Proceedings Paper:
Zhang, Y., Lanfranchi, V., Wang, X. et al. (2 more authors) (2023) Modeling Alzheimer's disease progression via amalgamated magnitude-direction brain structure variation quantification and tensor multi-task learning. In: 2022 IEEE International Conference on Bioinformatics and Biomedicine. 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2022), 06-08 Dec 2022, Las Vegas, NV, USA. Institute of Electrical and Electronics Engineers (IEEE) , pp. 2735-2742. ISBN 9781665468206

https://doi.org/10.1109/BIBM55620.2022.9995468

Modeling Alzheimer's Disease Progression via Amalgamated Magnitude-Direction Brain Structure Variation Quantification and Tensor Multi-task Learning

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

Vitaveska Lanfranchi
Department of Computer Science
University of Sheffield
Sheffield, UK
v.lanfranchi@sheffield.ac.uk

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

Menghui Zhou
Department of Software
Yunnan University
Kunming, China
menghuizhoucn@gmail.com

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

***Abstract*—Machine learning (ML) techniques for predicting the progression of Alzheimer's disease (AD) can greatly assist researchers and clinicians in establishing effective AD prevention and treatment strategies. The problems of monotonicity of data forms and scarcity of medical data are the main reasons that currently limit the performance of ML approaches. In this research, we propose a novel similarity-based quantification approach that simultaneously considers the magnitude and direction relationships of structural variations among brain biomarkers, and encodes quantified data as third-order tensors to solve problem of data form monotonicity, then combining tensor multi-tasking learning model to predict AD progression. In this model, the prediction of each patient is considered as a task, and each task shares a set of latent factors obtained by tensor decomposition, knowledge sharing between tasks can improve the generalization of the model and solve the problem of scarcity of medical data. The model can be utilised to efficiently predict the progression of AD integrating magnetic resonance imaging (MRI) data and cognitive scores of AD patients at different stages. To evaluate the effectiveness of the proposed approach, we conducted extensive experiments utilising MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The results reveal that the proposed model predicts AD progression more accurately and consistently than single-task and state-of-the-art multi-task regression approaches on various cognitive scores. The proposed approach can recognize brain structural variation in patients and apply it to reliably predict and diagnose AD progression.**

Keywords—Alzheimer's disease progression, amalgamated magnitude-direction quantification, brain structure variation, tensor multi-task learning

I. INTRODUCTION

In Alzheimer's disease, neurons and their connections deteriorate, leading to cognitive impairment and memory loss [1]. There is no cure for AD, which can be physiologically, psychologically and emotionally traumatic for patients and their families. For a better understanding and early diagnosis of AD, it is essential to comprehend the AD progression and identify pathological biomarkers.

Previous studies have concentrated on utilising biomarker data along with machine learning algorithms to calculate patients' cognitive scores as target for determining the severity of cognitive impairment in patients. Existing models for predicting AD progression include machine learning regression algorithms [2], deep learning approaches based on neural networks [3] and survival models based on statistical probabilities [4]. The above models face two main problems. The first is the problem of small dataset, it is difficult to obtain data on neurological diseases such as AD. The accuracy of traditional machine learning algorithms is limited, and it is difficult to train high-precision deep learning models with small dataset. The second problem is the monotonicity of the data form and the resulting loss of hidden information. The input features of the above models are represented as second-order matrices containing patient and biomarker information, which makes it difficult to predict and analyse disease progression from multiple dimensions (e.g., spatial and temporal dimensions). At the same time, since the second-order matrix generally focuses on a single biomarker, the correlation information between different AD biomarkers will be lost.

For the problem of small dataset, multi-task learning (MTL) can share knowledge and information across tasks, outperforms standard single-task learning approaches in terms of model accuracy, generalizability and interpretability, and is most

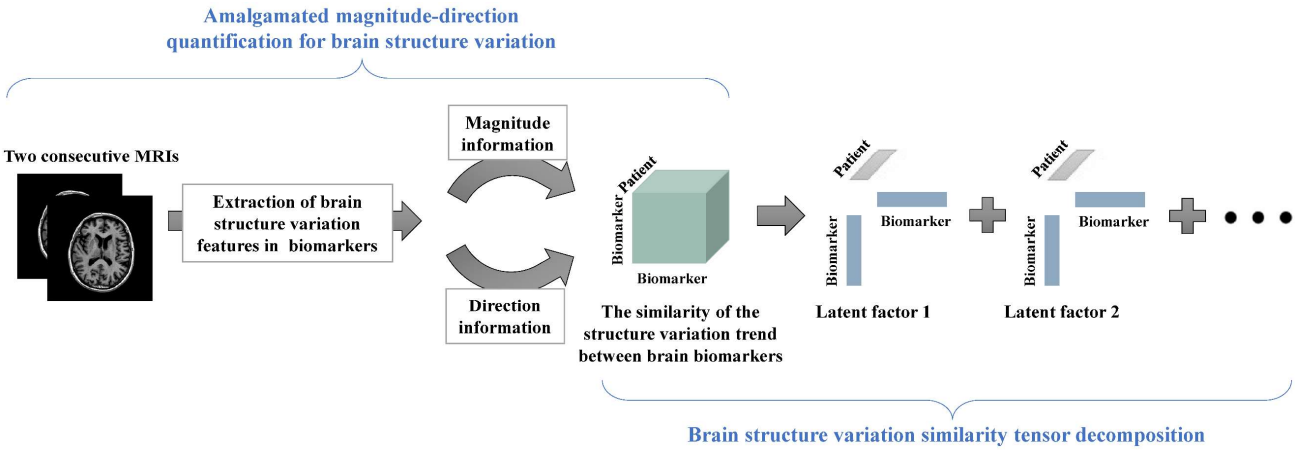

Fig. 1. CP decomposition on a similarity tensor representation based on the similarity of the structure variation trend between brain biomarkers.

efficient when sample numbers are small [5]. To address the problem of monotonicity of the data form and the resulting loss of hidden information, this paper proposes to construct a third-order tensor with three components per index to build an AD prediction model to better represent AD data in both spatial and temporal dimensions. The utilisation of tensor in regression algorithms can enhance prediction accuracy, stability, and interpretability by better representing AD biomarker features.

