Investigation of Multi-dimensional Tensor Multi-task Learning for Modeling Alzheimer's Disease Progression

Yu Zhang

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Department of Computer Science

January 2024

Acknowledgment

Firstly, I would like to express my sincere thanks to my supervisory team, Dr. Po Yang and Dr. Vitaveska Lanfranchi, for their advice on how to conduct my research, how to present my work and how to publish my findings, and thank you very much for all your support. Particular thanks to my principal supervisor Dr. Po Yang for his support and assistance, he has provided me with many invaluable opportunities and constant support throughout my PhD research, and has been a role model not only in academia but also in life.

Many thanks to all my friends and colleagues in the Pervasive Computing research group at the University of Sheffield, with whom I have had the privilege of working. Without their presence and involvement, I would not have been able to achieve such a fruitful career in my PhD research. Also, I am very thankful to my good friends Jiaming and Lu, research life has been challenging and difficult. However, thanks to all my companions, I did not feel lonely and bored during this journey.

I am sincerely thankful to my family for their unconditional and enduring support, they did their best to support and encourage me to do what I wanted to do. My father and mother always give me their love, care and words. Thank you greatly to my partner Dana, there is so much I want to say but so little that can be expressed, thank you for your continuing support and encouragement. Studying for a PhD is challenging but undoubtedly exciting. Thank you to everyone I have met along the way.

List of Publications

Conference Paper (published)

Zhang, Y., Yang, P. and Lanfranchi, V., 2021, July. Tensor multi-task learning for predicting alzheimer's disease progression using MRI data with spatio-temporal similarity measurement. In *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)* (pp. 1-8). IEEE.

Zhang, Y., Liu, T., Li, Y., Wang, R., Huang, H. and Yang, P., 2021, December. Spatio-temporal Tensor Multi-Task Learning for Precision fertilization. In *2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS)* (pp. 398-405). IEEE.

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

Zhang, Y., Liu, T., Li, Y., Wang, R., Huang, H. and Yang, P., 2022, October. Spatio-temporal Tensor Multi-Task Learning for Precision Fertilisation with Real-world Agricultural Data. In *IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society* (pp. 1-6). IEEE.

Zhang, Y., Lanfranchi, V., Wang, X., Zhou, M. and Yang, P., 2022, December. 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 (BIBM)* (pp. 2735-2742). IEEE.

Zhou, M., **Zhang, Y.**, Liu, T., Yang, Y. and Yang, P., 2023. Robust Temporal Smoothness in Multi-task Learning. *AAAI-23* (pp. 1-9).

Zhou, M., **Zhang, Y.**, Liu, T., Yang, Y. and Yang, P., 2022, October. Multi-task Learning with Adaptive Global Temporal Structure for Predicting Alzheimer's Disease Progression. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management* (pp. 2743-2752).

Wang, X., Bi, G., **Zhang, Y.**, Qi, J., Yang, Y. and Yang, P., 2020, December. Visual Analysis of Biomarkers Selected via Multi-Task Learning for Modeling Alzheimer's Disease Progression. In *2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)* (pp. 1328-1333). IEEE.

Zhou, M., Wang, X., Yang, Y., Nan, F., **Zhang, Y.**, Qi, J. and Yang, P., 2021, December. Modeling disease progression flexibly with nonlinear disease structure via multi-task learning. In *2021 17th International Conference on Mobility, Sensing and Networking (MSN)* (pp. 366-373). IEEE.

Journal Paper (published)

Zhang, Y., Liu, T., Lanfranchi, V. and Yang, P., 2022. 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*, *11*, pp.1-12.

Zhang, Y., Wang, X., Liu, T., Wang, R., Li, Y., Xue, Q. and Yang, P., 2023. Sustainable fertilisation management via tensor multi-task learning using multi-dimensional agricultural data. *Journal of Industrial Information Integration*, *34*, p.100461.

Zhou, M., **Zhang, Y.**, Liu, T., Yang, Y. and Yang, P., 2023. Efficient multi-task learning with adaptive temporal structure for progression prediction. *Neural Computing and Applications*, pp.1-16.

Notation

Represent tensors as italic capital letters, such as X or Y

Represent matrices as capital letters, such as A or B

Represent vectors as lowercase letters, such as x or y

Represent scalars as italic lowercase letters, such as a or b

$\|\cdot\|_F^2$ the Frobenius norm

$\|\cdot\|_2^2$ the $\ell 2$ norm

$\|\cdot\|_1$ the $\ell 1$ norm

$\|\cdot\|_{2,1}$ the $\ell 2,1$ norm

$[\![\cdot]\!]$ the tensor decomposition

$[\![\cdot]\!]_S$ the symmetric tensor decomposition

[...] set of elements

$[\dots]^T$ the transpose of set of elements

$[a, b]$ the interval value between a and b (include a and b)

$\hat{y}$ the predicted value of y

S^{-1} the reciprocal of S

$\odot$ the dot product of elements between matrices

$\circ$ the outer product operation between two vectors

$\in$ correlation between element and set

$\mathbb{R}$ the set of real numbers

$\sigma()$ the standard deviation

$\mathrm{Corr}()$ the correlation coefficient

Abstract

Machine learning (ML) techniques for predicting Alzheimer's disease (AD) progression can significantly assist clinicians and researchers in constructing effective AD prevention and treatment strategies. The main constraints on the performance of current ML approaches are prediction accuracy and stability problems in medical small dataset scenarios, monotonic data formats (loss of multi-dimensional knowledge of the data and loss of correlation knowledge between biomarkers) and biomarker interpretability limitations. This thesis investigates how multi-dimensional information and knowledge from biomarker data integrated with multi-task learning approaches to predict AD progression. Firstly, a novel similarity-based quantification approach is proposed with two components: multi-dimensional knowledge vector construction and amalgamated magnitude-direction quantification of brain structural variation, which considers both the magnitude and directional correlations of structural variation between brain biomarkers and encodes the quantified data as a third-order tensor to address the problem of monotonic data form. Secondly, multi-task learning regression algorithms with the ability to integrate multi-dimensional tensor data and mine MRI data for spatio-temporal structural variation information and knowledge were designed and constructed to improve the accuracy, stability and interpretability of AD progression prediction in medical small dataset scenarios. The algorithm consists of three components: supervised symmetric tensor decomposition for extracting biomarker latent factors, tensor multi-task learning regression and algorithmic regularisation terms. The proposed algorithm aims to extract a set of first-order latent factors from the raw data, each represented by its first biomarker, second biomarker and patient sample dimensions, to elucidate potential factors affecting the variability of the data in an interpretable manner and can be utilised as predictor variables for training the prediction model that regards the prediction of each patient as a task, with each task sharing a set of biomarker latent factors obtained from tensor decomposition. Knowledge sharing between tasks improves the generalisation ability of the model and addresses the problem of sparse medical data. The experimental results demonstrate that the proposed approach achieves superior accuracy and stability in predicting various cognitive scores of AD progression compared to single-task learning, benchmarks and state-of-the-art multi-task regression methods. The proposed approach identifies brain structural variations in patients and the important brain biomarker correlations revealed by the experiments can be utilised as potential indicators for AD early identification.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Predicting disease progression and identifying biomarkers that track disease progression is an important aspect of the application of computer science. Accurately predicting disease progression can assist clinicians and patients to make the best decisions. To achieve this goal, various prediction methods based on machine learning algorithms are being delivered, such as regression algorithms, kernel methods and deep learning. However, they all suffer from various drawbacks that make the accuracy, stability and interpretability of the algorithms not reaching the desired level for predicting disease progression, such as algorithm performance issues in medical small dataset scenarios, loss of multi-dimensional information in medical data, and interpretability of algorithm results in the medical domain. Therefore, it is important to explore new disease progression prediction algorithms to replace traditional prediction methods. Quantifying the data and constructing it into a multi-dimensional tensor combined with machine learning algorithms can exploit the multi-dimensional information to predict the target. The aim of this thesis is to explore how multi-dimensional tensor-based data combined with multi-task learning methods can be utilised to predict disease progression in the best way to achieve accurate, stable and interpretable disease progression predictions. This chapter briefly describes the research background, motivation, research issues, research objectives and contributions of this research.

1.1 Background

Alzheimer's disease is a serious primary neurodegenerative disease that causes neurons and their connections degenerate gradually, resulting in a comprehensive spectrum of dementia including memory loss, cognitive decline and executive dysfunction [1]. The etiology of AD is unclear, but may be related to a variety of factors including neurotransmitters, environment, genetics, inflammation and oxidative stress. Current clinical studies of AD are characterised by the pathology of neurofibrillary tangles and amyloid plaques in the cerebral cortex and hippocampus, leading to brain atrophy and neuronal death [2][3]. According to World Health Organization (WHO), it is estimated that there were more than 55 million dementia patients worldwide in 2023, with nearly 10 million new cases each year. AD is the most common cause of dementia, accounting for approximately 60-70% of dementia cases. There is presently no cure to treat or reverse the disease progression which causes massive psychological and emotional stress for patients and their families. The current diagnosis of AD primarily relies on clinical manifestations, neuropsychological cognitive tests, brain imaging and cerebrospinal fluid examination [4][5]. Neuropsychological cognitive tests can evaluate the cognitive function of patients, such as Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS- Cog) and Mini Mental State Examination (MMSE). Brain imaging can demonstrate characteristic changes such as cerebral atrophy, reduced cerebral metabolism, reduced cerebral blood flow and amyloid plaque deposition, such as magnetic resonance imaging (MRI), positron emission tomography (PET) and single photon emission computed tomography (SPECT). Numerous researches have been conducted to identify precise and sensitive biomarkers of early AD progression that can aid in early AD diagnosis to create, validate and evaluate current and novel treatment. And the discovery of the correlation between AD biomarkers can help doctors understand, prevent and detect AD clinically.

Numerous studies in last decade have been carried out for predicting the AD

progression and the identification of sensitive and specific AD biomarkers, existing AD progression models mainly utilise machine learning regression algorithms [6][7][8][9] [10][11][12], survival models based on statistical probabilities [13][14][15], and deep learning methods based on neural networks [16][17]. Clinical applications of AD progression models include 1) Early diagnosis [16][18]: The established prediction models and algorithms can assist clinicians in identifying the risk of AD before the onset of symptoms or at an early stage, enabling earlier intervention and treatment. 2) Disease progression prediction [13][19]: By utilising dynamic data and long-term follow-up to establish progression curves and prognostic indicators for AD, it can assist clinicians in evaluating the changes and future development of patients' conditions, as well as formulating personalised treatment plans and care protocols for patients. 3) Biomarker researches [20][21]: By analysing variations in biomarkers, researchers can identify specific indicators associated with AD. Variations in biomarkers can be utilised to predict the development and progression of the disease. 4) Research on disease mechanisms [22][23]: By utilising multi-disciplinary and multi-level data to establish the pathophysiological networks and causal relations of AD, it can assist researchers in exploring the mechanisms of disease development and influencing factors, providing researchers with new research ideas and targets. 5) Functional connectivity and network analysis [24][25]: The research results of brain networks contribute to the understanding of the connections and synergies between different brain regions. By analysing these functional connections, abnormal patterns associated with AD can be discovered. This approach can alternatively be utilised to construct predictive models to identify variations in a patient's brain network, providing a more comprehensive understanding of the disease progression. 6) Personalised healthcare [26][27]: Clinicians can utilise predictive models to better understand individual differences in patients and develop personalised treatment plans for each patient to improve treatment effectiveness. 7) Medicine research and development [28][29]: Predictive models can assist in the design of accurate drug trials and accelerate the discovery of new medicines. 8) Screening of high-risk groups [30][31]: By utilising demographics, genes, lifestyle

and other factors to establish a risk assessment tool for AD, it can assist in identifying those who are predisposed to the onset of the disease, and provide them with timely preventive measures and intervention programmes. 9) Assessment of treatment effects [32][33]: Models and algorithms can be utilised to assess the effects of different treatments on patients, including medication, cognitive training and lifestyle interventions. By analysing these data, researchers and clinicians can better understand the effectiveness of different treatment strategies and optimise patients' treatment plans. 10) Clinical decision support systems [34][35]: Embedding models into clinical decision support systems can provide real-time, individualised advice to clinicians. This can contribute to the optimisation of clinical workflows and enhance clinicians' management of AD patients. In historical algorithm researches and experiments, all algorithms aiming at disease progression prediction require to regard the following performance benchmarks:

Accuracy: The accuracy of prediction is a measure of how correct disease progression prediction algorithms, models and systems are. Statistical approaches (root mean square error, mean absolute error, mean square error, etc.) are typically utilised as performance metrics, which estimate the degree of error between the predicted and true values. Accuracy is a fundamental requirement for most disease progression prediction algorithms, models and systems. The higher the accuracy, the better the algorithm, model and system.

Stability: The stability of disease progression prediction algorithms, models and systems is to demonstrate the consistency of the predicted results with the actual results for different individuals. Accuracy only measures the statistical methodological error level. However, prediction stability considers and measures the consistency of prediction algorithms, models and systems as it reveals the variation in their performance over multiple trials.

Interpretability: The interpretability of approaches and results is as important in medical research as approach performance to enable clinicians and patients to understand the decision-making process of algorithms, models and systems and trust their predictions. In addition, high interpretability of disease progression prediction algorithms, models and systems can recognize important biomarkers in the data to assist clinicians in identifying patients with suspected disease for early prevention.

Scalability: The scalability of disease progression prediction algorithms, models and systems refers to their ability to perform consistently for different prediction targets. Furthermore, the algorithms can be applied to various types of disease progression.

Cost: The cost of disease progression prediction algorithms, models and systems can depend on a number of factors. Important factors include the generation and collection of biomarker data, the training and deployment of algorithms, models and systems.

Predicting the early detection and diagnosis of Alzheimer's disease is extremely imperative for patients and families. It not only helps patients to access early intervention and treatment, but also helps to reduce the adverse effects of the disease. At the same time, research into predicting Alzheimer's disease also helps to better understand the mechanisms by which the disease occurs, and ultimately to find a cure. As a result, research into AD prediction algorithms will continue to be of interest and will continue to be extensively researched and developed. Researchers will continue to strive to improve the accuracy, stability, interpretability and reliability of predictive algorithms to better support patients and families. Ultimately, through early detection and diagnosis of Alzheimer's disease, we can improve the quality of life of patients and help reduce the adverse effects of the disease.

1.2 Motivation

The current state-of-the-art AD prediction algorithms are classified into the following main categories. Imaging-based prediction: This type of algorithm predicts the risk of Alzheimer's disease based on brain imaging, such as MRI or PET. Biomarker-based prediction: These algorithms utilise biological markers, such as blood biology indicators and genetic features to predict the risk of Alzheimer's disease. Behavioural and psychological prediction: These algorithms predict the risk of Alzheimer's disease based on the behavioural and psychological characteristics of the individual, such as cognitive tests and mental status assessments. Multi-modal algorithms: This class of algorithms combines several of these approaches to improve the accuracy of prediction algorithms. Specifically, the biomarker prediction algorithms include logistic regression-based biomarker prediction models and support vector machine-based biomarker prediction models. Brain image analysis prediction algorithms include convolutional neural network-based and generative adversarial network-based brain image analysis prediction models. Behavioural and psychological assessment prediction algorithms include decision tree-based and random forest-based prediction models for behavioural and psychological assessment. Deep learning prediction algorithms include models based on convolutional neural networks, generative adversarial networks and graph neural networks. The choice and utilisation of these algorithms depends on the specific requirements of the prediction task, as well as the requirements for accuracy and efficiency of the prediction. For example, a deep learning algorithm can be chosen to be used if the requirement for accuracy is high, or a biomarker prediction algorithm can be used if the requirement for prediction efficiency is high. In addition, the performance of prediction algorithms is usually affected by the quality of the data and the amount of data. Therefore, before utilising a prediction algorithm, a detailed evaluation of the data should be carried out and ensure that the data quality is high and the data amount is large.

