. J. Lah, C. C. Meltzer, C. A. Holder, and H. Mao, "Alterations in cortical thickness and white matter integrity in mild cognitive impairment measured by whole-brain cortical thickness mapping and diffusion tensor imaging," *American Journal of Neuroradiology*, vol. 30, no. 5, pp. 893–899, 2009.