In this research, we propose an amalgamated magnitude-direction brain structure variation quantification with tensor-based MTL approach to predict AD progression by incorporating spatio-temporal information on relative structural variations between different brain biomarkers. Specifically, we propose a novel similarity computation-based quantification method to simultaneously assess the magnitude and direction of brain structural variations, which describes the similarity of morphological variation trends among different biomarkers as a third-order tensor and its dimensions correspond to the first biomarker, second biomarker and patient sample. The proposed approach then decomposes the tensor utilising a CANDECOMP/PARAFAC (CP) decomposition [6] and extracts a collection of rank one latent factors from the original data (Fig. 1). Each latent factor is represented by its first biomarker, second biomarker and patient sample dimensions to explain the latent factors that affect the variability of the data in an interpretable way that can serve as a predictor for training the MTL model. And these latent factors are shared by the predictions of each patient sample, which is the task in this study.

The main contributions of this research can be summarized as follows:

- We propose a novel similarity-based quantification approach that simultaneously considers the magnitude and direction correlations of structural variations between brain biomarkers, it contains comprehensive information on brain structural variations and can effectively distinguish between CN (cognitively normal elderly), MCI (mild cognitive impairment) and AD patients, and encoded MRI biomarker data as a third-order tensor to solve the problem of the data form monotonicity.
- We combine the above tensor with a multi-task learning algorithm for AD progression prediction, which uses tensor decomposition to learn task relevance from the original data. This allows all samples to share latent knowledge of biomarkers based on brain structural variations, and significantly improves the accuracy and stability of AD progression prediction.
- We identified and analysed important relative structural variation relationships between brain biomarkers in the prediction of AD progression that could be utilised as potential indicators for AD early identification.

The rest of this paper proceeds as follows: Section II introduces the related works. Section III presents the proposed approach. Section IV includes details on the ADNI datasets, pre-processing, and experimental procedures. Section V presents the experimental results for the ADNI datasets, along with a discussion. Section VI concludes this research.

II. Related Work

Abundant research in the field of brain science have focused on the differences in brain structure variations between CN, MCI and AD. [7] use imaging data from large human populations to find patterns of brain anatomy and function related with AD, CN, schizophrenia and abnormal brain development. [8] investigated the link between cerebrospinal fluid (CSF) and MRI biomarkers, clinical diagnosis, and cognitive function in individuals with CN, AD and aMCI (amnestic mild cognitive impairment). MRI was found to provide superior cross-sectional integrated cognitive and functional abilities, as well as enhanced cross-sectional grouping and discriminating. [9] used automated MRI analysis to assess cortical thickness in healthy older people, MCI patients, and AD patients. As the disease progressed from MCI to AD, patterns of cortical thinning were noticed, and it was discovered that the whole cortex thinned and extended considerably into the lateral temporal cortex. Based on the study of brain structural variation, the correlation between AD MRI biomarkers is also the focus of brain research. [10] enhanced the classification performance of AD and its precursor stages by merging relevant information with ROI-based data and correlating regional mean cortical thickness with multi-kernel support vector machines. [11] examined brain networks using graph theory by thresholding the cortical thickness correlation matrix for different areas. According to the study, there are differences in

the brain biomarkers for CN, MCI, and AD. It also studied and analysed the relationship between the course of Alzheimer's disease and biomarkers. The above studies demonstrated the importance and validity of structural brain variation and brain biomarker correlations for AD research, but they focus on a single biomarker or a single category of biomarkers, ignoring the correlation of structural variations in different categories of brain biomarkers, which is essential for characterizing AD.

To quantify the correlation of structural variation between different types of brain biomarkers, we represented the trend of morphological variation of MRI biomarkers as a vector and utilised a similarity-based calculation to calculate the similarity of structural variation between brain biomarkers. The similarity between vectors can be calculated in a number of approaches. The Euclidean distance is the simplest method to calculate the distance and is derived from the formula for the distance between two points in Euclidean space. The Mahalanobis distance is a reliable approach for determining the similarity of two unknown sample sets. It is not impacted by scale unlike the Euclidean distance, and the Mahalanobis distance between two points has no relation with the measuring unit of the original data; nonetheless, its drawback is that it exaggerates the effect of minor factors [12]. The cosine similarity measures the difference between two subjects by using the cosine value of the angle between two vectors in a vector space [13]. Cosine similarity, as opposed to Euclidean distance, focuses on the difference in direction of two vectors, whereas Euclidean distance quantifies the difference in values. The above are several standard, commonly used and effective similarity computation methods, but they all share the same problem of not being able to assess both the magnitude and the direction of the vector, whereas structural variation in the brain is a complex process that contains both magnitude and direction information, therefore, existing similarity computation methods cannot quantify all its variation information. For this reason, we propose a novel similarity quantification method that can simultaneously assess the magnitude and direction of the brain structural variation.

Multi-task learning aims to learn numerous related tasks together to ensure that the knowledge contained in one task may be utilised by other tasks, therefore boosting the generalisation performance of all tasks [5]. MTL methodology is extensively employed in the field of biomedical engineering; for our research case AD, multi-task learning offers a wide range of applications in numerous domains. For feature learning approach, existing approaches have focused on modelling task interactions through the use of novel regularisation methods [14][15][16]. And kernel approaches were added to the methodology to accommodate non-linear relationships [17][18]. For feature selection approach, [19] presented a multi multi-task learning strategy that identifies a common subset of multiple variable-related features from each modality and predicts multiple variables from multi-modal input concurrently. [20] presented a deep belief network-based multi-task learning AD classification algorithm that introduces a multi-task feature selection technique, evaluates the internal connection between several related tasks, and identifies feature sets relevant to all tasks. For low-rank approach, [21] presented a robust multi-task learning system that uses a low-rank structure to represent task links while identifying anomalous tasks using a group sparse structure. Different from the setting of the algorithm target as the task in the above methods, we believe that knowledge sharing across prediction tasks of various patients can enhance achievable performance, therefore, we establish the prediction task of a single patient as one task, which is a small-scale task setup approach.

III. Methodology

A. Denotation

For brevity, we represent tensors as italic capital letters, such as X or Y, and matrices by capital letters, such as A or B. Vectors are denoted by lowercase letters such as x whereas Scalars are denoted by italic lowercase letters such as a.