The algorithms described above all focus on converting patient data into second-order matrices for linear or non-linear prediction. In most AD progression prediction studies, different biomarkers for each patient are composed into a second-order matrix. While these typical technical algorithms can make acceptable predictions and identify biomarkers, there are two limitations that can affect the clarification of biomarker correlations and the accuracy of predictions. The first limitation is the monotonic data form, which in most of these methods is utilised and displayed in the form of a second-order matrix. In the AD progression prediction model described above, the algorithm uses a second-order input matrix formed by clinical data or biomarker data from different AD patients combined with a second-order target matrix formed by the patient's AD cognitive test scores (e.g., ADAS-Cog and MMSE) at different time periods to obtain a second-order weight matrix of biomarkers at different time periods. However, the second-order matrix of AD biomarker data can only be predicted and analysed in one dimension, making it difficult to predict and analyse the disease progression in multiple dimensions (e.g., spatial and temporal). The second limitation is the correlation between AD biomarkers. Numerous studies have demonstrated that applying relationships between biomarkers to machine learning algorithms can improve prediction accuracy and understanding of biomarkers and disease progressions, but second-order matrices can only contain data on the biomarkers themselves and cannot include correlations between biomarkers. Therefore, the utilisation of higher-order tensor combined with machine learning algorithms for AD prediction, biomarker correlation and importance detection remain a challenging problem.

Multi-dimensional spatio-temporal data, including brain images, biomarkers and behavioural assessment data, have a wide range of applications in Alzheimer's disease prediction algorithms. By utilising information from multi-dimensional spatio-temporal data, algorithms can better assess the risk of Alzheimer's disease and predict the progression of the disease. For example, by analysing brain imaging data, algorithms can identify damage and degenerative changes in the brain and assess them;

by analysing behavioural assessment data, algorithms can assess a patient's cognitive function, such as memory and language skills. At the same time, multi-dimensional spatio-temporal data can also help the algorithm to identify early warning signs of disease, thus providing assistance for early intervention and prevention. The utilisation of multi-dimensional spatio-temporal data for Alzheimer's disease prediction can provide more valuable information and thus improve the accuracy of predictions. However, there are certain challenges. The first challenge is that for spatio-temporal data, temporally and spatially relevant features need to be extracted from the data. This can be a technically challenging task as complex temporal and spatial correlations need to be considered. A second challenge is the interpretability of algorithms in multi-dimensional spatio-temporal data. In multi-dimensional spatio-temporal data, the decisions of the algorithm are difficult to interpret, which can have an impact on the trust in the decision. Therefore, when utilising multi-dimensional spatio-temporal data for prediction, there is a need to consider how to enhance the interpretability of algorithm results in the context of medical knowledge and clinical scenarios. The third challenge is that complex spatio-temporal correlations, where there are complex correlations between spatio-temporal features, require prediction algorithms with superior model complexity and robustness.

In addition, a common problem with AD prediction algorithms is the size and diversity of dataset. The small dataset problem is a common problem with AD prediction algorithms. As AD is a chronic disease, the time taken to collect relevant data is long and the prevalence of AD is relatively low, therefore current datasets are typically small for the training and testing of algorithms. Numerous researchers have relied on small datasets to assess the performance of algorithms, but this has led to instability in the results. Utilising small datasets can also lead to overfitting of the model as it does not capture the true distribution of the data well. This can lead to the algorithm making inaccurate predictions for specific types of samples, thus reducing overall accuracy. In addition, the utilisation of small datasets can also lead to algorithms that do not have

generalisation capabilities and therefore can perform poorly on new data.

Multi-task learning can provide great advantages in small data set scenarios. Multi-task learning is a machine learning technique for learning models of multiple related tasks, which aims to exploit the commonality of multiple tasks to improve the performance of the model. In small dataset scenarios, the advantage of multi-task learning is that it utilises the full information of the dataset, thus compensating to a certain extent for the lack of a dataset. Due to the limited amount of data in small datasets, it is difficult to fit data from a single task, but with multi-task learning, the model can be trained together utilising data from multiple related tasks, thus overcoming the difficulties of single-task learning on small datasets. Multi-task learning in small data set scenarios can make full utilise of information from related tasks, thus improving the generalisation ability of the model. For example, if there are two tasks that are both AD predictions, then by learning both tasks at the same time, the model can take full advantage of the features common to both tasks to improve the accuracy of the model. Thus, multi-task learning can overcome the problem of low generalisation ability of the model in small data set scenarios. In multi-task learning, certain common knowledge can be shared by multiple tasks, which can reduce the number of parameters and prevent overfitting. At the same time, this sharing mechanism allows the model to obtain better generalisation on small datasets, thus achieving greater improvements in the accuracy of the prediction results.

Interpretability is an important challenge in AD prediction algorithms. The interpretability of an AD prediction algorithm refers to the readability and transparency of the algorithm's results and decision-making process. Interpretability is a key issue for AD prediction algorithms as the results of the algorithm can be wrong or biased and the lack of readability and transparency can lead to a lower level of trust in the algorithm by doctors and patients. The results of a model cannot just be considered a black box model, the results must be understandable and interpretable by humans. Machine learning techniques can generate accurate predictions, but if the decisions of the model

cannot be interpreted, they cannot be accepted by doctors or patients. Furthermore, if the algorithm's decision-making process is not readable, then the algorithm cannot be easily checked for any bias or errors, which could lead to unjust results. Due to the complex and multi-dimensional nature of AD, various machine learning models can struggle to fully understand the model's decision-making process when making predictions. This leads numerous doctors and specialists to doubt the results of machine learning models and consequently places limitations on their application. Furthermore, if the results of the models are not interpretable, then they cannot be used for revision or improvement, thus the issue of interpretability in AD prediction algorithms is an important one that needs to be addressed. The problems with the interpretability of AD prediction algorithms are mainly manifested in: 1) Model complexity: most of the current AD prediction algorithms utilise deep learning and machine learning techniques, and the model complexity is high and difficult to interpret. 2) Lack of transparency: due to the complexity of the model, it is difficult to know how the algorithm arrives at the prediction result, and there is a lack of transparency. 3) Lack of clarity in judgement criteria: it is difficult to judge whether the prediction results are reliable as there are no clear criteria to evaluate the accuracy of the prediction algorithms. 4) Data deviation: if there is deviation in the training data, the prediction results can also be affected and the root cause of the deviation is difficult to trace. Overall, the issue of interpretability of Alzheimer's disease prediction algorithms is an important challenge that needs further research. Therefore, there is a need for continued efforts to improve the interpretability and reliability of AD prediction algorithms in order to better assist clinicians and patients in the prevention and treatment of the disease.

Consequently, the motivation for this research work is to investigate the utilisation of multi-dimensional information knowledge from biomarker data combined with multi-task learning approaches for AD progression prediction. This project focuses on the three main issues mentioned above including prediction accuracy and stability in small dataset scenarios, monotonic data format (loss of multi-dimensional knowledge of the

data and loss of association knowledge between biomarkers) and biomarker interpretability limitations. The aim is to investigate the quantification and construction of AD biomarker data into a high-dimensional tensor combined with multi-task learning approaches to construct algorithms with high accuracy, stability and interpretability.

1.3 Aim and Objectives of Research

The following section describes the main aim and objectives of the proposed project:

a) Aim

The aim of this project is to investigate the quantification and construction of AD biomarker data as a higher-order tensor, combining multi-task learning and machine learning regression approaches to construct algorithms with multi-dimensional spatio-temporal characteristics of AD biomarker data to predict disease progression in AD patients at multiple future time points. To achieve this aim, the research conducts extensive experiments and tests to construct predictive models based on the above algorithms, which can simultaneously comprise feature correlations of different dimensions to improve prediction accuracy, stability and biomarker interpretability.

b) Objectives

The aim of this project is to investigate the utilisation of a multi-task learning approach, machine learning regression algorithms combined with multi-dimensional tensor-based data features to predict the disease progression in AD patients and to assist in understanding the importance and correlation of biomarker features. In order to investigate it, the research provides a number of objectives, which are defined as follows.

- To review research and applications on multi-dimensional tensor, multi-task learning and machine learning regression algorithms.
- To study AD biomarker data and investigate approaches to quantify and construct multi-dimensional tensors.
- To analyse and evaluate the optimal type of multi-dimensional tensor for AD biomarker data.
- To combine multi-dimensional tensor AD data, multi-task learning and regression algorithms to construct prediction algorithms and models.
- To implement algorithms and models, evaluate their performance.
- To utilise data analysis and visualisation methods to enhance the interpretability of algorithms and biomarker features.
- To analyse and compare the importance of AD biomarkers from different dimensions and discover biomarkers and biomarker correlations that can be utilised as potential indicators for early AD identification.

1.4 Contribution to Knowledge

To summarize, the main knowledge contributions are:

1. This research proposed a novel approach to analyse and quantify MRI biomarkers, which utilises similarity calculations to reflect the spatio-temporal variability between brain biomarkers, which simultaneously considers the magnitude and directional correlation of structural variation between brain biomarkers, the results of ablation studies have demonstrated that the multi-dimensional tensor multi-task learning regression model with the proposed quantification approach outperforms the mainstream correlation calculations for AD progression predictions, it contains comprehensive knowledge of brain structural variation and can effectively differentiate between CN (cognitively normal elderly), MCI (mild cognitive impairment) and AD patients, and constructs MRI biomarker data as a third-order tensor to address the problem of monotonic data forms.

2. The research proposed novel multi-dimensional tensor-based multi-task learning regression algorithms for the Alzheimer's disease progression prediction. The algorithms utilise tensor decomposition techniques to learn task correlations from raw data, which allows all samples to share latent knowledge of biomarkers based on brain structural variation. The algorithm is designed to seamlessly integrate spatio-temporal information and knowledge based on structural brain variants and their biomarker latent factors, experimental results have demonstrated that the proposed approach outperforms comparative methods in AD progression prediction based on various cognitive scores, significantly improves the accuracy and stability of AD progression prediction in medical small dataset scenarios.

3. This research identified important structural variation correlations between brain biomarkers and presents detailed analyses in terms of brain biomarker function, distributional differences in early MRI biomarker quantification values along with relative structural variation information of brain biomarkers between cognitively impaired and non-cognitively impaired individuals at various specific AD progression time points. These biomarker correlations can be exploited to predict AD progression and can be utilised as potential indicators for early identification of AD.

4. This research visualised and analysed the biomarker latent factors learned from raw data with tensor decomposition techniques to identify brain regions affected in the AD progression. The visualisation results have demonstrated that for different AD progression time points, the latent factors have different spatial distributions, which indicates that they capture different aspects of the spatial variability in the data.

1.5 Organization of the thesis

This thesis is divided into eight major chapters. The first chapter is the introduction, motivation and knowledge contribution to this research work. Chapter 2 begins with an introduction to researches on various AD prediction algorithms, a literature review of multi-task learning algorithms, and algorithm applications based on multi-task learning. Then it presents the multi-dimensional data, multi-dimensional tensor, tensor decomposition techniques and correlation quantification approaches that are relevant to this research. Chapter 3 elaborate on the methodology to accomplish the aim of this research. Chapter 4 proposes, designs and constructs an approach for the construction and quantification of a biomarker multi-dimensional tensor with the ability to mine and incorporate multi-dimensional spatio-temporal information and biomarker correlation knowledge. Chapter 5 designs and develops the multi-dimensional tensor multi-task learning regression algorithm and model to depict the utilisation of multi-task learning algorithms integrated with multi-dimensional tensor data for AD progression prediction. It exploits multi-task learning concepts incorporated with quantitative multi-dimensional tensor of spatio-temporal structural variation information of MRI brain biomarkers to enhance the accuracy, stability and interpretability of AD progression prediction in medical small data set scenarios. And then presents the experimental configurations and processes required for the multi-dimensional tensor multi-task learning regression algorithm and model, coupled with the testing and analysis of the algorithm and model results. Chapter 6 presents a multifaceted interpretability analysis for the experimental results and proposed algorithm. Chapter 7 designs and constructs various multi-dimensional tensor multi-task learning regression based time-continuous algorithms to predict disease progression at different time points in neurological disease prediction scenarios in order to simultaneously overcome the problems of monotonic data forms, small datasets and the scarcity of time-continuous data. The final chapter concludes the research work in this thesis and discusses possible future research, development work and enhancements.

Chapter 2

Literature review

There are four main aspects to this chapter: first, an overview of the research areas that have influenced the research. This includes AD clinical overview, research on various AD prediction algorithms, a review of multi-task learning algorithms and applications of algorithms based on multi-task learning. It discusses the limitations of existing and state-of-the-art prediction algorithms in medical small dataset scenarios and why multi-task learning approaches can potentially be utilised to enhance the performance of algorithms and models in this scenario. The second part of the chapter reviews current tensor decomposition techniques and their latent factors and applications in algorithms. It also analyses the range of applications of multi-dimensional tensor-based prediction algorithms. It discusses the limitations of traditional matrix data in the application of prediction algorithms and why high-dimensional tensor can be utilised to enhance the performance of AD prediction algorithms. The third part of this chapter reviews current research and applications in the field of brain science on the differences in brain structural variation among CN, MCI and AD along with the correlation between AD MRI biomarkers. The final part of this chapter reviews the major approaches utilised for biomarker correlation calculations and their associated techniques, including Cosine similarity, Euclidean distance and Mahalanobis distance.

2.1 Alzheimer's disease clinical overview

Alzheimer's disease is a critical primary neurodegenerative disorder that leads to a comprehensive spectrum of dementia, including memory impairment, visuospatial

perception impairment, language impairment, anosognosia, prosopagnosia and executive dysfunction, along with non-cognitive psychiatric symptoms including anxiety, depression, delusions, hallucinations and aggressive behaviour [1][36][37]. The etiology and pathogenesis of AD have not been fully clarified, but current research indicates that it may be related to the following aspects: 1) Genetic factors [2][3]: AD is associated with mutations or variations in certain genes, such as amyloid precursor protein, apolipoprotein E and presenilin. Abnormalities in these genes affect the production and removal of amyloid from the brain, leading to amyloid deposition and aggregation and the formation of neurotoxic amyloid plaques. Amyloid deposition can cause inflammatory response, oxidative stress and apoptosis in neurons, further impairing brain function. 2) Environmental factors [38][39][40]: The occurrence of AD can be affected by a variety of environmental factors, such as level of education, traumatic brain injury, aluminium poisoning, air pollution etc. These factors can directly or indirectly interfere with the metabolism, neuroprotective mechanisms and signal transmission in the brain, thus increasing the risk of AD. 3) Neurotransmitter system dysfunction [41][42]: AD patients have declining levels of various neurotransmitters in the brain, such as glutamate, acetylcholine, dopamine, etc. These neurotransmitters are closely associated with cognitive functions including learning and memory, and their reduction can lead to communication disorders between neurons and loss of synaptic function. In addition, excessive phosphorylation of Tau protein exists in the brain of AD patients, leading to the detachment of Tau protein from microtubules and the formation of neurogenic fibre tangles, which affects the transport function and structure of neurons. 4) Oxidative stress and free radical damage [43][44]: Oxidative stress and overproduction of free radicals are frequently observed in the brains of AD patients. Oxidative stress leads to increased production of free radicals inside and outside the cell, and these highly reactive molecules can damage cell membranes, nucleic acids and proteins, ultimately accelerating neuronal degeneration.

Diagnosis of Alzheimer's disease is a complex process that requires a variety of methods including physical examinations, comprehensive medical histories, cognitive

tests, brain imaging and blood tests [4][45][46]. Cognitive tests (e.g., MMSE, ADAS-Cog, etc.) can evaluate the patient's memory, language, attention, executive and visuospatial abilities. These tests can assist the clinician in determining if the patient has MCI or AD, along with the severity of condition [47][48]. Brain imaging tests (e.g., MRI, PET, SPECT, etc.) can observe the brain for structural and functional abnormalities such as ischemia, atrophy and decreased metabolism. These tests can assist the clinician to exclude other brain disorders that can cause cognitive impairment, such as Parkinson's disease, cerebrovascular disease, frontotemporal lobe dementia, etc [49][50]. Blood tests (e.g., routine blood, immunological, biochemical, hormonal, etc.) can exclude other systemic diseases that can affect cognition, such as anaemia, vitamin deficiencies, abnormal thyroid function, etc [51][52]. By conducting a variety of tests, clinicians can comprehensively evaluate whether a patient has AD, along with the staging and prognosis of the condition. Diagnosis of AD is a continual process that requires periodic review and subsequent monitoring to detect variations in the condition and the effectiveness of treatment. However, there is no cure to treat or reverse the disease progression, which causes immense psychological and emotional stress to patients and their families. Current treatment focuses on slowing the progression of the disease, relieving symptoms, and improving the quality of life for patients [4][53][54].

2.2 State-of-the-art in Alzheimer's disease prediction algorithm

Numerous researches have been performed in order to detect precise and sensitive biomarkers of early AD progression that can assist in early AD diagnosis to develop, assess, examine current and novel treatments. Previous researches have focused on utilising biomarkers combined with ML algorithms to predict patients' cognitive test scores as target data for diagnosing the degree of cognitive impairment in patients. The purpose of the AD prediction algorithm is to predict the cognitive scores (e.g., ADAS-Cog and MMSE) at different future time points based on the currently available information.

2.2.1 Regression algorithm

AD prediction is achieved with machine learning regression algorithms. The regression algorithm starts from a set of data, determines the quantitative relationships between certain variables, builds a mathematical model and estimates the unknown parameters. The aim of regression is to predict numerical target values; it aims to accept continuous data, find the equation that best fits the data and be able to predict a particular value. Algorithmic approaches include linear regression-based approaches, support vector regression-based approaches, regression tree-based approaches and neural network-based approaches. The advantages and disadvantages of these approaches are summarised in Table 2.1 and sections 2.2.1.1 to 2.2.1.4 present a review of each approach and discuss in detail the applications and research directions, advantages and disadvantages of these approaches. These methods work by building models and training them with training data to identify predictive markers and make predictions about future conditions. Specifically, Alzheimer's disease prediction algorithms utilise different data sources and methods to predict an individual's risk of developing Alzheimer's disease, including imaging-based algorithms that utilise brain imaging techniques, such as MRI or PET scans, to identify biological markers in the brain that can be associated with Alzheimer's disease. Biology-based algorithms utilise biological data associated with Alzheimer's disease, such as genetic data, to predict the risk of the disease. Behaviour-based algorithms utilise behavioural data related to cognitive function, such as memory assessment, to predict disease risk. Fusion algorithms, on the other hand, combine multiple data sources and methods to improve the accuracy of predictions.

To sum up, the key advantages of AD regression algorithms are 1) Fast diagnosis: predictive algorithms can analyse large amounts of data in a short time, faster than

Table 2.1: Advantages and disadvantages of various Alzheimer's disease prediction models and algorithms.

Approaches	Advantages	Disadvantages
Linear regression-based approaches	1) Interpretability 2) Computational efficiency 3) Robustness 4) Model simplicity 5) Do not require big data	1) Linear assumptions 2) High-dimensional data 3) Unbalanced data 4) Possible overfitting problems
Support vector regression-based approaches	1) Non-linear modelling capability 2) Robustness 3) Sparsity 4) Multi-scale learning	1) Computational complexity 2) Lack of interpretability 3) Kernel function selection
Regression tree-based approaches	1) Interpretability 2) Processes non-linear relationships 3) Effective model visualisation	1) Instability 2) Requires large amounts of data 3) Processing of high-dimensional data 4) Variable correlation 5) Processing of continuous outputs
Neural network-based approaches	1) Processing of high-dimensional data 2) Multi-modal data integration 3) Model complexity and flexibility 4) End-to-end learning	1) Lack of interpretability 2) Large amount of data required 3) Hardware dependency 4) Long training times

manual analysis. 2) Unbiased: predictive algorithms are not influenced by human bias and can evaluate data more objectively. 3) More accurate diagnosis: predictive algorithms can make predictions based on a variety of factors, more accurately than relying on human judgement alone. 4) Provides more Information: the prediction algorithm can provide more information based on different factors to help doctors make more informed decisions. However, there are a number of disadvantages. 1) Data limitations: prediction algorithms require a large amount of data as training samples, and prediction results are not satisfactory when there is insufficient data. 2) High algorithm complexity: there are complex calculations within the prediction algorithm, and it can take a long time to produce results for certain complex data. 3) High requirement for expertise: prediction algorithms require a certain level of understanding of relevant expertise, it can be difficult to understand the principles and results of the algorithm without the relevant knowledge.

2.2.1.1 Linear regression-based approaches

Linear regression is a statistical approach utilised to predict numerical data based on a linear relationship between one or more independent variables and a dependent variable [55][56]. In the case of AD, it implies that the model will attempt to discover linear equations describing how to predict disease progression or clinical scores based on one or more predictive variables (e.g., MRIs, magnitude of brain structures, cerebrospinal fluid biomarkers, genetic traits, psychological assessments, etc.). These data can provide models with information and knowledge about structural, functional and chemical changes in the brain that are critical for early diagnosis and prediction of disease progression. By linear regression modelling, researchers can identify a number of key factors associated with the progression of AD. These models can help in the formulation of individualised treatment strategies and provide valuable information for medicine development.

Current researches on linear regression-based AD prediction models and algorithms focus on: 1) Ensemble learning [57][58]: combining predictions from multiple models to improve accuracy and stability. 2) Novel regularisation methods [12][59][60]: application of regularisation methods can assist in dealing with multi-collinearity and overfitting problems. 3) Linear fusion of multi-modal data [61][62][63]: considering that AD researches can involve data from different sources (e.g., MRI, PET scans, genetic data, etc.), linear regression models can be utilised to fuse the data from these sources. 4) Segmented linear regression [64][65]: the AD progression can exhibit different rates at different stages. Segmented linear regression can be utilised to capture this non-linear relationship while maintaining a linear model within each segment. 5) Linear mixed effects models [66][67]: these models consider random effects between individuals, which is particularly valuable when considering genetic family data or other factors that can cause correlations.

In summary, the advantages of linear regression-based AD prediction models and

algorithms contain: 1) Interpretability: the parameters of linear regression models can be interpreted intuitively as the relationship between each feature and the response variable, which allows biological or medical implications to be derived. 2) Computational efficiency: linear regression is a relatively simple model that requires fewer computational resources and is faster to train. 3) Robustness: linear regression is relatively less susceptible to slight data variations, thus resulting in a more stable model. 4) Model simplicity: Compared to neural networks or complex ensemble methods, linear regression provides a concise model framework, making it easy to implement and validate. 5) Do not require big data: as opposed to models such as deep learning, linear regression models do not require a large amount of data to obtain reasonable results. Disadvantages contain: 1) Linear assumptions: linear regression assumes a linear relationship between the dependent and independent variables, but the reality of AD progression can be more complex than the model, and there can be non-linear relationships or interaction effects. 2) High-dimensional data: e.g., MRI images and genetic data can provide a large number of features, leading to dimensional catastrophe. 3) Unbalanced data: data from early AD can be sparse, leading to models that are biased towards those conditions that are more frequently encountered. 4) Possible overfitting problems: especially if the number of features is much larger than the number of samples, linear regression can overfit. While this can be mitigated by regularisation, there is a limited degree of overfitting that can be alleviated.

2.2.1.2 Support vector regression-based approaches

Support vector regression (SVR) is based on the concept of support vector machines. The objective of SVR is to find a hyperplane that allows most of the data points to fall within a margin of this hyperplane. The core is to maximise the distance from the hyperplane to the nearest data point while ensuring that most of the data fits within this margin. In nonlinear problems, SVR can implicitly map the input data to a higher-dimensional space by utilising a kernel function that allows it to become linearly separable in the new feature space. Mainstream kernel functions include linear kernels,

polynomial kernels, radial basis kernels and sigmoid kernels.

Current researches on SVR-based AD prediction models and algorithms focuses on: 1) Ensemble learning SVR [68][69]: researchers have explored approaches that combine SVR with various machine learning models to form ensemble systems, which integrate multiple models to merge their predictions to provide more accurate and robust predictions. 2) Domain adaptation and transfer learning [70][71]: there can be distributional differences for neuroimaging data from different sources or devices. Domain adaptation techniques aim to enable models to be trained on one domain (or dataset) and to make effective predictions on another domain. 3) Multi-modal data fusion SVR [72][73]: considering the complexity of Alzheimer's disease, researchers have been exploring approaches to fuse data from multiple sources (e.g., MRI, fMRI, cerebrospinal fluid markers, etc.). By fusing these data, models can provide a comprehensive understanding of the disease from multiple perspectives. 4) Variants of kernel techniques [74][75]: although radial basis function kernels are the prevalent ones, researches have explored a variety of novel kernel functions such as polynomial, sigmoid and Laplace kernels, with combinations and variants of these kernel functions. 5) Graph-based SVR [76][77]: incorporating graph data of brain structure and function into SVR. 6) Multi-kernel learning SVR [78][79]: due to the complexity of AD and the multi-modal nature of the data, the model can simultaneously utilise multiple kernel functions, each corresponding to a different data type or feature, and integrate different sources of information through weight allocation.

In summary, the advantages of SVR-based AD prediction models and algorithms contain: 1) Non-linear modelling capability: SVR has a high degree of non-linear modelling capability through various kernel functions (e.g., radial basis kernel, polynomial kernel, etc.), which allows it to process complex biomedical datasets that may not conform to simple linear assumptions. 2) Robustness: SVR is robust on datasets in the presence of noise, since its optimisation strategy is based on structural

risk minimisation, aiming to obtain a high generalisation degree to unseen data. 3) Sparsity: the solutions of SVR are typically sparse, implying that only a portion of the training samples (i.e., the support vectors) are utilised to determine the decision boundaries or regression curves. It helps to reduce the computational complexity and storage requirements of the model. 4) Multi-scale learning: SVR can incorporate different kernel functions (such as radial basis functions, polynomial kernels, etc.) to process multi-scale data. Disadvantages contain: 1) Computational complexity: SVR training can be time-consuming for large datasets, as it requires solving a quadratic optimisation problem involving the entire dataset. 2) Lack of interpretability: although support vectors provide a certain degree of interpretability, SVR can be considered as a black-box model when complex nonlinear kernel functions are utilised, which makes it difficult to interpret its internal working mechanism. 3) Kernel function selection: choosing the appropriate kernel function can be a challenge, as it greatly affects the performance of the model. No single kernel function is suitable for all applications.

2.2.1.3 Regression tree-based approaches

The regression tree is a type of decision tree that is utilised to solve regression problems. In contrast to the classification tree (which outputs discrete labels), the objective of a regression tree is to predict a continuous output value. A regression tree is a process of recursively partitioning a dataset into subsets, where each subset corresponds to a leaf node of the decision tree. The model starts at the root node and splits the data into different sub-nodes based on specific decision criteria. These decisions are based on features and thresholds that minimise the mean square error of the output. When certain stopping conditions are met (e.g., a predetermined node depth is reached), the node is no longer split and becomes a leaf node. At each leaf node, the predicted output value is the average of the target values of all data points within that subset. By considering nonlinearities, interaction effects and combining ensemble methods, its prediction performance can be further improved to provide solid data support for early diagnosis

and treatment of AD.

Current researches on regression tree-based AD prediction models and algorithms are focused on: 1) Application and research on random forests [80][81][82]: random forests are combinatorial models based on multiple decision trees, which enhance the generalisation ability of the model by random sampling of features and data. For AD, researchers have utilised random forest models to integrate multi-modal data (such as MRI images and cerebrospinal fluid biochemical indicators) to improve prediction accuracy. 2) Feature fusion and tree splitting [83][84]: researches have considered feature fusion in regression tree splitting for AD. It means that the model may consider the combination of multiple features or their interaction effects to determine the split, instead of splitting based on a single threshold of one feature. 3) Deep forests [85][86]: a combination of deep learning and decision trees that performs feature transformation and representation learning through multi-layer forests to enhance model performance.

In summary, the advantages of regression tree-based AD prediction models and algorithms contain: 1) Interpretability: regression trees provide explicit decision paths, which assist clinicians and researchers in understanding which factors have the most influence on prediction and provide clear advice to patients. 2) Processes non-linear relationships: there is no requirement to have prior knowledge of the relationships between data, as regression trees can automatically capture non-linearities and higher-order interactions. 3) Effective model visualisation: compared to some complex machine learning algorithms, regression trees provide decision paths that can be visualised, assisting clinicians and patients in understanding the model decision basis. Disadvantages contain: 1) Instability: small data changes can result in the generation of completely different trees. 2) Requires large amounts of data: AD involves multiple complex biological processes and constructing an accurate regression tree model can require a large number of samples. 3) Processing of high-dimensional data: if the number of features is enormous (e.g., gene expression data), the regression tree can

encounter the curse of dimensionality, leading to a decrease in model performance. 4) Variable correlation: if multiple variables are highly correlated, the regression tree can only select one of them and ignore other correlated variables, which can result in misleading interpretations. 5) Processing of continuous outputs: regression trees can be imprecise in splitting continuous output values, especially when the data range is wide.

2.2.1.4 Neural network-based approaches

Neural networks are based on artificial neurons that mimic biological neurons. Each neuron receives multiple inputs, which are aggregated by weight parameters and output to the next layer through an activation function and deep learning is the multi-layer neural network. It is trained with data to optimise the weights and perform tasks such as regression, classification, recognition etc. Neural networks and deep learning techniques have made significant progress in medical image analysis and biomarker prediction. The mainstream applications are processing medical image data (such as MRIs and PETs) by learning image features that enable the identification of abnormal patterns and image markers associated with AD, such as brain atrophy, brain plaque accumulation, etc. Processing time series data, such as cognitive decline data and brain function data over time (e.g., fMRI). Processing long sequence data, such as gene sequences or long-term medical records.

Current researches on neural network and deep learning-based AD prediction models and algorithms are focussed on: 1) Multi-modal fusion [87][88][89][90]: multi-modal data (e.g., MRI, PET, CSF biomarkers, and neuropsychological assessment data) are being integrated to present a more comprehensive view of the patient. Deep learning models (especially deep fusion strategies) can automatically capture correlations between various data types. 2) 3D convolutional neural networks (CNN) [91][92][93]: these are designed to process 3D brain image data. Compared to 2D CNNs, 3D CNNs can better capture the 3D spatial characteristics of brain structures and can be utilised to identify brain variations associated with AD. 3) Graph neural networks (GNNs)

[94][95][96]: the topology and functional connectivity data of brain networks can be analysed by GNNs, which can capture complex patterns at nodes (brain regions) and edges (connections), providing novel information and knowledge for AD diagnosis. 4) Interpretable deep learning [97][98][99]: in order to make deep learning models more reliable for medical applications, researchers are increasingly focusing on the interpretability of the models. For example, it is feasible to visualise the decision-making process of a model with techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping) or SHAP (SHapley Additive exPlanations), which can reveal key brain regions that influence diagnosis. 5) Attention-driven models [100][101][102]: with the attention mechanism based on transformer, models can concentrate on essential parts of the brain image or sequence data, thus enhancing the prediction accuracy. This approach performs especially well with multi-modal data fusion, as it automatically weighs the importance of various data sources. 6) Continual learning for longitudinal data [103][104]: these models can be updated as new data becomes available, especially for tracking longitudinal data of patients. 7) Reinforcement learning for treatment strategies [105][106][107]: although primarily applied to the decision-making process, reinforcement learning can be utilised to identify the optimal diagnostic pathway or treatment strategy, especially with limited data resources. 8) Model distillation [108][109]: since computational resources on medical devices can be limited, model distillation can transfer knowledge from large deep learning models to smaller models, thus reducing computational requirements while maintaining accuracy. 9) Few-shot and zero-shot learning [110][111][112]: since high-quality medical labelled data can be scarce, models can be trained on minimal data and still achieve good generalisation with few-shot or zero-shot learning approaches.