B. Amalgamated magnitude-direction quantification for brain structure variation

Correlation of structural variation between brain biomarkers was calculated utilising two consecutive MRI examinations. To calculate the rate of change and velocity of brain biomarkers, we utilised baseline BL (the date the patient was first tested in hospital) and M06 (the six-month time point after the first visit) MRI, where x is the value of the rate of interest brain biomarker and t is the date of the MRI examination. The rate of change is $\frac{x_{M06}-x_{BL}}{x_{BL}}$, the velocity is $\frac{x_{M06}-x_{BL}}{t_{M06}-t_{BL}}$ per month. The rate of change and velocity of each brain biomarker was then utilised to create a vector to describe its trend of structural variation.

We then propose a two-stage quantitative approach that simultaneously assesses the magnitude and direction of structural variation among brain biomarkers. The Mahalanobis distance was first utilised to calculate the similarity of the absolute values of the two vectors to reflect the similarity in the magnitude of the structural variation of the two MRI biomarkers. Mahalanobis distance is utilised because it is scale independent when dividing by the covariance matrix. The Mahalanobis distance between the absolute values of vectors x_i and x_j is stated as: $\mathrm{Ma}(|\mathrm{x}_i|,|\mathrm{x}_j|)=\sqrt{(|\mathrm{x}_i|-|\mathrm{x}_j|)^{\mathrm{T}}\mathrm{S}^{-1}(|\mathrm{x}_i|-|\mathrm{x}_j|)}$, where S is covariance matrix. The quantified value of Mahalanobis distance is between 1 and 0, where 1 is completely similar and 0 is completely dissimilar. Then we add the direction information to the values. We noticed that for the two brain biomarkers, there were only five cases of their structural variation direction relationship: 1) both grow, 2) both decline, 3) one grows and the other declines, 4) one changes and the other does not change, 5) both remain unchanged. We set cases 1) and 2) to be synchronous variation, case 3) to be asynchronous variation, and cases 4) and 5) to be completely irrelevant. A mapping function (1) is then utilised to map the values previously calculated utilising the Mahalanobis distance to values between 1 and -1 to add directional information. Where 1 means completely relevant in the case of synchronous variation, 0 means completely irrelevant and -1 means completely relevant in the case of asynchronous variation.

$$\begin{cases} x = x, \text{if two biomarkers varied synchronously} \\ x = -x, \text{if two biomarkers varied asynchronously} \\ x = 0, \text{if two biomarkers are not relevant} \end{cases} \quad (1)$$

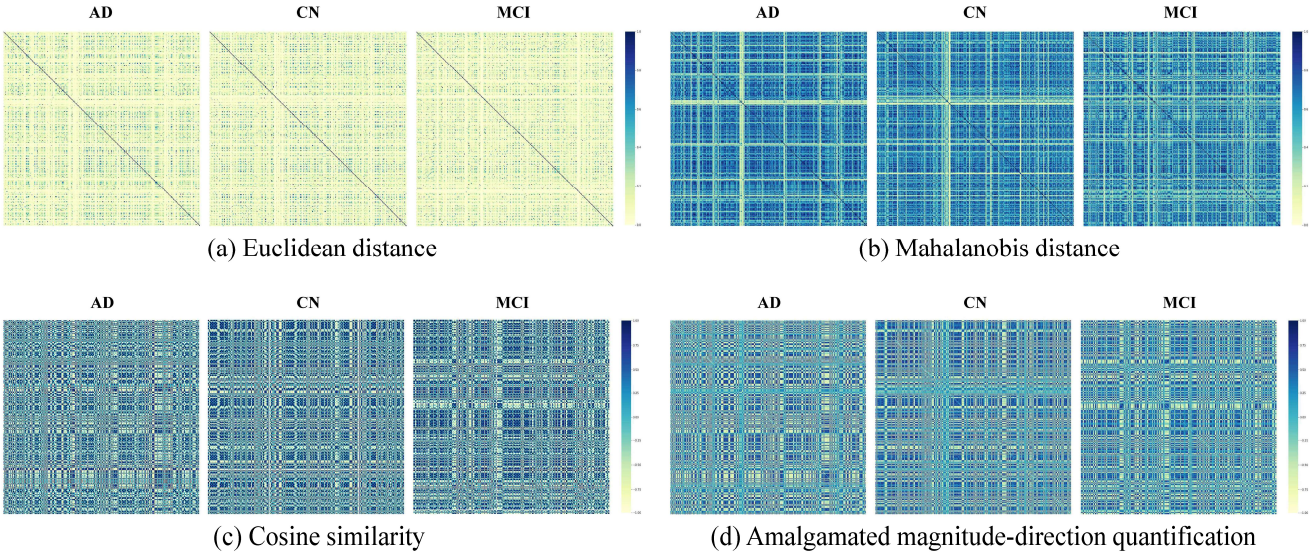

Fig. 2. Examples of (a) Euclidean distance, (b) Mahalanobis distance, (c) Cosine similarity and (d) Amalgamated magnitude-direction quantification matrix distribution for AD, CN and MCI brain structure variation quantification. (The scale for (a) Euclidean distance and (b) Mahalanobis distance from top to bottom is 1.0, 0.8, 0.6, 0.4, 0.2, 0.0. The scale for (c) Cosine similarity and (d) Amalgamated magnitude-direction quantification from top to bottom is 1.00, 0.75, 0.50, 0.25, 0.00, -0.25, -0.50, -0.75, -1.00).

Fig. 2 shows the structural variation correlations of brain biomarkers for AD, CN and MCI expressed by Euclidean distance, Mahalanobis distance, Cosine similarity and our proposed Amalgamated magnitude-direction quantification. We observed that our proposed quantification approach showed the greatest difference in matrix distribution across disease stages compared to Euclidean and Mahalanobis distances, with the data for Euclidean distances being too sparse and the data for Mahalanobis distances being more uniform. The distribution of cosine similarity data across disease stages is similar to our proposed quantification approach, but the data has more maxima and minima because it only contains information on the direction of structural variation in brain biomarkers, whereas our approach contains both magnitude and direction information, resulting in a similar matrix distribution to cosine similarity, with a smooth data distribution and diverse data characteristics. It allows the AD progression prediction process to include more comprehensive information on brain structural variations, while enhancing the interpretability of brain biomarker correlations in AD progression at the results analysis process.