In summary, the advantages of neural network and deep learning based AD prediction models and algorithms contain: 1) Processing of high-dimensional data: e.g. 3D MRI scans, deep learning models can process high-dimensional data and capture complex spatial patterns. 2) Multi-modal data integration: deep learning models are can integrate

information and knowledge from different data sources (e.g., MRI, PET, genetic data, etc.). 3) Model complexity and flexibility: the depth and structure of neural network models can be adapted as required, enabling the capture of more complex patterns and correlations. 4) End-to-end learning: models can learn directly from raw data to output results without intermediate steps or human intervention. Disadvantages contain: 1) Lack of interpretability: while certain techniques exist to explain the decisions of deep learning models, they are considered black boxes models and cannot be as easily interpreted as other methods. 2) Large amount of data required: deep learning models generally require large amounts of labelled data to avoid overfitting. 3) Hardware dependency: although deep learning hardware has made great strides, the training and deployment of large models still requires advanced graphics processing units (GPUs) or other specialised hardware, which can raise costs and complexity. 4) Long training times: especially for extremely large and deep models, training can require long periods of time despite the availability of high-performance hardware.

2.2.2 Regularisation

Regularisation techniques are a common method utilised in machine learning to help limit the complexity of a model to avoid overfitting problems. When fitting data with a complex model, it is easy to overfit (good performance in the training set and poor performance in the test set), which can lead to a reduction in the generalisation ability of the model. Regularisation techniques can also be widely utilised in AD prediction algorithms. Specifically, regularisation methods usually introduce a penalty term to limit the range of values of certain parameters in the model, thus making the model more robust. In addition, regularisation techniques can be utilised for feature selection, i.e., selecting the most influential features for AD prediction. In AD prediction algorithms, there can be a large number of irrelevant features that can affect the accuracy of the model. By incorporating regularisation, features can be filtered by limiting their importance, thus reducing the impact of irrelevant features on the model.

Commonly utilised regularisation methods contain $\ell 1$ regularisation (lasso) and $\ell 2$ regularisation (ridge). $\ell 1$ regularisation realizes regularisation by adding the sum of the absolute values of all features to the original objective function. The $\ell 2$ regularisation realizes regularisation by adding the sum of squares of all features to the original objective function. Both limit the parameter size by adding a sum term, but they have different effects, $\ell 1$ regularisation is more appropriate for feature selection, while $\ell 2$ regularisation is more appropriate for forestalling model overfitting. These two approaches and their combinations and variants are frequently applied to AD prediction algorithms. For instance, [6][11][62][113] applied $\ell 2,1$-norm to capture task-specific features for different time points and a small number of features for all tasks. In [9][114], presented approaches use inter-vector correlation among regression coefficient vectors and intra-block correlation in each regression coefficient vector. [12] [60] proposed regularisation of three relations ('response-response' relation, 'feature-feature' relation and 'sample-sample' relation) to improve the performance of model.

Overall, regularisation is an important technique in Alzheimer's disease prediction algorithms, which can effectively improve the generalisation performance of models, reduce the number of features, accelerate the speed of prediction, and improve the interpretability of models. Therefore, regularisation is regarded as an important component in Alzheimer's disease prediction algorithms.

2.2.3 Multi-task learning

Multi-task learning is a machine learning method that aims to improve prediction performance by learning multiple related tasks simultaneously. Multi-task learning has been widely utilised in the prediction of disease progression in a variety of diseases. In AD prediction research, multi-task learning algorithms have been used to predict multiple tasks such as brain images, behavioural scores and diagnoses. The algorithms are utilised to obtain more accurate predictions by utilising multi-modal data, such as

images, speech and behavioural scores.

Multi-task learning was first proposed in 1997. It extracts the relationship between multiple similar tasks by establishing a statistical model, which proves that multi-task learning can improve the performance of the model [115]. Since then, it has attracted widespread attention in various algorithms, with applications ranging from bioinformatics to finance [116][117]. This new research direction has brought encouraging performance improvements in various fields, including but not limited to development of probability and statistical models [118][119], multi-task learning utilising kernel methods [120], feature selection [12][121][122], explaining task relationship [123][124][125] and feature hashing [126].

In the ML regression algorithm, the usual focus is on optimizing specific metrics. In order to accomplish it, a model or a combination of models is usually trained to implement the target task. Then, fine-tune models until the model results cannot continue to be optimised. Although this method can usually be utilised to achieve acceptable performance, because the focus is on a single task, this approach cannot take into consideration other information that can help optimise the metrics. Multi-task learning can make the model better summarize the original task by sharing the representation between related tasks.

For AD, the prediction problem can be considered as a multi-task regression problem [11], where the number of subjects is small and the number of input features (e.g., AD biomarkers) is large, the traditional assumption is to utilise a linked analysis of the tasks from different time points is assumed to improve performance. Compared with single-task learning, multi-task learning utilises regression models to predict the patients' future state at different multiple time points. The essential premise of models is that there is an intrinsic correlation between various subject information records and capturing the intrinsic correlation can enhance prediction model generalisation.

Multi-task learning attempts to jointly learn various related tasks to ensure that the knowledge contained in one task can be applied by other tasks, ultimately improving the generalisation performance of all tasks [127][128]. MTL technology is extensively implemented in the biomedical engineering field, for our research case AD, MTL provides a wide range of applications in numerous domains. In terms of feature selection approach, [61][72][129][130] developed multi-modal multi-task learning methods that choose the equivalent subset of numerous variable-related features from each modality and concurrently predicts multiple variables from multi-modal data. [131][132] presented deep learning network-based multi-task learning AD prediction algorithms that introduce multi-task feature selection techniques, evaluate the internal connection between several related tasks and choose feature sets relevant to all tasks. In terms of feature learning approach, existing approaches have focused on modelling task interactions with novel regularisation techniques [133][134][135][136][137][138]. And kernel approaches were combined with the methodology to support non-linear relationships [6][61][139]. In terms of low-rank approach, [140][141][142] presented robust multi-task learning methods that employ low-rank structures to preserve task connections while recognizing anomalous tasks utilising group sparse structures. In contrast to the above approach, we assumed that knowledge sharing between prediction tasks for different patients is expected to improve achievable performance, and therefore we set up prediction task for a single patient as one task, which is a small-scale manner of task setting.

2.3 Multi-dimensional tensor-based data

All of the aforementioned classic AD prediction approaches utilise a second-order matrix to represent input features with patient and biomarker feature dimensions. In contrast to the second-order matrix, which has two components for each index, the third-order tensor has three components for each index. The third-order tensor is like a

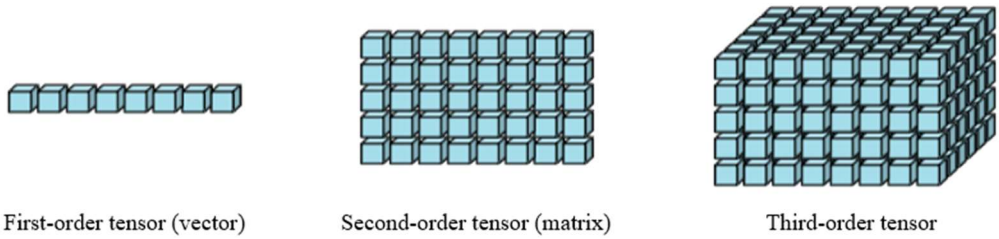


Figure 2.1: Vector, matrix and tensor.

cube matrix composed of multi-dimensional arrays. There will be more knowledge and information in various dimensions as there are more components. Applying multi-dimensional tensor to ML regression algorithms can integrate multi-dimensional knowledge and information to improve prediction accuracy and feature understanding. This section will review the techniques and applications of tensors, tensor decomposition techniques and latent factors.

2.3.1 Multi-dimensional Tensor and its application

Tensor is a general tool of multiple linear algebra, and matrix is a specific linear algebra tool, tensors have a larger data coverage than matrices. The matrix is a two-dimensional array, in which each element is determined by two indexes and a tensor is an array with more than two dimensions. In fact, a scalar is a zero-dimensional tensor, a vector is a one-dimensional tensor, and a matrix is a two-dimensional tensor. Figure 2.1 shows the difference in shape of tensors of different orders.

Tensor technology is widely utilised in the fields of spatio-temporal data analysis, neural image analysis [143], signal processing [144] and machine learning [145][146][147]. The application of tensor technology in the field of spatio-temporal data analysis mainly includes spatio-temporal data mining, analysis and modelling. It can help to process, analyse and mine complex spatio-temporal data to draw valuable findings [148][149][150][151][152][153][154]. Specifically, 1) The multi-dimensional structure of tensor allows for the simultaneous processing of multi-dimensional spatio-temporal data, thereby obtaining more comprehensive information. 2) Tensor

technology supports a variety of data analysis approaches, allowing for the selection of suitable analysis methods and efficient data analysis according to actual application requirements, thereby reducing computation temporal and spatial complexity. 3) Tensor technology can enhance the interpretability of data analysis and generate interpretable analysis results, thus facilitating the understanding and application of data analysis results.

Neural image analysis is an important task of neuroscience, and neural images can be regarded as tensor data. For example, electroencephalogram (EEG) is a second-order tensor; MRI is a third-order tensor; functional magnetic resonance image (fMRI) is a fourth-order tensor. In order to study the relationship between neural images and clinical results, and to comprehend the inner working mechanism of the brain, [155] used the matrix Logistic model to classify the electroencephalogram, [156][157] used Tucker tensor regression model and CP tensor regression model to explore the relationship between MRI and fMRI and clinical results for attention deficit hyperactivity disorder (ADHD). Tensor techniques have important applications and advantages in the field of signal processing. Firstly, tensors can represent multi-dimensional signals, which facilitates better capture of complex structures and relationships in signals [158][159][160]. For example, in image processing, tensors can be used to represent features such as colour and brightness of an image, while in speech recognition, tensors can be used to represent the frequency and time domain features of an audio signal. Secondly, tensor techniques can be applied in signal processing in a variety of mathematical approaches, which can help to extract important features and relationships in the signal [161][162][163][164]. For example, tensor decomposition can be used to extract features in images, while tensor-dense networks can be used to recognize speech signals.

In the machine learning prediction model, incorporating tensor data into the framework of regression analysis not only expands the methods and techniques of tensor data

analysis, but also further enriches theoretical results of regression analysis. Traditional linear regression models can be extended to tensor regression models to process high dimensional data. These models can capture features in multiple dimensions and provide greater accuracy in prediction [165][166][167][168]. Neural networks can be expanded into tensor regression networks that process tensor data. These network structures can learn complex associations of multi-dimensional features and can provide powerful predictive capabilities [169][170][171]. The tensor kernel approach combines the concepts of tensor analysis and kernel methods for dealing with non-linear relationships. It can capture complex relationships and improve the performance of regression analysis by mapping tensor data to a high-dimensional feature space [172][173][174]. In applications such as image processing, text mining and recommender systems, where target variables or features can have a non-negativity constraint, non-negative tensor regression considers this constraint and develops corresponding models and algorithms that are essential for the interpretation and analysis of the data [175][176][177][178]. Interpretation is a critical requirement for decision making and prediction results in regression analysis. Tensor regression approaches can provide understanding and interpretation of regression models and output results in a variety of dimensions. It assists researchers in understanding the rationale and key features underlying the model and enhances trust and acceptance of the results of regression analysis [179][180][181].

Overall, the utilisation of multi-dimensional tensors in machine learning has numerous advantages and is widely utilised in various fields. 1) Model construction: multi-dimensional tensors can be utilised as input and output to machine learning models, thus facilitating the construction of complex models. Multi-dimensional tensor can perform complex mathematical calculations in different machine learning algorithms, such as matrix multiplication, tensor multiplication, etc. These calculations help to learn complex relationships of data in a multi-dimensional space. 2) Data correlations: multi-dimensional tensor can represent complex data correlations, such as the correlation

between multiple features, which is extremely useful in solving practical problems. 3) Data representation: multi-dimensional tensor can represent complex data, such as images, speech and video, as a structured data type for convenient storage and processing. 4) Feature extraction: multi-dimensional tensor can be utilised to extract features from data, for example using convolutional neural networks to extract features from images in image recognition. 5) Computational efficiency: mathematical operations of multi-dimensional tensor can be implemented using matrix operations, improving computational efficiency.

2.3.2 Tensor decomposition and Latent factor

Tensor decomposition is essentially a high-order generalisation of matrix decomposition. Matrix decomposition has three obvious uses, namely, implicit relationship mining, missing data filling (or "sparse data filling") and dimensionality reduction processing. Tensor decomposition can also satisfy these uses. In the prediction algorithm, the hidden relationship mining is extremely important. It can discover the hidden information in the data to optimise the performance of the algorithm. Tensor decomposition can assist algorithms to learn data better by decomposing factors in a tensor to capture the relationships between different dimensions. Tensor decomposition techniques include different approaches such as non-negative tensor decomposition (NMF), latent semantic decomposition (LSA), etc., which are applicable to different data types and task scenarios [182][183][184][185][186]. The major techniques for tensor decomposition are Tucker and CP tensor decomposition, which are discussed in detail in this section.

The latent factor obtained by tensor decomposition can facilitate a better understanding of complex, and these latent factors can represent certain intrinsic patterns and associations of the data in multiple dimensions. They reflect the hidden structure and internal relationships of the original data. Latent factors can be utilised for a variety of

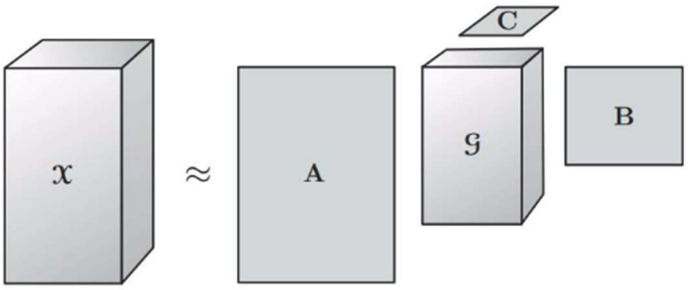


Figure 2.2: Tucker tensor decomposition.

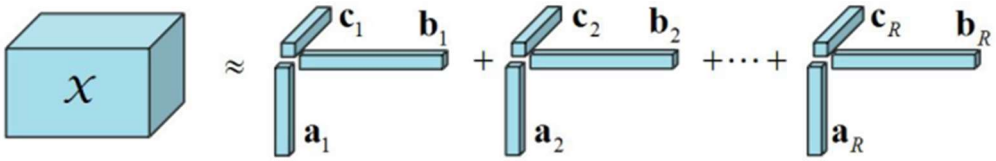


Figure 2.3: CP tensor decomposition.

tasks, such as data visualisation, dimensionality reduction, clustering, regression, classification, etc. [187][188][189][190]. They can extract meaningful information from high-dimensional data and improve the performance of prediction algorithms. In addition, latent factors can be utilised to explore potential characteristics of the original data in order to better understand it. They can be used in numerous areas of data analysis, such as bioinformatics, information retrieval, recommender systems, image processing, etc. [191][192][193][194].