C. Tensor multi-task learning regression

The above approach quantifies each individual's BL and M06 MRI data into a structural variation similarity matrix containing information on both magnitude and direction, and then integrates the similarity matrices of all individuals into a similarity tensor. This is combined with our proposed tensor multi-task learning regression algorithm to perform AD progression prediction. Preliminary versions of the regression algorithm have been reported [22][23], and this research extends it and combines it with the newly proposed quantification method to include more comprehensive information on brain structural variation in the AD progression prediction process.

To forecast future cognitive scores (e.g., MMSE and ADAS-Cog). Consider the following tensor multi-task regression problem for n training samples with d_1 and d_2 features in t time points. Let $X \in \mathbb{R}^{d_1 \times d_2 \times n}$ be the input tensor from two successive MRI detections and it is the combination of amalgamated magnitude-direction quantified matrix for all n samples $\mathrm{X}_n \in \mathbb{R}^{d_1 \times d_2}$, $\mathrm{Y} = [\mathrm{y}_1, \cdots, \mathrm{y}_t] \in \mathbb{R}^{n \times t}$ be the targets and $\mathrm{y}_t = [y_1, \cdots, y_n] \in \mathbb{R}^n$ is the corresponding target (cognitive scores) at various future time points. We utilise the operator $\odot$ as follows: $\mathrm{Z} = \mathrm{M} \odot \mathrm{N}$ denotes $z_{ij} = m_{ij} n_{ij}$, for all i, j.

Because the association calculation between biomarkers is paired and half of the data is duplicated, the input tensor for the similarity of morphological variation trends in brain biomarkers is a symmetric tensor. Duplicate data leads to an increase in computational complexity and computational cost. To solve the problem of duplicate data, the study proposes the duplicate data correction matrix:

$$\mathrm{K} = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{d_1 \times d_2} \tag{2}$$

For t-th prediction time point, the proposed approach's objective function can be described as follows:

$$L_t(X, \mathrm{y}_t) = \min_{\mathrm{W}_t, \mathrm{A}_t, \mathrm{B}_t, \mathrm{C}_t} \frac{1}{2} \|\hat{\mathrm{y}}_t - \mathrm{y}_t\|_{\mathrm{F}}^2 + \frac{\lambda}{2} \|X - [\![\mathrm{A}_t, \mathrm{B}_t, \mathrm{C}_t]\!]\|_{\mathrm{F}}^2 + \beta \|\mathrm{W}_t, \mathrm{A}_t, \mathrm{B}_t, \mathrm{C}_t\|_1 \tag{3}$$

$$\hat{y}_n = \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \mathrm{U}_{ij}$$

$$\text{where } \mathrm{U} = (\mathrm{A}_t \mathrm{B}_t^{\mathrm{T}}) \odot \mathrm{K} \odot \mathrm{W}_t \odot \mathrm{X}_n \text{ , } \mathrm{U} \in \mathbb{R}^{d_1 \times d_2}$$

where the first term evaluates empirical error for the training data, $\hat{\mathrm{y}}_t = [\hat{y}_1, \cdots, \hat{y}_n] \in \mathbb{R}^n$ are the predicted values, $\mathrm{A}_t \in \mathbb{R}^{d_1 \times r}$ is the latent factor matrix for the first biomarker dimension and $\mathrm{B}_t \in \mathbb{R}^{d_2 \times r}$ is the latent factor matrix for the second biomarker dimension with r latent factors, $\mathrm{W}_t \in \mathbb{R}^{d_1 \times d_2}$ is the model parameter matrix for t-th prediction time point, λ

and β are the regularization parameters. Acquiring latent factors by optimising CP tensor decomposition objective function $\|X - [\![\mathrm{A}_t, \mathrm{B}_t, \mathrm{C}_t]\!]\|_{\mathrm{F}}^2$, where $X = [\![\mathrm{A}_t, \mathrm{B}_t, \mathrm{C}_t]\!] = \sum_{i=1}^{r} \mathrm{a}_i^t \circ \mathrm{b}_i^t \circ \mathrm{c}_i^t$ where $\circ$ represent the outer product operation between two vectors, while $\mathrm{a}_i^t, \mathrm{b}_i^t$ and c_i^t correspond to the vectors related with the *i*-th latent factor for *t*-th prediction time point. $\|\mathrm{W}_t, \mathrm{A}_t, \mathrm{B}_t, \mathrm{C}_t\|_1$ applying an $\ell 1$-norm on the, A_t, B_t, C_t and $\mathrm{W}_t$matrices respectively.

For all prediction time points, the objective function is as follows:

$$L(X, \mathrm{Y}) = \min_{\mathrm{W}_f} \sum_1^t L_t(X, \mathrm{y}_t) + \theta \|\mathrm{W}_f \mathrm{P}(\alpha)\|_{\mathrm{F}}^2 \qquad (4)$$

where $\|\mathrm{W}_f \mathrm{P}(\alpha)\|_{\mathrm{F}}^2$ is the generalized temporal smoothness term, the model parameter matrix $\mathrm{W}_f \in \mathbb{R}^{(d_1 \times d_2) \times t}$ is the temporal dimension unfolding for model parameter tensor $W \in \mathbb{R}^{d_1 \times d_2 \times t}$, θ is regularization parameter. The generalised temporal smoothing states that while diagnosing Alzheimer's disease, the expert examines not only the patient's present symptoms, but also their previous symptoms. As a result, we can utilise matrix multiplication to derive the more realistic temporal smoothness assumption:

$$\mathrm{WP}(\alpha) = \mathrm{WHD}_1(\alpha_1)\mathrm{D}_2(\alpha_2) \cdots \mathrm{D}_{t-2}(\alpha_{t-2}) \qquad (5)$$

where $\mathrm{H} \in \mathbb{R}^{t \times (t-1)}$ has the following description: $\mathrm{H}_{ij} = -1$ if $i = j + 1$, $\mathrm{H}_{ij} = 1$ if $i = j$ and $\mathrm{H}_{ij} = 0$ otherwise. $\mathrm{P}(\alpha)$ represents the association between progress, it contains the hyperparameters α, which denotes the interactional degree of the present progression and all previous progressions; the interactional degree criteria differ for each stage of disease progression because the influence of each stage on the stage after it is not always consistent, and it is dependent on the results of cross-validation. $\mathrm{D}_i(\alpha_i) \in \mathbb{R}^{(t-1)\times(t-1)}$ is an identity matrix and the value of $\mathrm{D}_{i_{m,n}}(\alpha_i)$ is replaced by $1 - \alpha_i$ if $m = n = i + 1$, the value of $\mathrm{D}_{i_{m,n}}(\alpha_i)$ is replaced by α_i if $m = i, n = i + 1$.

Latent factors $A \in \mathbb{R}^{d_1 \times r \times t}$, $B \in \mathbb{R}^{d_2 \times r \times t}$, $C \in \mathbb{R}^{d_3 \times r \times t}$ and model parameter $W \in \mathbb{R}^{d_1 \times d_2 \times t}$ can be learned by optimising the objective function iteratively for each group of variables to be solved. Because not all parts of the objective function are differentiable, we utilise proximal gradient descent to resolve each subproblem. The parts in our objective function relating Frobenius norms are differentiable, but the parts involving the sparsity $\ell 1$-norms are not differentiable. To construct the proximal problem for a non-smooth objective function, the proximal approach is extensively used [24][25][26], By substituting the smooth function for the quadratic function, we can get sum of the smooth and non-smooth functions. Its quadratic functions can be constructed in a number of ways applying Taylor series, and the proximate problems that result are typically simpler to solve than the original models. The strategy can hasten optimization convergence and simplify the construction of distributed optimization algorithms.

TABLE I. DEMOGRAPHIC CHARACTERISTIC OF THE STUDIED SUBJECTS VALUED ARE SPECIFIED AS MEAN±STANDARD DEVIATION.

Time point	Attribute	MMSE	ADAS-Cog
M12	Sample size (CN, MCI, AD)	1332 (358, 725, 249)	1299 (351, 716, 232)
	Gender(f/m)	579/753	565/734
	Age	75.0±7.1	75.1±7.1
M24	Sample size (CN, MCI, AD)	1110 (330, 617, 163)	1066 (321, 602, 143)
	Gender(f/m)	484/626	464/602
	Age	76.0±7.2	73.±7.1
M36	Sample size (CN, MCI, AD)	710 (192, 509, 9)	677 (186, 485, 6)
	Gender(f/m)	307/403	292/385
	Age	76.7±7.0	73.5±7.0
M48	Sample size (CN, MCI, AD)	456 (120, 334, 2)	431 (114, 315, 2)
	Gender(f/m)	195/261	185/246
	Age	77.1±6.9	73.1±6.9
M60	Sample size (CN, MCI, AD)	262 (88, 174, 0)	245 (86, 159, 0)
	Gender(f/m)	108/154	102/143
	Age	78.7±6.7	73.9±6.7

IV. EXPERIMENTS

A. Dataset

The Alzheimer's Disease Neuroimaging Initiative (ADNI) database was used to obtain data for this investigation (adni.loni.usc.edu). A University of California, San Francisco (UCSF) team employed the FreeSurfer image analysis system (http://surfer.nmr.mgh.harvard.edu/) to conduct volumetric segmentations and cortical reconstruction using imaging data from the ADNI database, which contains all ADNI subprojects (ADNI 1, 2, GO, 3). We collected MRI data from the ADNI website and then conducted through the following pre-processing steps: 1) Removal of image records with failed quality control; 2) Individuals who lacked BL and M06 MRIs were eliminated; 3) Removal of features with more than half of the samples having missing values; 4) Fill in the remaining missing data by taking the average of the features.

There are 313 MRI features in total after pre-processing operations, and they can be divided into five categories: the volumes of specific white matter parcellations (SV), the total surface area of the cortex (SA), the volumes of cortical parcellations (CV), average cortical thickness (TA) and standard deviation in cortical thickness (TS). The demographic details of the ADNI MRI data used in the research are shown in Table I.

B. Evaluation metrics

The similarity tensor of structural variation between brain biomarkers was utilised to construct predictive models for each target. We randomly split the data into a training and test set in

TABLE II. COMPARISON OF THE RESULTS FROM OUR PROPOSED METHODS WITH STATE-OF-THE-ART METHODS FOR MMSE AT TIME POINTS M12 TO M60. THE BEST RESULTS ARE BOLDED.

Target: MMSE	nMSE	wR	M12 rMSE	M24 rMSE	M36 rMSE	M48 rMSE	M60 rMSE
Ridge	2.4874±4.8534	0.2522±0.0963	4.4886±1.9690	5.6765±1.2738	6.0979±2.0777	7.5200±1.7215	7.2992±1.8936
Lasso	0.8947±0.0909	0.4307±0.0814	1.9714±0.2192	2.3112±0.2860	2.8583±0.6391	3.5472±0.6325	4.5034±1.0273
TGL	0.4020±0.0521	0.8184±0.0301	1.3829±0.2115	1.4261±0.1613	1.7775±0.3474	2.1648±0.5556	2.7628±0.3295
nCFGL1	0.3562±0.0415	0.8405±0.0138	1.3357±0.2420	1.4220±0.1461	1.7461±0.2546	2.1255±0.3665	2.7130±0.5322
cFSGL	0.3274±0.0238	0.8347±0.0248	1.3529±0.1276	1.4616±0.2137	1.6610±0.2686	2.0098±0.3717	2.6400±0.6849
FL-SGL	0.4988±0.0513	0.7475±0.0312	1.6909±0.1817	1.7383±0.1374	2.1410±0.5211	2.6623±0.4586	3.0783±0.5438
NC-CMTL	0.4973±0.0600	0.7670±0.0426	1.5728±0.1654	1.5987±0.1718	1.8805±0.1196	2.6200±0.4209	4.1446±1.2208
AMDQ-TMTL	**0.2771±0.0096**	**0.8739±0.0063**	**1.3146±0.0311**	**1.3638±0.1419**	**1.3317±0.1813**	**1.7059±0.3142**	**2.3047±0.3744**

TABLE III. COMPARISON OF THE RESULTS FROM OUR PROPOSED METHODS WITH STATE-OF-THE-ART METHODS FOR ADAS-COG AT TIME POINTS M12 TO M60. THE BEST RESULTS ARE BOLDED.