Tucker decomposition and CANDECOMP/PARAFAC (CP) decomposition are the major techniques for tensor decomposition, tucker decomposition is a high-level principal component analysis (PCA), which expresses a tensor as a core tensor and multiplies each mode by a matrix (Figure 2.2). i.e., $X \approx [\![G \times \mathrm{A} \times \mathrm{B} \times \mathrm{C}]\!] = \sum_{m=1}^{r_1}\sum_{n=1}^{r_2}\sum_{l=1}^{r_3}(g_{mnl}a_m \circ b_n \circ c_l)$, where $\circ$ denote the outer product operation between two vectors. The size of tensor *G* is $r_1 \times r_2 \times r_3$, also called core tensor, the size of matrix A is $m \times r_1$, the size of matrix B is $n \times r_2$, the size of matrix C is $l \times r_3$. Although it provides a more comprehensive representation, it is difficult to interpret latent factors since the amount of latent factors can be different for each mode [195]. Tucker tensor decomposition captures the structure of complex tensors in

multiple dimensions and therefore works effectively with complex data types.

The tensor form of CP decomposition is to represent a tensor as the sum of a limited number of rank-one tensors (Figure 2.3). i.e., $X \approx [\![\mathrm{A} \times \mathrm{B} \times \mathrm{C}]\!] = \sum_{i=1}^{r} \mathrm{a}_i \circ \mathrm{b}_i \circ \mathrm{c}_i$, where $\circ$ denote the outer product operation between two vectors, and A, B and C correspond to the vectors associated with the *r*-th latent factor. Given a tensor *X* of the size $n_1 \times n_2 \times n_3$, the size of matrix A, B and C is $n_1 \times r, n_2 \times r$ and $n_3 \times r$ respectively. CP tensor decomposition has a lower parameter complexity and is faster than tucker tensor decomposition in the case of large data. In addition, it is robust to the number of features in the tensor, making it suitable for the analysis of noisy data. It is highly sparse, which makes it effective when dealing with sparse data. In addition, it can capture the principal components of the data, which is especially useful for tasks such as feature selection and data visualisation.

To sum up, both Tucker and CP tensor decomposition are extremely flexible and can be adapted to the characteristics of the data and the purpose of the analysis in order to obtain the best results. Common advantages are as follows: 1) Scalability: both Tucker and CP tensor decomposition can be easily extended to higher dimensional tensors, which allows them to handle complex data structures. 2) Interpretability: both Tucker and CP tensor decomposition are relatively easy to interpret, making it easy to understand the characteristics and distribution of the data. 3) Parallelism: both Tucker and CP tensor decomposition can be computed in parallel utilising distributed systems, which greatly improves computational efficiency. 4) Algorithm stability: both Tucker and CP tensor decomposition are relatively stable algorithms, which do not suffer from convergence 5) Support for non-linear analysis: both Tucker and CP tensor decomposition can support non-linear analysis, which allows them to be applied to a variety of complex data types. 6) Data visualisation: the results of Tucker and CP tensor decomposition can be easily visualised, which helps users to understand the distribution and characteristics of the data. 7) Algorithm practicality: both Tucker and CP tensor

decomposition are widely utilised algorithms with good practical results in various fields, such as bioinformatics, computer vision, natural language processing, etc. 8) Solving complex problems: Tucker and CP tensor decomposition can solve complex problems in multi-dimensional data analysis and can effectively extract important features of the data. 9) Algorithm universality: Tucker and CP tensor decomposition are universally applicable algorithms that can be utilised on a diverse variety of data types and data sources.

2.4 Research and application of AD brain structural variation

Researches in the field of brain science have proven that the brain structural variants of AD are closely related to the progression of the disease. In AD patients, there are two characteristic variants in the brain: a reduction in neurons and a lack of connectivity; and a build-up of neuropathic proteins such as beta-amyloid. These mutations lead to a progressive loss of brain function. Investigating brain structural variation in Alzheimer's disease provides opportunities for prevention and early intervention as well. For example, through screening and monitoring of patients at high risk, brain structural variations can be detected in advance and preventive measures can be taken to reduce the progression of the disease. By researching brain structural variation in AD, we can better understand the function of the human brain and the mechanisms of disease progression, which can contribute to the research and treatment of other neurodegenerative diseases.

In machine learning algorithms, data from the research on brain structural variation in AD can be utilised as essential input for training models, thus improving the accuracy and efficiency of the models and allowing the development of novel diagnostic models based on brain structural variation to diagnose AD quickly and accurately. By investigating brain structural variation in AD, we can obtain a large amount of data on brain structural variation that can contribute to ML algorithms to accurately identify

and predict the progression of AD.

2.4.1 The differences in brain variation structural between CN, MCI and AD

CN, MCI and AD are states with major brain structural variations, each of which has specific brain structural variations that can be identified during the diagnostic process utilising several techniques such as MRI and cognitive assessment tests. MRI can reveal the internal structure and function of the brain and is particularly beneficial in the diagnosis of MCI and AD. Cognitive assessment tests can assess memory, attention and other cognitive functions to facilitate the diagnosis of CN, MCI and AD. The difference between CN, MCI and AD brain structural variations is the degree to which each stage affects the brain differently. In CN, the brain structure remains healthy and functions normally. In MCI, cognitive function has begun to diminish, but the individual can still live independently. In AD, the brain structure has been severely affected, resulting in severe cognitive impairment and inability to live independently. In the CN state, the connectivity network in the brain remains normal and there is sufficient neuronal activity to maintain normal cognitive function. On the other hand, in MCI and AD, the connectivity network gradually breaks down and neuronal activity is reduced, leading to brain dysfunction.

Numerous previous researches in the brain science field have concentrated on the distinctions between the brain structures of AD, MCI (mild cognitive impairment) and CN (cognitively normal older individuals). [196][197][198][199] designed various frameworks for simulating the effects of AD and ageing on the evolution of the brain's morphology, focusing on certain morphological alterations in the brain that could be used to pinpoint clinical disorders. [200][201][202][203] examined the connection between PET, CSF, MRI and fMRI biomarkers and clinical diagnosis and cognitive performance in individuals with CN, MCI and AD. In result, cross-sectional integrated

cognitive and functional skills were better connected with cross-sectional MRI, which offered stronger cross-sectional grouping and discriminating. [204][205][206][207] assessed cortical thickness in CN, AD and MCI patients. As the disease progressed from MCI to AD, patterns of cortical thinning were detected as a consequence of the malignant transformation, and it was found that the entire cortex thinned and expanded noticeably into the lateral temporal cortex.

To sum up, applications of brain structural variation information in Alzheimer's disease prediction algorithms has the following advantages: 1) Early diagnosis: the utilisation of brain structural variation information can facilitate early diagnosis of AD and provide earlier intervention for patients. 2) Multi-dimensional features: brain structural variation information has multi-dimensional features, allowing for a more comprehensive assessment of the patient's condition and increasing flexibility of the prediction algorithm. 3) Reliability: brain structural variation information is based on imaging techniques, providing reliable evidence and enhancing the reliability of the prediction results. variation information is obtained based on imaging techniques, which can provide reliable evidence and enhance the reliability of prediction results. 4) Simplicity: obtaining information on brain structural variation requires the utilisation of imaging techniques such as MRI or computed tomography, which are medically mature and therefore easy to obtain information on brain structural variation. 5) Comprehensive: information on structural brain variation is comprehensive and can consider variations in various areas of the brain as well as the impact of multiple factors within the brain on Alzheimer's disease.

2.4.2 Correlations between AD biomarkers.

Correlations between AD biomarkers represent the relative variations in these biomarkers in individuals with AD. Their correlations are important for the diagnosis and monitoring of AD as they can assist doctors in assessing the progression of the

disease more accurately. In addition, the correlation of biomarkers can be utilised as a powerful tool in the research of AD, helping scientists to understand the mechanisms and trends in the progression of the disease. By analysing the correlation of biomarkers, clinicians can evaluate patients' disease more accurately and develop more effective treatment plans and timely adjustments to treatment protocols. In addition, the correlation of biomarkers can serve as important indicators to assess the efficacy of medicines and help evaluate whether the medicines are having a positive impact on patients.

Researches on brain structural variation has concentrated on the correlations between AD MRI biomarkers. [208][209][210][211] enhanced the model performances for distinguish of AD and its precursor phases by correlating regional mean cortical thickness to incorporate pertinent information with ROI-based data. [212][213][214][215] used graph theory and correlation matrix to analyse and organise brain networks for various regions. The researches mentioned above assessed and examined the correlation between the Alzheimer's disease progression and brain biomarkers, it revealed that there are differences in the brain structure for CN, MCI and AD. However, aforementioned researches only concentrate on a single biomarker or the same category of biomarkers, lacking the connection and correlation of spatio-temporal variation between distinct categories of biomarkers, which is crucial for Alzheimer's disease feature representation.

To sum up, correlation between multiple biomarkers can improve predictive accuracy since they collectively reflect the biological mechanisms of the disease. In addition, correlation between biomarkers can improve predictive stability, as individual biomarkers can have a certain level of uncertainty, but multiple biomarkers combined can be more robust. In addition, correlation between biomarkers provides greater sensitivity and specificity in AD prediction algorithms. It implies that by assessing biomarker correlations, patients with AD can be identified more effectively and

misdiagnosis of healthy populations can be reduced. Finally, correlations between biomarkers can provide information on the biology of AD and contribute to the identification and treatment of new targets for AD.

2.5 Quantification of biomarker correlations

The quantification of biomarker correlations can be quantified and calculated utilising a variety of methods, specifically, 1) Correlation coefficient analysis: by calculating the correlation coefficient between two variables to assess their correlation (e.g., Spearman's rank correlation coefficient, Pearson's correlation coefficient, etc.) [216][217]. 2) Discriminant analysis: by identifying the differences between two or more biomarkers using linear discriminant analysis or non-linear discriminant analysis [218][219]. 3) Multivariate statistical analysis: to assess the relationship between multiple biomarkers by using principal component analysis, cluster analysis, factor analysis, etc. [220][221]. 4) Bioinformatics methods: to assess the correlation between biomarkers by using bioinformatics techniques such as proteomics, genomics, cluster analysis, genetic algorithms, etc. [222][223]. 5) Network analysis: to assess the correlation between biomarkers by studying the interrelationship between biomarkers and building a network of correlations between biomarkers [224][225]. 6) Molecular biology: to assess the correlation between biomarkers by using molecular biology methods such as polymerase chain reaction (PCR), deoxyribonucleic acid (DNA) sequencing [226][227]. 7) Fractal analysis: to assess the correlation between biomarkers by examining the complexity of biomarkers [228][229]. 8) Statistical regression analysis: to assess the predictive value of biomarkers by identifying the correlations between biomarkers and disease occurrence [230][231].

However, all the above quantification and calculation approaches require large and continuous amounts of patient biomarker data, and data from AD patients are unstable in terms of detection time, the data are discrete and sparse, and this is not a problem

only for AD, but for most neurological disorders, as it is difficult to require patients to go to the hospital (or research facility) for frequent tests. Therefore, none of the above quantification and computational approaches can be applied in neurological disorders, therefore we need to investigate a novel quantification approach to quantify the correlation between AD biomarkers. We propose similarity calculation-based quantification approach, which quantify biomarkers as vectors with multi-dimensional information, and then utilise similarity calculations to quantify biomarker correlations. Relevant to this thesis, there are three similarity calculation approaches that have been extensively utilised in the literature and that are utilised in this thesis. They are described in detail in Appendix A Mainstream similarity calculation approaches.

2.6 Summary

This chapter reviews the research fields that have impacted the research. Section 2.1 presented a clinical overview for AD and section 2.2 discusses the major and state-of-the-art researches on various AD prediction algorithms, a review of multi-task learning algorithms and applications of multi-task learning algorithms. It discusses limitations of existing and state-of-the-art prediction algorithms in medical small data set scenarios and the reasons that multi-task learning approaches can be utilised to enhance the performance of algorithms and models in such scenarios. The reviews in Sections 2.3 and 2.4 show that multi-dimensional data, multi-dimensional tensor and tensor decomposition techniques have been a hot topic for research and applications in various fields, especially in machine learning and bioinformatics. However, to date no research has been able to effectively provide highly accurate, stable and interpretable algorithms for predicting disease progression based on multi-dimensional tensor. The main difficulty is that data from patients with neurological disorders such as AD are unstable in terms of detection time, the data can be discrete and sparse, as it is difficult to get patients to go to hospital (or research facilities) for frequent tests. Since this problem combined with the reviews in Section 2.5 indicate that mainstream correlation

quantification approaches are unable to quantify the spatio-temporal and biomarker correlation information in AD biomarker data as a multi-dimensional tensor, we propose to utilise similarity calculation-based approaches to quantify the multi-dimensional information of biomarkers, and discuss the mainstream similarity calculation approaches in Section 2.5. The next chapter will elaborate on the methodology to accomplish the aim of this research.

Chapter 3

Methodology

In order to investigate how to accomplish the aim of this research which is the quantification and construction of AD biomarker data as a higher order tensor, combining multi-task learning and machine learning regression approaches to construct algorithms with multi-dimensional spatio-temporal characteristics of AD biomarker data to predict AD progression at multiple future time points. It is first necessary to design and construct the quantification approach for the AD biomarker higher-order tensor, and then to design and construct multi-task learning algorithms and models to integrate and mine the spatio-temporal variation information and knowledge of MRI data. In this chapter, the idea, design and composition of the proposed method are first presented, followed by the analysis and definition of the design philosophy of each composition. Then the dataset utilised and the design of the experimental approach are presented. Finally, the concept and process of investigating the interpretability of algorithm and experimental results is presented.

3.1 Research Design

For the aim presented in Section 1.3, the research has been divided into two parts. The first part investigates the quantification and construction of AD biomarker data as a higher order tensor. For this part, we proposed a novel similarity-based quantification approach that considers both the magnitude and direction correlations of structural variation between brain biomarkers and encodes the quantified data as a third order tensor to address the problem of monotonic data form and the associated loss of hidden

information. The design goals, fundamentals and concepts are detailed in Section 3.2.

The second part investigates the combination of quantified multi-dimensional tensors with multi-task learning and machine learning regression approaches to construct algorithms with multi-dimensional spatio-temporal characteristics of AD biomarker data to predict disease progression in AD patients at multiple future time points. For this part, we proposed a novel tensor multi-task learning algorithm, where each patient prediction is regarded as one task, each task shares a set of latent factors obtained by tensor decomposition techniques. Knowledge sharing between tasks improves model generalisation and addresses scarcity problem of medical data, and its design goals, fundamentals and concepts are detailed in Section 3.3.

To demonstrate the effectiveness and performance of the proposed AD progression prediction approach, a series of experiments were designed in this research for comparison and validation. The datasets utilised and the design and procedures of the experiments are detailed in Section 3.4. In medical research, the interpretability of approach and results is as important as the performance of the approach. The concepts and process of interpretability research for proposed algorithm and experimental results are elaborated in Section 3.5.

3.2 Multi-dimensional tensor construction of biomarkers

3.2.1 Design goals

Considering the benchmarks for quantification and construction of multi-dimensional tensor listed in Chapter 1, this section discusses the goal of constructing a multi-dimensional tensor utilising spatio-temporal information and correlation information from MRI biomarkers.

Ability to mine and contain multi-dimensional spatio-temporal information and knowledge: The main goal of this quantification approach is to provide a viable solution for extracting multi-dimensional spatio-temporal information and knowledge of AD biomarkers.

Ability to mine and integrate biomarker correlation information: In order to address the problem of monotonic data formats and the resulting loss of hidden information, the proposed quantification approach must integrate biomarker correlation information in order to mine comprehensive AD biomarker knowledge.