Target: ADAS-Cog	nMSE	wR	M12 rMSE	M24 rMSE	M36 rMSE	M48 rMSE	M60 rMSE
Ridge	2.3098±0.3173	0.2505±0.1042	11.1110±1.5294	11.1444±1.4665	13.5850±2.0725	14.9580±2.7276	18.9206±2.2966
Lasso	0.8855±0.1250	0.4065±0.0951	6.1045±1.2568	7.0128±1.2575	8.1833±1.5708	11.0483±1.8357	12.8538±2.5683
TGL	0.3180±0.0666	0.8568±0.0306	3.8917±0.7509	3.5174±0.4375	3.9497±0.6136	4.3520±1.0386	8.5801±1.3462
nCFGL1	0.2846±0.0715	0.8753±0.0395	4.1608±0.6385	3.6252±0.7720	3.6492±0.6487	5.8214±2.4860	6.6881±2.0821
cFSGL	0.3208±0.0717	0.8582±0.0275	3.9790±0.8610	3.5344±0.4348	4.0206±0.2632	4.6390±0.9821	8.0030±1.2742
FL-SGL	0.4817±0.0927	0.7841±0.0079	4.7609±1.0911	4.3630±0.4959	4.3774±0.3522	6.0582±1.0090	9.7911±1.6663
NC-CMTL	0.2393±0.0323	0.8826±0.0222	3.9980±0.6890	3.6433±0.7796	3.8006±0.5050	4.7475±1.9865	7.2107±1.6098
AMDQ-TMTL	**0.1585±0.0470**	**0.9102±0.0335**	**1.8916±0.4561**	**1.9747±0.1502**	**1.9433±0.2124**	**2.4944±0.7428**	**3.1906±0.4281**

a 9:1 ratio. Because the number of model parameters (λ, β and θ), the hyperparameters α and the latent factor r must be chosen during the training phase, we choose them utilising 5-fold cross-validation on the model training process.

The study compares the predicted ability of multiple approaches for each single time point using the root mean square error (rMSE) as the major assessment metric. For the overall regression performance evaluation, we use normalised mean square error (nMSE), which is used in multi-task learning research [27], and weighted correlation coefficient (wR), which is used in medical literatures to resolve AD progression problems [28]. The rMSE, nMSE and wR are defined as follows:

$$\mathrm{rMSE}(\mathrm{y}, \hat{y}) = \sqrt{\frac{\|\mathrm{y}-\hat{\mathrm{y}}\|_2^2}{n}} \tag{6}$$

$$\mathrm{nMSE}\ (\mathrm{Y}, \widehat{\mathrm{Y}}) = \frac{\sum_{i=1}^{t} \|\mathrm{Y}_i-,\widehat{\mathrm{Y}}_i\|_2^2 \Big/ \sigma(\mathrm{Y}_i)}{\sum_{i=1}^{t} n_i} \tag{7}$$

$$\mathrm{wR}(\mathrm{Y}, \widehat{\mathrm{Y}}) = \frac{\sum_{i=1}^{t} \mathrm{Corr}\ (\mathrm{Y}_i, \widehat{\mathrm{Y}}_i) n_i}{\sum_{i=1}^{t} n_i} \tag{8}$$

where y is the ground truth of target at a single time point and $\hat{\mathrm{y}}$ is the corresponding prediction from a prediction model for the rMSE. For the nMSE and wR, Y_i is the ground truth of target at time point i and $\widehat{\mathrm{Y}}_i$ is the corresponding prediction from a prediction model, Corr is the correlation coefficient between two vectors. The mean and standard deviation of 20 experimental iterations of the different data splits are reported.

V. RESULTS AND DISCUSSION

A. Comparison with the state-of-the-arts

We utilised the proposed amalgamated magnitude-direction quantification combined with the tensor multitask learning regression algorithm (AMDQ-TMTL) to compare with the following single-task learning and state-of-the-art multitask learning, which were selected as competing methods in the research of predicting clinical deterioration. Including Ridge regression (Ridge) [29], Lasso regression (Lasso) [30], Temporal Group Lasso (TGL) [2], Non-convex Fused Sparse Group Lasso (nFSGL1) [31], Convex Fused Sparse Group Lasso (cFSGL) [2], Fused Laplacian Sparse Group Lasso (FL-SGL) [32] and Non-Convex Calibrated Multi-Task Learning (NC-CMTL) [33]. Experimental results for MMSE prediction are shown in Table II and for ADAS-Cog prediction are shown in Table III.

The proposed approach has a lower rMSE than other models for all individual time points. In terms of overall regression performance, our proposed approach exceeds state-of-the-art algorithms in terms of nMSE and wR for MMSE and ADAS-Cog, demonstrating that our method outperforms competitors. Our observations are as follows: 1) The proposed AMDQ-TMTL approach outperforms single-task learning models and state-of-the-art MTL models, which validates the application of similarity calculation containing both magnitude and direction information for brain structural variation and the exploitation of the tensor latent factor hypothesis in our MTL formulation. 2) The proposed AMDQ-TMTL algorithm significantly improves prediction stability. The results obtained through 20 iterations of the experiment had a lower standard deviation than the state-of-

TABLE IV. THE TOP-10 RANK BRAIN BIOMARKER CORRELATIONS IN TIME POINT M12 FOR THE AMDQ-TMTL MODEL ON MMSE PREDICTION.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorParietal - CTA. of R.InferiorParietal	1.2248
CTA. of R.Postcentral - Vol(C). of R.Insula	1.0616
CTA. of R.RostralMiddleFrontal - Vol(C). of R.LateralOccipital	1.0461
Vol(C). of R.Precentral - Vol(C). of L.RostralMiddleFrontal	0.8838
Vol(WM). of CorpusCallosumAnterior - CTA. of R.InferiorParietal	0.8760
CTA. of R.Postcentral - CTA. of R.Precentral	0.8503
CTStd. of L.Postcentral - CTA. of R.InferiorParietal	0.8168
CTA. of L.ParsOpercularis - CTA. of L.RostralMiddleFrontal	0.8167
Vol(C). of R.Paracentral - Vol(C). of L.SuperiorFrontal	0.7991
CTA. of R.Pericalcarine - Vol(C). of R.InferiorTemporal	0.7875

TABLE V. THE TOP-10 RANK BRAIN BIOMARKER CORRELATIONS IN TIME POINT M24 FOR THE AMDQ-TMTL MODEL ON MMSE PREDICTION.

Brain biomarker correlation	Weight
Vol(C). of R.TransverseTemporal - CTStd. of R.Lingual	1.0065
CTA. of R.TransverseTemporal - CTA. of L.Precuneus	0.9556
CTA. of R.SuperiorParietal - Vol(C). of L.SuperiorFrontal	0.9520
CTA. of L.Precentral - CTA. of L.SuperiorTemporal	0.9135
Vol(C). of R.ParsTriangularis - Vol(C). of R.Precuneus	0.9068
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.8881
CTA. of L.InferiorParietal - CTA. of L.ParsOrbitalis	0.8698
Vol(C). of R.ParsTriangularis - Vol(C). of L.InferiorTemporal	0.8350
CTA. of L.CaudalMiddleFrontal - CTA. of L.SuperiorTemporal	0.8210
CTA. of R.Precuneus - Vol(C). of L.InferiorTemporal	0.8194

TABLE VI. THE TOP-10 RANK BRAIN BIOMARKER CORRELATIONS IN TIME POINT M36 FOR THE AMDQ-TMTL MODEL ON MMSE PREDICTION.

Brain biomarker correlation	Weight
CTA. of R.RostralMiddleFrontal - Vol(C). of R.LateralOccipital	1.4163
Vol(C). of R.SuperiorTemporal - CTA. of L.Pericalcarine	1.3244
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	1.2301
CTA. of L.ParsOpercularis - CTA. of L.RostralMiddleFrontal	1.0588
Vol(C). of L.MiddleTemporal - CTA. of L.SuperiorFrontal	0.9816
CTA. of R.RostralAnteriorCingulate - CTA. of L.CaudalMiddleFrontal	0.9768
CTA. of R.RostralAnteriorCingulate - Vol(C). of L.CaudalMiddleFrontal	0.9729
CTA. of R.RostralMiddleFrontal - CTA. of L.Supramarginal	0.9712
Vol(C). of R.SuperiorTemporal - CTA. of R.MiddleTemporal	0.9630
Vol(C). of L.IsthmusCingulate - CTA. of R.Fusiform	0.9447

TABLE VII. THE TOP-10 RANK BRAIN BIOMARKER CORRELATIONS IN TIME POINT M48 FOR THE AMDQ-TMTL MODEL ON MMSE PREDICTION.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	1.3938
Vol(C). of R.SuperiorTemporal - CTA. of L.Pericalcarine	1.2555
Vol(C). of R.SuperiorTemporal - CTA. of R.MiddleTemporal	1.1114
Vol(C). of R.TransverseTemporal - CTStd. of R.Lingual	1.1095
Surf. Area of R.Pericalcarine - CTA. of R.Pericalcarine	1.0473
Vol(C). of R.SuperiorTemporal - CTA. of L.Paracentral	1.0407
CTA. of L.InferiorParietal - CTA. of L.ParsOrbitalis	1.0220
CTA. of L.LateralOccipital - CTA. of R.LateralOccipital	0.9934
Vol(C). of L.MiddleTemporal - Vol(C). of L.SuperiorFrontal	0.9706
CTA. of R.SuperiorParietal - Vol(C). of L.SuperiorFrontal	0.9528

TABLE VIII. THE TOP-10 RANK BRAIN BIOMARKER CORRELATIONS IN TIME POINT M60 FOR THE AMDQ-TMTL MODEL ON MMSE PREDICTION.

Brain biomarker correlation	Weight
Vol(C). of R.Paracentral - Vol(C). of L.SuperiorFrontal	1.5628
CTA. of L.InferiorParietal - CTA. of L.ParsOrbitalis	1.5124
CTStd. of R.Postcentral - Surf. Area of L.PosteriorCingulate	1.5091
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	1.4332
Vol(C). of L.SuperiorFrontal - Surf. Area of R.InferiorTemporal	1.3372
Vol(C). of R.Paracentral - CTA. of L.CaudalMiddleFrontal	1.3096
Vol(C). of R.Precentral - Vol(C). of L.RostralMiddleFrontal	1.2670
Vol(C). of R.Paracentral - Vol(C). of L.MiddleTemporal	1.2659
CTA. of R.Precentral - CTA. of L.MedialOrbitofrontal	1.2594
Vol(C). of R.Postcentral - Vol(WM). of R.Hippocampus	1.2586

the-art comparison technique. This may be due to the proposed quantification method incorporates information on global brain variability and the addition of brain biomarker latent factors to the prediction algorithm to improve stability.

B. Interpretability

The interpretability of approach and results is as crucial in medical research as model performance. Because there is presently no cure for AD, the key to current treatment is early detection and prevention of the disease. Therefore, identifying significant brain biomarker structural variation relationships in early MRI data can help clinicians recognize individuals with suspected AD for early prevention. Since the MMSE dataset has a larger sample size than the ADAS-Cog dataset at each time point, it presents a more comprehensive range of samples. The top 10 brain biomarker structural variation correlations of the proposed AMDQ-TMTL method are shown in descending order of the weighted parameter values by MMSE prediction at different time points in Tables IV, V, VI, VII and VIII. Higher values imply a greater impact on the final prediction. The important brain biomarker relationships identified can be utilised as potential indicators for early identification of AD.

We discovered one of the brain structure variation correlations between brain biomarkers were important at most of time points (M24, M36, M48 and M60). That is Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal.

The superior frontal gyrus is a gyrus of the frontal lobe of the brain, occupying one-third of the size of the frontal lobe. It is an important area that controls many functions such as movement, working memory, resting state and cognition. And the superior frontal gyrus on one side of the cerebral hemisphere is responsible for planning various complex movements on the other side of the body [34].

The correlation between the cortical volume variation of the superior frontal gyrus in the right and left hemispheres of the brain can be a factor in the symptoms of memory loss, physical function deterioration, and cognitive dysfunction in AD patients.

VI. CONCLUSION

We propose a new quantification approach of brain structural variation for AD prediction scenarios combined with tensor multitask learning regression methods to predict AD progression at various time points to overcome variability and instability in prediction accuracy. Specifically, our quantitative approach considers both the magnitude and directional correlations of structural variation among brain biomarkers and quantifies them as third-order tensors to address data form monotonicity problem, and tensor multitask learning regression

uses tensor latent factors as multitask relationships to share knowledge and improve model generalisation to address small dataset problem. The experimental results demonstrate that the proposed approach can be utilized to identify brain structural differences in individuals with AD, MCI and CN, that it has the ability to predict and diagnose AD progression, and it only requires MRI data from patient to achieve superior prediction performance.

REFERENCES

[1] Z. S. Khachaturian, "Diagnosis of Alzheimer's Disease," *Arch Neurol*, vol. 42, pp. 1097–1105, November 1985.