Analysability and interpretability: The analysability and interpretability of the quantification approach means that the approach not only provides a practical and optimised utilisation solution to quantify and construct AD biomarkers as multi-dimensional tensor, but also to mine the information and knowledge of AD biomarkers in multiple dimensions, and to indicate and extract brain biomarker correlation and biomarker regions important for disease progression.

Feasibility and generalisability: The feasibility and generalisability of the quantification approach implies that this approach cannot be limited to specific types of disease or modalities. Different requirements for different diseases or modalities can be fully applicable.

3.2.2 Fundamentals

3.2.2.1 Dimensional selection of the multi-dimensional tensor

In disease progression prediction models, a multi-dimensional tensor can capture comprehensive biomarker information and knowledge, and selecting the right dimensions is important to properly analyse the data. The following are the main dimension selections in disease progression prediction models in addition to the biomarker dimension and the patient sample dimension: 1) Time dimension: time is

one of the commonest dimensions, as it allows researchers to track disease progression [232][233]. The time dimension can be further broken down into smaller time intervals (e.g., days, hours or minutes), which can provide more detailed information. 2) Spatial dimension: the spatial dimension allows us to compare disease progression between different locations or regions [234][235]. 3) Demographic dimension: we can consider the impact of factors such as the age, gender and ethnicity of the patient on the spread of the disease [236]. 4) Disease characteristics dimension: the disease itself has a number of characteristics, such as infectiousness, duration and severity [237]. These characteristics can be utilised as another dimension of the multi-dimensional tensor in order to analyse the impact of different characteristics on the progression of the disease. 5) Medical resources dimension: we can consider information on the number of hospitals in different regions, the number of health care workers and the distribution of medical equipment, which can influence the treatment and control of the disease [237]. 6) Environmental dimension: factors such as air pollution and climate change can make certain diseases more likely to spread [238][239]. Therefore, considering the environmental dimension can help us better understand the spread and control of disease. 7) Behavioural and social dimensions: factors such as the manner in which people interact, travel behaviour and personal hygiene practices can influence the spread of diseases [240][241]. Therefore, considering behavioural and social dimensions can help researchers to better understand the dynamics of disease transmission. 8) Health service utilisation dimension: different populations can utilise different types of health services, which can affect the diagnosis and treatment of diseases [237][242]. Therefore, considering the health service utilisation dimension can assist researchers to better understand spread and control of disease. 9) Health risk factor dimension: factors such as smoking, diet and physical activity can influence people's risk of developing a disease [243][244]. Therefore, considering the health risk factor dimension can help researchers to better understand occurrence, progression and control of disease.

However, none of above dimensional selections work optimally in predictive models of neurological disease progression such as AD. Since most neurological diseases have no cure and cannot be reversed, the key to predictive models of neurological disease progression is to utilise early data to detect and predict disease progression earlier in order to intervene and delay disease earlier. All the above dimensional selections require a large amount of data to construct a complete multi-dimensional tensor, but data from each AD patient is unstable at the time of detection, data can be discrete and sparse, and this is not just a problem for AD, but for most neurological diseases as it is difficult to get patients to go to the hospital (or research facility) frequently for tests. Therefore, we incorporate biomarker correlation knowledge into the multi-dimensional tensor construction, with the dimensions of tensor as the first biomarker dimension, second biomarker dimension and patient sample dimension.

3.2.2.2 Biomarker correlation quantification approaches

In disease progression prediction models, biomarker correlations can be analysed by a variety of quantitative approaches. The following are the main approaches to quantifying biomarker correlations in disease progression: 1) Correlation coefficient analysis: it is a measure of the degree of correlation between two variables, for example, the Pearson correlation coefficient and the Spearman correlation coefficient [245][246]. These methods can be utilised to analyse the relationship between biomarkers and disease progression, as well as the interrelationship between biomarkers. 2) Survival analysis: it is an analysis of temporal data that can be utilised to assess the relationship between biomarkers and disease progression. Survival analysis can use Kaplan-Meier curves to show the incidence of events and Cox regression models to assess the effect of variables on survival time [247][248]. 3) Factor analysis: it can be utilised to identify patterns between biomarkers and combine them into fewer variables or factors [249][250]. This approach can be utilised to determine which biomarkers provide the best explanation for variation in disease progression and use them as input to predictive models. 4) Network analysis: it is an analytical method used to explore complex

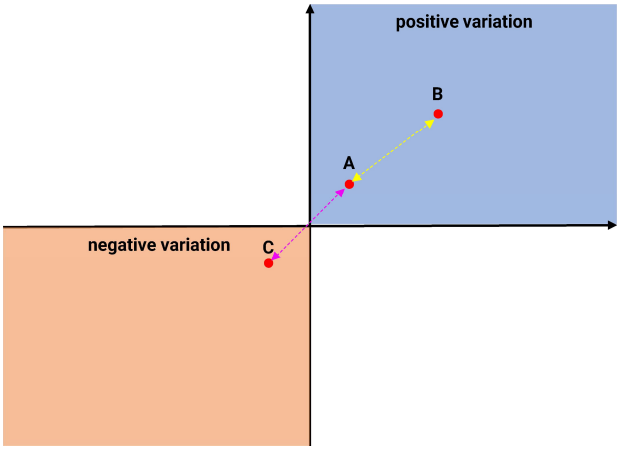


Figure 3.1: Calculate the similarity between two vectors to express the similarity of spatio-temporal variations of two MRI biomarkers.

systems and can be utilised to study the interactions between biomarkers [251][252]. This approach can reveal the topology of biomarker networks and which biomarkers in the network are important for disease progression.

However, none of the above quantification approaches work optimally in predictive models of neurological disease progression such as AD. All the above quantification approaches require a large amount of data to complete the quantification calculations, but the timing of data testing for each AD patient is erratic, the data can be discrete and sparse, and this is not just a problem for AD, but most neurological diseases have the same problem because it is difficult to get patients to go to the hospital (or research institutions) for tests. Therefore, we introduced multi-dimensional knowledge vector construction and similarity calculation into the quantification of biomarker correlations.

3.2.3 Design and construction approach

The mainstream and commonly utilised similarity calculation methods share a common blind spot, which is that the quantification and calculation process cannot be sensitive to both direction and value. Figure 3.1 shows a graphical representation of calculating the similarity between two vectors to express the similarity of spatio-temporal variation of two MRI biomarkers, from which it can be seen that the distances of biomarkers A

and B and A and C are the same. However, the direction of variation for biomarkers A and C is different while the direction of variation for A and B is the same, distance-based similarity calculation methods (e.g., Euclidean distance, Mahalanobis distance, etc.) can only sensitive to the value of the vector and their results indicate that the similarity of biomarker A and B is the same as the similarity of A and C. In contrast, for direction-sensitive similarity calculation methods (e.g., cosine similarity), they cannot be sensitive to the value of the vector.

In the research of brain structural variation, both the magnitude and direction of the biomarker variation are equally important for disease progression prediction and interpretability, therefore we need to design and construct a similarity-based quantification approach that can incorporate both the magnitude and direction information of the vectors.

3.3 Multi-dimensional tensor multi-task learning regression

3.3.1 Design goals

Considering the performance benchmarks for multi-task learning presented in Chapter 1, this section discusses the goals of algorithms and models that incorporate multi-task learning with multi-dimensional tensor of MRI biomarker spatio-temporal variation information.

Provide practical capabilities for predicting AD progression in medical small dataset scenarios: the main goal of the algorithms and models is to provide a feasible solution for AD progression prediction and modelling based on multi-task learning and multi-dimensional tensor.

Ability to improve the accuracy and stability of AD progression prediction:

Accuracy and stability are two key concerns in assessing the performance of AD progression prediction. Viable solutions proposed by algorithms and models based on multi-dimensional tensor and multi-task learning must provide higher accuracy and stability than traditional multi-task learning based AD prediction algorithms.

Analysability and interpretability: The analysability and interpretability of the algorithms, models and results means that the algorithms and models can not only provide practical optimised utilisation solutions to improve the accuracy and stability of Alzheimer's disease progression prediction, but can also indicate important multi-dimensional correlations between brain biomarkers. The integration of tensor decomposition and latent factor analysis allows the extraction of important brain biomarker regions.

Feasibility and generalisability: The feasibility and generalisability of algorithms and models means that they cannot be restricted to specific types of disease or modalities. Different requirements for different diseases or modalities can be fully applicable.

3.3.2 Fundamentals

3.3.2.1 Single task learning regression

Let $X = [x_1, x_2 \dots, x_n]^T \in \mathbb{R}^{n \times d}$ be the input data matrix, $W = [w_1, w_2 \dots, w_t] \in \mathbb{R}^{d \times t}$ be the weight parameter matrix, and $Y = [y_1, y_2 \dots, y_n]^T \in \mathbb{R}^{n \times t}$ be the target matrix with *n* patient samples, *t* time points and *d* biomarker features. The problem of predicting various clinical cognitive scores (e.g., ADAS-Cog and MMSE) at future multiple points can be expressed by solving different regression models:

$$\min_W \|XW - Y\|_F^2 + \theta_1 \|W\|_F^2 \tag{3.1}$$

where the first term measures the empirical error on the training data, $\theta_1 > 0$ is a regularisation parameter, and the penalty term $\|W\|_F^2$ controls the quantity for contraction of parameters and force the variance close to zero to reduce the mean square

error. The above regression method is the ridge regression, another solution is to utilise the lasso regression to minimize the constraints as follows:

$$\min_{W} \|XW - Y\|_F^2 + \theta_1 \|W\|_1 \tag{3.2}$$

For this formula, increasing θ_1 brings most of the W parameters associated with features considered unimportant to close to zero and shrinking non-zero coefficients at the same time. The difference between above two regression methods is that the square of the $\ell 2$ norm in ridge regression is utilised as a penalty term, and $\ell 1$ norm in lasso regression is utilised as a penalty term, which enhances the weight coefficients sparsity. Both the $\ell 1$ and $\ell 2$ penalties are utilised to control overfitting of machine learning models. The $\ell 1$ penalty is applied to model parameters by adding up absolute values of each parameter, i.e., the $\ell 1$ penalty is the sum of absolute values for model parameters. The $\ell 2$ penalty makes the model parameters as small as possible, but does not completely set them to zero. The choice of the $\ell 1$ and $\ell 2$ penalties depend on the nature of the particular problem and the characteristics of the dataset. When there are a large number of features, the $\ell 1$ penalty term typically produces better results because it allows for feature selection of the model parameters, while when feature correlation is high, the $\ell 2$ penalty term typically produces better results because it distributes the impact of the parameters as evenly as possible across each feature. The $\ell 1$ penalty term is more sensitive to outliers, as the presence of outliers can cause certain parameters to deviate from zero, thus affecting the performance of the model. In contrast, the $\ell 2$ penalty term is less sensitive to outliers as it takes all parameters into consideration and distributes the weights equally.

Single-task learning models are the basic machine learning models, but they have the following disadvantages: 1) Inability to handle multiple tasks simultaneously: single-task learning can only solve one task, but not multiple tasks at the same time. If multiple tasks need to be solved, the model needs to be trained separately for each task. 2) Lack of generalisation: single-task learning models are usually trained for a specific dataset and task; therefore, they can lack generalisation, i.e., performance degradation across

different datasets and tasks. This can require the adoption of more generalised models and algorithms to improve generalisation capabilities. 3) Inability to learn correlations between tasks: single-task learning models are typically trained for a single task and are unable to learn correlations between different tasks. 4) Difficulty in handling complex relationships: certain tasks can involve complex relationships between multiple inputs. Single-task learning models can have difficulties capturing these complex relationships, leading to performance degradation. 5) Limited amount of data: single-task learning requires a large amount of labelled data to train the model. If the data amount is limited, the model can over-fit or fail to capture key features in the data. 6) Not suitable for non-static environments: in dynamic environments, data distribution and features can vary. Single-task learning models cannot adapt themselves to new data distributions and features, resulting in reduced performance. 7) Sensitive to noise: single-task learning models can be overly sensitive to noise in the input data, leading to performance degradation. Noise can arise from incompleteness, inaccuracy or inconsistency of the data, etc. 8) Limited by model capability: single-task learning models can be limited by their capability and expressiveness, and thus unable to handle several complex tasks and data.

3.3.2.2 Definition of tasks and task relationships for multi-task learning

The key to designing and constructing multi-task learning algorithms and models is how to define tasks and task relationships. Mainstream and commonly utilised definition approaches include: 1) Defining the tasks by the labels of tasks [253][254]. 2) Defining the tasks by their input features [125][255][256]. 3) Defining the tasks by their optimisation goals [257][258]. 4) Defining the tasks by the domain knowledge of tasks [259][260]. 5) Defining the tasks by their hierarchical structure [261][262]. 6) Defining the tasks by their temporal and spatial relationships [263][264]. 7) Defining the tasks by the constraint relationships between tasks [265][266]. 8) Defining the tasks by similarities between tasks [267][268]. 9) Defining the tasks by model structure and parameter sharing between tasks [269][270]. 10) Defining the tasks by the probability

distribution relationship between tasks [271][272].

For disease progression prediction researches, different prediction tasks can be learned as multiple tasks. Relationships between tasks can be defined in various ways: 1) Simultaneous learning tasks: all tasks are trained simultaneously and share model parameters [273][274]. 2) Predicting progression at different time points: predicting progression at 6 months, 12 months, 24 months, etc. There is a temporal relationship between these tasks, i.e., the earlier task provides valuable information for the subsequent tasks [11][142][275]. 3) Predicting disease progression in different subtypes of patients: disease progression in each subtype is a task [276]. For example, task one can be to predict the dementia progression of younger patients and task two can be to predict the dementia progression of older patients. The relationship between these tasks is hierarchical, as certain tasks can be more generic than others. 4) Dependency learning: the outcome of certain tasks can have an impact on others [277][278], e.g., age can have an impact on the disease progression prediction task. 5) Sequential learning: learning different tasks in a certain order [279][280], e.g., learning the age prediction task first and then the disease progression prediction task. 6) Hierarchical learning: different tasks are divided into different levels and the tasks in each level can share particular parameters [281][282], e.g., age prediction and gender prediction tasks are placed in the bottom level and disease progression prediction tasks are placed in the top level. 7) Combinatorial learning: combining multiple tasks to form a higher-level task [283][284], e.g., combining age prediction, gender prediction and education level prediction into a baseline model and then adding a disease progression prediction task. 8) Corporate learning: there can be common features between different tasks that can be learned together to improve performance and efficiency of the model [274][281], e.g., training the disease progression prediction and age prediction tasks together and learning the feature representation of brain imaging data and patient age together. 9) Alternate learning: alternating different tasks to improve generalisation ability and stability of the model [269][285], e.g., alternating the disease progression prediction

task and the age prediction task, training only one of them at a time. 10) Aggregate learning: aggregating the prediction results of multiple tasks to obtain the final prediction result [269][281], e.g., using a weighted average method to aggregate the prediction results of disease progression prediction, age prediction, gender prediction and education level prediction tasks to obtain the final disease progression prediction result.

To sum up, the design of task definition and task relationships in multi-task learning approaches for disease progression prediction applications requires determination based on the specific application scenario and data characteristics. Through effective task definition and task relationships design, multi-task learning can exploit the correlation and knowledge sharing between different tasks to improve the predictive capabilities, efficiency and generalisation of the model, resulting in higher accurate predictions for disease progression, management and treatment.

3.3.2.3 Multi-task learning regression for disease progression prediction

In the field of disease progression prediction, the major and important multi-task learning approaches define task associations through a diverse variety of correlations between various time points of disease progression. The most commonly used approach is to assume that during disease progression, the difference in cognitive scores (e.g., MMSE and ADAS-Cog) is small for two consecutive time points, also known as temporal group lasso regularisation (TGL) [11].

$$\min_{\mathrm{W}} \|\mathrm{XW}-\mathrm{Y}\|_{\mathrm{F}}^{2}+\theta_{1}\|\mathrm{W}\|_{\mathrm{F}}^{2}+\theta_{2}\|\mathrm{WH}\|_{\mathrm{F}}^{2}+\delta\|\mathrm{W}\|_{2,1} \tag{3.3}$$

Where $\|\mathrm{W}\|_{2,1}=\sum_{i=1}^{d}\sqrt{\sum_{j=1}^{t}\mathrm{W}_{ij}^{2}}$, the δ is a regularisation parameter. When t = 1, i.e., there is only one task for the model, the above penalty is simplified to lasso. When t > 1, the penalty utilises $\ell 2$-norm to group the weights of all tasks for a feature, and utilises $\ell 1$-norm to further group all features. Therefore, the $\ell 2$,1-norm attends to choose features ground on the importance of feature for all different t tasks. Another

formulation is to solve the problem of convex optimization with the convex fused sparse group Lasso (cFSGL) [11].

$$\min_{W} \|XW - Y\|_F^2 + \lambda_1 \|W\|_1 + \lambda_2 \|RW^T\|_1 + \lambda_3 \|W\|_{2,1} \quad (3.4)$$

where the first term measures empirical error with training data, $\|RW^T\|_1$ is the fused lasso penalty and $\|W\|_1$ is the lasso penalty. R = H^T is a $(t-1) \times t$ sparse matrix which utilised to describe the relationship between different tasks. The association of lasso and group lasso penalty is called the sparse group lasso penalty, which utilises mutual feature selection for all different tasks at the same time, and select a specific feature set for each task and fusion lasso penalty is used to make it smooth in time.

In addition, researchers have proposed a wide variety of multi-task learning regression approaches based on temporal correlation for disease progression prediction. [275][286][287] developed deep multi-task learning methods with the assumption that there is an intrinsic correlation between cognition, clinical diagnosis and ventricular volumes at each time point of prediction task. [288][289] proposed multi-task exclusive relational learning models, which utilise exclusive lasso regularisation and relational induced regularisation approaches that can select the most discriminative features for different tasks and model intrinsic correlations between different time points. [233][290][291] presented high-order multi-task learning models to explore the temporal correlation existing in data features and regression tasks through structured sparsity-inducing norms. [283][292] presented graph regularisation multi-task learning regression frameworks to capture relationship between different tasks and penalise large changes in the model at successive future time points.

3.3.3 Design and construction approach

3.3.3.1 Definition of tasks and task relationships for multi-dimensional tensor multi-task learning

Multi-task learning aims to learn a large number of related tasks together to ensure that the knowledge contained in one task can be utilised by other tasks, in order to accomplish knowledge sharing and enhance the generalisation performance of all tasks [128][274][253][293]. It can be learnt that the key to multi-task learning is the form, scope and depth of knowledge sharing. In contrast to the algorithmic task setting in the approach in Section 3.3.2.2, we assume that knowledge sharing across patient prediction tasks improves achievable performance, and therefore we set a single patient prediction task as a one task, which is a small-scale task setting approach. All supervised machine learning algorithms share certain knowledge (e.g., model weight parameters) between sample prediction tasks, but the scope and depth of shared knowledge is limited, this research intends to increase the scope and depth of knowledge sharing in this specific form of task setting. For task relationships, unlike traditional multi-task learning where task relationships are defined mainly through various assumptions between tasks, the approach in this research is to learn task relationships from the data itself by tensor decomposition of the raw data. We utilise the approach proposed in Chapter 4 to simultaneously assess and quantify the magnitude and direction information of brain structural variation utilising a similarity-

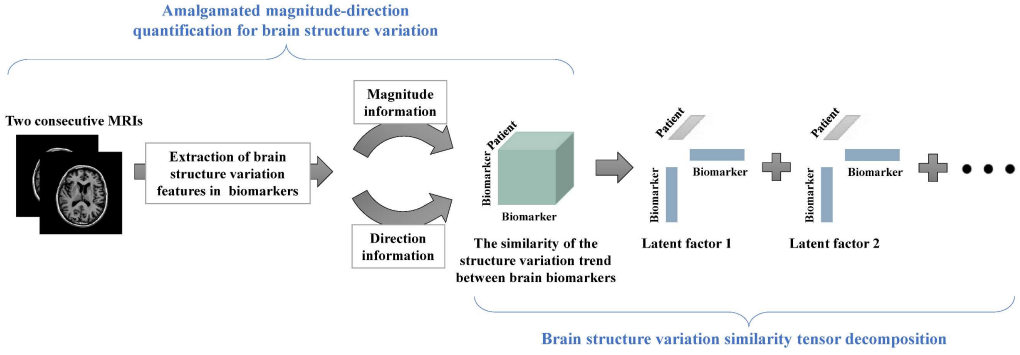


Figure 3.2: CP decomposition on a similarity tensor representation based on the similarity of the structure variation trend between brain biomarkers.

calculation-based amalgamated magnitude-direction quantification, which depicts the similarity of morphological trends between different biomarkers as a third-order tensor with dimensions corresponding to the first biomarker, the second biomarker and the patient sample. Then, the proposed approach utilises tensor decomposition to extract the set of first-order latent factors from the raw data (Figure 3.2). Each latent factor represented by its first biomarker, second biomarker and patient sample dimensions, interprets potential elements affecting the variability of the data in an interpretable manner and can be utilised as predictors for training the prediction model, and these latent factors are shared by the predictions for each patient sample. To sum up, the prediction of each patient sample in the tensor is a task, and all tasks share a set of biomarker latent factors (i.e., task relationships) from the tensor decomposition.

3.3.3.2 Algorithm design

The first key inspiration for the algorithm design in this research is not to utilise second-order matrices as model input features, but to propose building third-order tensors to construct AD progression prediction models that better present information and knowledge from multiple aspects of AD data along spatial and temporal dimensions. With the enhanced representation of AD biomarkers, the utilisation of multi-dimensional tensor in regression algorithms can enhance prediction accuracy, model stability, interpretability and feature comprehension. Alternatively, for AD prediction models, MTL can share information and knowledge across tasks, outperforming traditional single-task learning methods in terms of prediction accuracy, stability, interpretability and generalisability. It demonstrates superior performance when dealing with situations of small sample numbers. Therefore, the second key inspiration for the algorithm design of this research is to design and construct a tensor-based MTL approach to predict AD progression by integrating spatial and temporal information and knowledge of brain structural variation.

Specifically, the algorithm and model are designed to calculate the rate of change and

velocity of each biomarker with two consecutive MRI detections, forming a vector representing the trend of morphological variation of the biomarker; then the correlation of the spatio-temporal variation trend between two biomarkers is calculated and the data is encoded into a three-dimensional tensor, the proposed algorithm inputs the three-dimensional tensor and integrates tensor decomposition techniques in the algorithm to decompose the input tensor and extract a set of rank-one latent factors, which are shared by all sample predictions as a multi-task relationship and are utilised as predictors for training the prediction model.

3.4 Data and Experiments

3.4.1 Dataset

Data utilised in this research was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI started as a public-private partnership in 2003, directed by Principal Investigator Michael W. Weiner, MD. The fundamental purpose of ADNI has been to explore whether serial MRI, PET, other biological markers, clinical and neuropsychological assessments can be combined to measure, track and monitor the progression of MCI and early AD. The FreeSurfer image analysis software (http://surfer.nmr.mgh.harvard.edu/) was utilised by a team from the University of California, San Francisco (UCSF) to conduct volumetric segmentations and cortical reconstruction on imaging data from the ADNI database, which contains all ADNI subprojects (ADNI 1, 2, GO and 3). The FreeSurfer is an open-source software for processing and analysing MRI data of the human brain. It can process structural, functional and diffusion neuroimaging data, from which it extracts information on brain anatomy, cortical thickness, cortical partitioning, cortical folding, white matter fibre bundles, etc., and generates 3D models and quantitative results. It can also perform cortical surface reconstruction, partitioning and alignment. FreeSurfer is widely utilised in the fields of neuroscience, neurology and psychiatry to investigate

variations in the structure and function of the brain and correlations with cognition, disease, behaviour, emotion and other factors.

ADNI constituted the following Mini Mental State Examination (MMSE) to interpret the AD boundary:

- 30 is represented as cognitively no dementia,
- 29 to 26 is represented as questionable dementia,
- 25 to 21 is represented as mild dementia,
- 20 to 11 is represented as moderate dementia,
- 10 to 0 is represented as severe dementia.

The total ADAS-Cog score ranges from 0 to 70 with higher scores suggesting greater impairment. RAVLT TOTAL is the sum of the total number of words that the test subject can remember after learning a set of words in RAVLT (Rey Auditory Verbal Learning Test), the RAVLT TOTAL typically ranges from 0 to 75. FLU (Fluency test) scores are typically calculated based on the number of words that participants present. For tests on animals, normal people can normally present 15 to 20 different animal names in one minute, while people with certain cognitive impairments or neurological disorders can demonstrate poorer results.

The MMSE is a simple, quick neuropsychological test utilised to assess a patient's cognitive function. The MMSE involves a series of simple questions and tasks, including quizzes, naming objects, recalling words, counting, indicating time and place. The test typically takes between 5 and 10 minutes and is scored out of a total of 30 points, with the score for each item varying according to correctness and speed of completion. The advantages of the MMSE comprise its speed, simplicity, ease of use and analysis, enabling widespread usage in clinical, research and care settings. It has a high degree of reliability and validity, particularly in assessing cognitive function in older people. The MMSE can be utilised to examine cognitive function in patients with

AD, Parkinson's disease, stroke, craniocerebral injury, psychiatric disorders, etc. In clinical applications, clinicians can utilise MMSE to diagnose and monitor the condition of patients with cognitive impairment and to develop individualised treatment plans.

The ADAS-Cog is a commonly used scale for assessing cognitive function in people with dementia. It contains 11 items divided into the following 5 cognitive domains: memory, orientation, comprehension, language skills and attention. These include number string recall, object recognition, object names, orientation terms, object orientation, naming objects, understanding instructions, reading and answering questions, constructing simple sentences, connecting numbers and letters. The ADAS-Cog has numerous advantages in assessing cognitive function in patients with AD, such as a wide range of measures including several cognitive domains; high reliability and validity to effectively differentiate the level of cognitive ability between patients with AD and healthy older adults, good sensitivity and specificity. Its test procedure is simple, rapid and easy to administer, applicable for usage by clinicians and researchers.

The RAVLT is a commonly used psychometric instrument to assess an individual's memory ability. The RAVLT consists of two phases: learning and recall. During the learning phase, participants hear a list of 15 words and repeat the words as many times as possible. During the recall phase, participants are expected to recall as many words as possible that they heard in the learning phase. There are several indicators of RAVLT, the most commonly utilised of these is the RAVLT TOTAL score, which represents the total score obtained by the participant during the learning and recall phases. The RAVLT TOTAL is typically utilised to evaluate memory capacity in pathological conditions such as dementia, brain injury and other neurological disorders. A high RAVLT TOTAL can indicates better memory capacity, while a low RAVLT TOTAL can indicate memory problems or cognitive impairment.

The FLU is a test that evaluates language ability and cognitive function by being required to name a specific category of words, such as animal, fruit or colour in a certain amount of time. In the category of animals (FLU ANIM), the participant is requested to name as many animals as possible within a certain amount of time. Researches have demonstrated that for the FLU ANIM, normal adults can list 10 to 15 words in one minute. In contrast, older adults, patients with cognitive impairment and other neurological disorders typically perform at lower levels than normal adults. The FLU can be utilised to assess the impact of different neurological disorders on language and cognitive abilities, such as Parkinson's disease and AD. By applying the test, researchers and clinicians can better understand the manner in which these disorders affect patients' language and cognitive abilities, and develop better treatments.

3.4.2 Experimental design

In order to validate the performance of the proposed approach, the proposed model is first trained and tested by randomly splitting the data into training and testing sets in a ratio of 9:1. Since the values of the regularisation parameters, hyperparameters and the number of latent factors must be selected during the training phase, this research utilises 5-fold cross-validation on the training data to select them. This research compares the proposed approach with single-task learning, benchmarks and state-of-the-art MTL methods that have been selected as competing methods in clinical deterioration prediction researches. The research compares the prediction ability of multiple approaches for each single future time point utilising the root mean square error (rMSE) as the major assessment metric. For the overall regression performance evaluation, we utilise normalised mean square error (nMSE), which is typically utilised in MTL research [294], and weighted correlation coefficient (wR), which is typically utilised in medical literatures to resolve AD progression problems [295]. Figure 3.3 shows an overview of the construction, training and testing procedures for the proposed multi-dimensional tensor multi-task learning model.

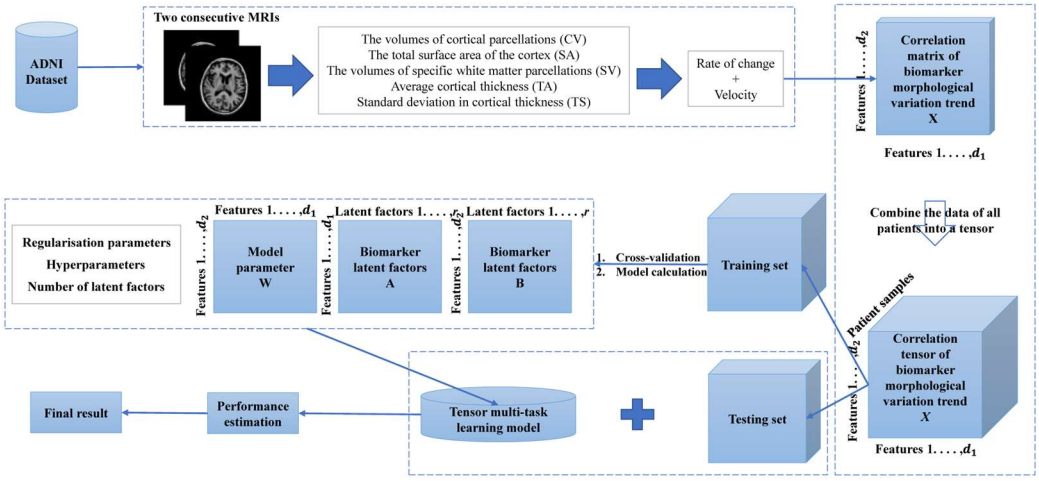


Figure 3.3: Overview of the construction, training and testing procedures for the proposed multi-dimensional tensor multi-task learning model.

3.4.3 Evaluation metrics

The multi-dimensional tensor multi-task learning regression model was developed with the quantified correlation tensor of morphological variation trends between MRI brain biomarkers. The evaluation metrics rMSE, nMSE and wR are defined as follows:

$$\mathrm{rMSE}(\mathrm{y},\hat{y}) = \sqrt{\frac{\|\mathrm{y}-\hat{\mathrm{y}}\|_2^2}{n}} \tag{3.5}$$

$$\mathrm{nMSE}\ (\mathrm{Y},\hat{\mathrm{Y}}) = \frac{\sum_{i=1}^{t} \|\mathrm{Y}_i-,\hat{\mathrm{Y}}_i\|_2^2 \Big/ \sigma(\mathrm{Y}_i)}{\sum_{i=1}^{t} n_i} \tag{3.6}$$

$$\mathrm{wR}(\mathrm{Y},\hat{\mathrm{Y}}) = \frac{\sum_{i=1}^{t} \mathrm{Corr}\ (\mathrm{Y}_i,\hat{\mathrm{Y}}_i) n_i}{\sum_{i=1}^{t} n_i} \tag{3.7}$$

where y is the ground truth of target at a single time point and $\hat{y}$ is the corresponding prediction from a prediction model for the rMSE. For nMSE and wR, Y_i is the ground truth of target at time point *i* and $\hat{\mathrm{Y}}_i$ 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 with different randomised data splits are reported.