[2] J. Zhou, J. Liu, V. A. Narayan, and J. Ye, "Modeling disease progression via multi-task learning," *NeuroImage*, vol. 78, pp. 233–248, September 2013.

[3] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," *IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 1015–1018, July 2014.

[4] R. S. Doody, V. Pavlik, P. Massman, S. Rountree, E. Darby, and W. Chan, "Predicting progression of Alzheimer's disease," *Nature BMC*, vol. 2, pp. 1–9, February 2010.

[5] Y. Zhang and Y. Qiang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1-20, 2021.

[6] T. G. Kolda, and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Review*, vol. 51, pp. 455–500, 2009.

[7] PM. Thompson, KM. Hayashi, ER. Sowell, N. Gogtay, JN. Giedd, JL. Rapoport, GI. Zubicaray, AL. Janke, SE. Rose, J. Semple, DM. Doddrell, Y. Wang, TGM. Erp, TD. Cannon and AW. Toga, "Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia," *NeuroImage*, vol. 23, no. 1, pp. S2-S18, 2004.

[8] P. Vemuri, H. J. Wiste, S. D. Weigand, L. M. Shaw, J. Q. Trojanowski, M. W. Weiner, D. S. Knopman, R. C. Petersen, C. R. Jack, "MRI and CSF biomarkers in normal, MCI, and AD subjects," *Neurology*, vol. 73, pp. 294–301, 2009.

[9] V. Singh, H. Chertkow, J.P. Lerch, A.C. Evans, A.E. Dorr, and N.J. Kabani, "Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease," *Brain*, vol. 129, pp. 2885–2893, 2006.

[10] CY. Wee, PT. Yap, and D. Shen, "Prediction of Alzheimer's Disease and Mild Cognitive Impairment Using Cortical Morphological Patterns," *Human Brain Mapping*, vol. 34, pp. 3411–3425, 2013.

[11] Y. He, Z. Chen, and A. Evans, "Structural Insights into Aberrant Topological Patterns of Large-Scale Cortical Networks in Alzheimer's Disease," *The Journal of Neuroscience*, vol. 28, pp. 4756–4766, 2008.

[12] R. De. Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The Mahalanobis distance," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, pp. 1–18, January 2000.

[13] P. Xia, L. Zhang, and F. Li, "Learning similarity with cosine similarity ensemble," *Information Sciences*, vol. 307, pp. 39–52, June 2015.

[14] P. Jiang, X. Wang, Q. Li, L. Jin, and S. Li, "Correlation-Aware Sparse and Low-Rank Constrained Multi-Task Learning for Longitudinal Analysis of Alzheimer's Disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, pp. 1450–1456, 2019.

[15] 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.

[16] M. Zhou, Y. Zhang, T. Liu, Y. Yang, and P. Yang, "Multi-task Learning with Adaptive Global Temporal Structure for Predicting Alzheimer's Disease Progression," *The 31st ACM International Conference on Information & Knowledge Management*, pp. 2743–2752, 2022.

[17] P. Cao, X. Liu, J. Yang, D. Zhao, M. Huang, and O. Zaiane, "$\ell 2$,$1-\ell 1$ regularized nonlinear multi-task representation learning based cognitive performance prediction of Alzheimer's disease," *Pattern Recognition*, vol. 79, pp. 195–215, 2018.

[18] 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.

[19] D. Zhang and D. Shen, "Multi-Modal Multi-Task Learning for Joint Prediction of Multiple Regression and Classification Variables in Alzheimer's Disease," *NeuroImage*, vol. 59, no. 2, pp. 895-907, 2012.

[20] N. Zeng, H. Li and Y. Peng, "A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease," *Neural Computing and Applications*, pp. 1-12, 2021.

[21] J. Chen, J. Zhou and J. Ye, "Integrating low-rank and group-sparse structures for robust multi-task learning," *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 42-50, 2011.

[22] Y. Zhang, P. Yang and V. Lanfranchi, "Tensor Multi-Task Learning for Predicting Alzheimer's Disease Progression using MRI data with Spatio-temporal Similarity Measurement," *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*, pp. 1-8, 2021.

[23] Y. Zhang, M. Zhou, T. Liu, V. Lanfranchi, and P. Yang, "Spatio-temporal Tensor Multi-Task Learning for Predicting Alzheimer's Disease in a Longitudinal study," *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 979-985, 2022.

[24] P. Gong, J. Zhou, W. Fan, and J. Ye, "Efficient multi-task feature learning with calibration," *KDD*, pp. 761–770, 2014.

[25] L. Han and Y. Zhang, "Learning multi-level task groups in multi-task learning," *AAAI*, pp. 2638–2644, 2015.

[26] Y. Zhang, T. Liu, V. Lanfranchi, and P. Yang, "Explainable Tensor Multi-task Ensemble Learning Based on Brain Structure Variation for Alzheimer's Disease Dynamic Prediction," *IEEE Journal of Translational Engineering in Health and Medicine*, pp. 1–12, 2022.

[27] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Mach. Learn*, vol. 73, pp. 243–272, January 2008.

[28] C. Stonnington, C. Chu, S. Klöppel, C. Jack Jr., J. Ashburner, and R. Frackowiak, "Predicting clinical scores from magnetic resonance scans in Alzheimer's disease," *Neuroimage*, vol. 51, pp. 1405– 1413, July 2010.

[29] A. Hoerl and R. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, vol. 12, pp. 55-67, 1970.

[30] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Ser. B Methodol*, vol. 58, pp. 267–288, 1996.

[31] J. Zhou, J. Liu, V.A. Narayan, and J. Ye, "Modeling Disease Progression via Fused Sparse Group Lasso," *KDD*, pp. 1095-1103, August 2012.

[32] X. Liu, P. Cao, A. Gonçalves, D. Zhao, and J. Banerjee, "Modeling Alzheimer s Disease Progression with Fused Laplacian Sparse Group Lasso," *ACM Transactions on Knowledge Discovery from Data*, vol. 12 pp. 1–35, December 2018.

[33] F. Nie, Z. Hu, and X. Li, "Calibrated Multi-Task Learning," *KDD*, pp. 19-23, August 2018.

[34] W. Li, W. Qin, H. Liu, L. Fan, J. Wang, T. Jiang, and C. Yu, "Subregions of the human superior frontal gyrus and their connections," *NeuroImage*, vol. 78, pp. 46-58, September 2013.