3.4.4 Experimental parameter and computational infrastructure settings

The following are the range of values for each parameter attempted during the research development and experiments. This research utilises 5-fold cross-validation on the training data to select them.

- Hyperparameters:

λ [0.001 0.01 0.1 1 10 100 1000 10000]

β [0.001 0.01 0.1 1 10 100 1000 10000]

θ [0.001 0.01 0.1 1 10 100 1000 10000]

α [0, 1]

- The number of latent factors:

r [1 2 3 4 5 6 7 8 9 10]

We maintained the number of latent factors in a small range to save computation time for the model training phase. It is challenging to determine the precise rank of a tensor since it is an NP-hard problem in most instances [296][297]. As a result, the rank is typically obtained in practice by fitting numerous CP decompositions with different ranks until a reasonably "good" rank is derived.

The following are the computational infrastructures were utilised to conduct experiments in this research.

- Used software: MATLAB R2023a
- GPU: NVIDIA GeForce RTX 2070 with Max-Q Design

- CPU: Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz
- Amount of memory: 16GB
- Operating system: Windows 10

Since one of the problems targeted in this research is the small dataset problem under medical scenarios, computational complexity and training time were not considered as evaluation criteria in this research.

3.4.5 Selection of comparison methods

This research compares the proposed approach with the following single-task learning, benchmarks and state-of-the-art multi-task learning approaches, which were selected as competing methods in the research of predicting clinical deterioration. Including Ridge regression (Ridge) [55], Lasso regression (Lasso) [56], Temporal Group Lasso (TGL) [11], Non-convex Fused Sparse Group Lasso (nFSGL1) [127], Convex Fused Sparse Group Lasso (cFSGL) [11], Fused Laplacian Sparse Group Lasso (FL-SGL) [133], Non-Convex Calibrated Multi-Task Learning (NC-CMTL) [123], Joint feature and task aware multi-task feature learning (FTS-MTFL) [134], Group Asymmetric Multi-Task Learning (GAMTL) [122] and Dual feature correlation guided multi-task feature learning (dMTLc) [114].

For medical research, comparison methods all require a high level of interpretability. Specifically, Ridge and Lasso are classical single-task learning methods. TGL, nFSGL1 and cFSGL are benchmark algorithms that investigated and applied multi-task learning concepts to AD progression prediction. NC-CMTL and GAMTL are state-of-the-art general multi-task learning methods published in the top artificial intelligence conferences. FL-SGL, FTS-MTFL and dMTLc are state-of-the-art multi-task learning

methods applied to AD progression prediction published in the artificial intelligence and medical cross-discipline research top conferences and journal.

3.5 Interpretability of algorithm results

In medical research, the interpretability of approaches and results is as important as the model performance. The performance and reliability of a model is critical to the lives and health of patients, as the decisions based on the model can involve patients' survival and healthcare [298][299]. Clinicians and patients require an understanding of why a particular diagnosis or treatment option is being recommended in order to be capable of trusting and accepting it. If models are merely black boxes that cannot provide explanations or justifications for their decisions, clinicians and patients can find it difficult to trust their accuracy and validity and thus become sceptical to the decisions [300][301].

This research investigates and analyses three aspects of the interpretability of the proposed approach and predicted results. The first aspect is the brain biomarker latent factor, which is utilised in the algorithm as the multi-task relationship, and their visualisation and analysis can provide understanding of their different spatial distribution at different time points. It can contribute to identifying the brain regions that are implicated in the AD progression. The second aspect is the structural variation correlation between brain biomarkers. Extracting structural variation correlations that have a significant impact on prediction at different time points can be utilised in medical applications to assist clinicians to further investigate the pathogenesis and progression mechanisms of AD. The third aspect is the potential indicators for AD early detection. Since there is no cure for AD, the key to treatment is early detection and prevention of the disease. Therefore, recognizing important structural variation correlations between brain biomarkers in early MRI data can assist clinicians in identifying individuals with

suspected AD for early prevention. It further investigates the structural variation correlations between brain biomarkers extracted in the second aspect, and then evaluates and analyses the differences in the distribution of spatio-temporal structural variation correlation quantification values and differences in the distribution of relative structural variation status between biomarkers in cognitively impaired and non-cognitively impaired individuals at different time points, to discover potential indicators that can facilitate the identification of cognitively impaired individuals at an early stage from the magnitude and direction information of relative structural variation between biomarkers.

3.6 Summary

This chapter explores in detail how the aim of this research was achieved and the detailed procedures of how the research was conducted. The AD progression prediction approach proposed in this research is divided into two main parts, the first part is to design and construct the quantification approach for the higher-order tensor of AD biomarkers, and the second part is the design and construct the tensor multi-task learning algorithm and model, combining the two parts to achieve the research aim of integrating and mining spatio-temporal information and knowledge of MRI data to predict AD progression at multiple future time points. This chapter elaborates on the idea, design and components of the proposed approach, along with the analysis and specification of the design philosophy for each component. The dataset utilised and the design of the experimental method are then presented. Finally, the concept and process of investigating the interpretability of algorithms and experimental results is presented. The following sections present, demonstrate and discuss in detail the approach components, the experimental results, the interpretability of the algorithm and the experimental results.

Chapter 4

Multi-dimensional tensor construction of biomarkers

In order to quantify the spatio-temporal information and knowledge of AD brain biomarkers into a multi-dimensional tensor, it is of utmost importance to evaluate the type of tensor dimensions that are optimally suited to the AD prediction model and the approach to quantify and construct it into a multi-dimensional tensor. In this chapter, the proposed quantification approach based on biomarker correlations for constructing multi-dimensional tensor is described in detail in this chapter.

4.1 multi-dimensional knowledge vector construction

To quantify correlations between biomarkers with early AD data and approaches based on similarity calculations, we first constructed each MRI biomarker as a vector containing spatio-temporal information, which in the case of AD brain data represents the process of brain structural 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 $\frac{x_{M0}\ -x_{BL}}{t_{M06}-t_{BL}}$ 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.

The content of the vector is extensible, if more parameters regarding brain structural variation are considered in the future they can be integrated directly into the vector, or if there is a desire to investigate parameters in other areas or fields, they can be integrated directly into the vector and will not affect subsequent biomarker quantification, multi-dimensional tensor construction and multi-dimensional tensor multi-task learning algorithms.

4.2 Amalgamated magnitude-direction quantification for brain structure variation

We propose a two-stage quantitative approach that simultaneously assesses the magnitude and direction of structural variation among brain biomarkers (Figure 4.1). 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: $\mathrm{Ma}\left(|\mathrm{x}_i|, |\mathrm{x}_j|\right) = \sqrt{\left(|\mathrm{x}_i| - |\mathrm{x}_j|\right)^{\mathrm{T}} \mathrm{S}^{-1} (|\mathrm{x}_i| - |\mathrm{x}_j|)}$, 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 varies and the other does not varies, 5) both remain unvaried. 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 (4.1) 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.

$$\begin{cases} x = x, \text{if two biomarkers varied synchronously} \\ x = -x, \text{if two biomarkers varied asynchronously} \\ x = 0, \text{if two biomarkers are not relevant} \end{cases} \tag{4.1}$$

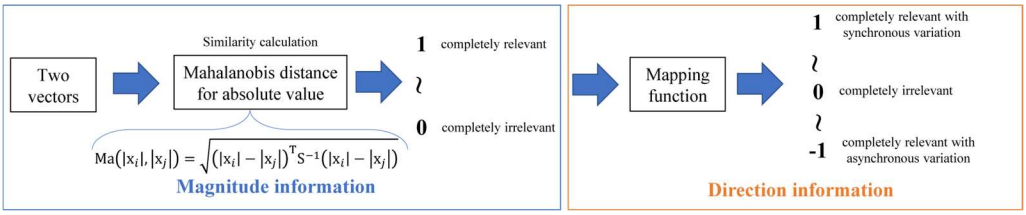
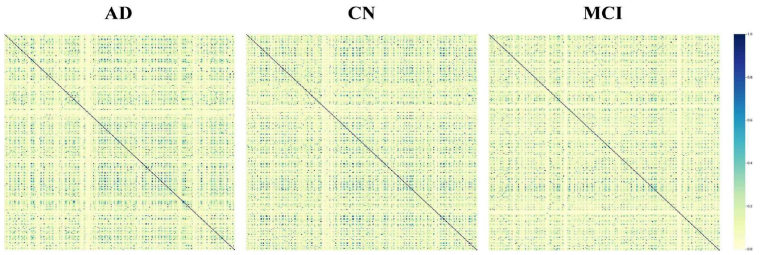
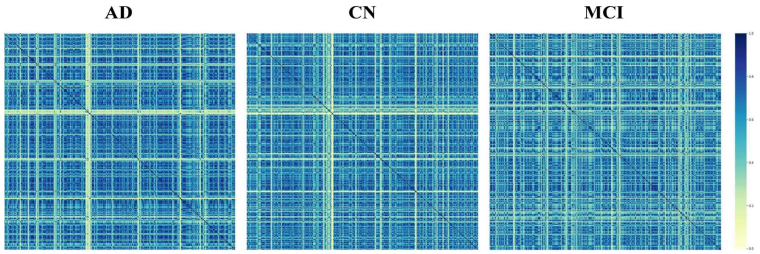


Figure 4.1: The two-stage quantitative approach that simultaneously assesses the magnitude and direction of structural variation among brain biomarkers.

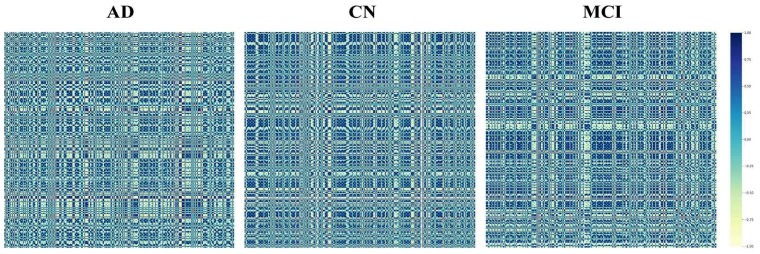
4.3 Comparison with mainstream and commonly used similarity calculation approaches



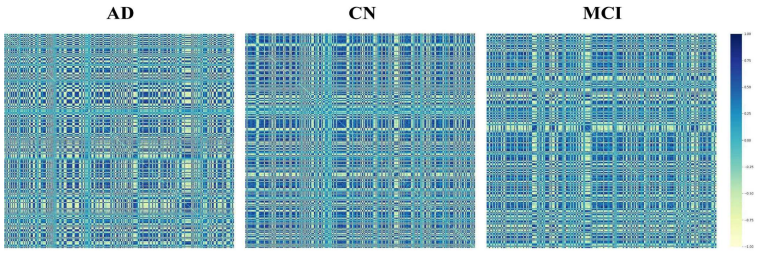
(a) Euclidean distance



(b) Mahalanobis distance



(c) Cosine similarity



(d) Amalgamated magnitude-direction quantification

Figure 4.2: 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).

Figure 4.2 depicts the structural variation correlations of brain biomarkers for AD, CN and MCI quantified by Euclidean distance, Mahalanobis distance, Cosine similarity and our proposed Amalgamated magnitude-direction quantification. We observed that our proposed quantification approach showed the largest 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.

4.4 Summary

This chapter designs and constructs an approach for the construction and quantification of a multi-dimensional tensor for biomarkers. The goal of the construction and quantification approach is to have the ability to mine and contain multi-dimensional spatio-temporal information and biomarker correlation knowledge. It is a novel quantification approach based on similarity calculations to simultaneously assess and quantify the magnitude and direction of brain structural variation. It describes the similarity of morphological trends between different biomarkers as a third-order tensor with dimensions corresponding to the first biomarker, the second biomarker and the patient sample to address the problem of the monotonic data form. It contains comprehensive information on brain structural variation and can effectively

differentiate between CN, MCI and AD patients. The quantification approach is detailed in two sub-components: the multi-dimensional knowledge vector construction and the amalgamated magnitude-direction quantification for brain structural variation. The detailed approach of each component has been described and discussed in sections 4.1 and 4.2. Chapter 5 would present the proposed multi-dimensional tensor multi-task learning regression algorithm for disease progression prediction.

Chapter 5

Multi-dimensional tensor multi-task learning regression

To investigate how to predict AD progression with multi-dimensional tensor data combined with multi-task learning regression, the main task is to design multi-task learning algorithms and models to incorporate and mine the spatio-temporal variation information and knowledge of MRI data in AD progression prediction in conjunction with the quantitative multi-dimensional tensor from the previous chapter in order to enhance the prediction accuracy, stability and interpretability in medical small dataset scenarios. This chapter first describes the components and workflow of the multi-dimensional tensor multi-task learning regression algorithm. The algorithm consists of three parts: supervised symmetric tensor decomposition to extract biomarker latent factors, tensor multi-task learning regression, and algorithmic regularisation terms. Secondly, the chapter defines, experiments, and analyses the overall algorithm and model along with their various components.

5.1 Supervised symmetric tensor decomposition

Our proposed formula requires an understanding of the latent factors of the correlation tensor of morphological variation trends between MRI brain biomarkers. These latent factors are represented by factor matrices A and B, which can be derived utilising tensor decomposition approaches. There are two mainstream standard approaches for tensor decomposition, specifically the Tucker and CANDECOMP/PARAFAC (CP)

decomposition. Tucker decomposition decomposes the tensor into the result of the core tensor and factor matrix for each mode. Although it presents a more comprehensive statement, it is difficult to interpret the latent factors since the number of latent factors can be different for each mode. In contrast, CP decomposition decomposes the tensor into a set of rank-one tensors. i.e., $X \approx [\![\mathrm{A}, \mathrm{B}, \mathrm{C}]\!] = \sum_{i=1}^{r} \mathrm{a}_i \circ \mathrm{b}_i \circ \mathrm{c}_i$, where $\circ$ denote the outer product operation between two vectors, while $\mathrm{a}_i, \mathrm{b}_i$ and c_i correspond to the vectors related with the *i*-th latent factor. Given a tensor X of the size $n_1 \times n_2 \times n_3$, the size of matrix A, B and C is $n_1 \times r, n_2 \times r$ and $n_3 \times r$ respectively.

Since the proposed amalgamated magnitude-direction quantification approach quantifies and constructs the MRI data as a three-dimensional symmetric tensor, we incorporate a symmetry trick into the tensor decomposition approach to make it a symmetric tensor decomposition. In order to extract and mine tensor latent factors specific to a particular model prediction task, we incorporate the symmetric tensor decomposition in the form of a regularisation term into the algorithm design. Acquiring biomarker latent factors by optimising the symmetric CP tensor decomposition objective function:

$$\| X - [\![\mathrm{A}_t, \mathrm{B}_t, \mathrm{C}_t]\!]_{\mathrm{S}} \|_{\mathrm{F}}^2 \tag{5.1}$$

where $X = [\![\mathrm{A}_t, \mathrm{B}_t, \mathrm{C}_t]\!]_{\mathrm{S}} = \sum_{i=1}^{r} \frac{1}{2} (\mathrm{a}_i^t \circ \mathrm{b}_i^t \circ \mathrm{c}_i^t + \mathrm{b}_i^t \circ \mathrm{a}_i^t \circ \mathrm{c}_i^t)$ where $\circ$ denote the outer product operation between two vectors for *t*-th prediction time point.

In addition to the characteristics of tensor data we are quantifying, symmetric tensor decomposition has the following advantages over tensor decomposition: 1) Symmetry of the result: symmetric tensor decomposition preserves the symmetric nature of the original tensor. 2) Greater computational efficiency: compared to general tensor decomposition approaches, symmetric tensor decomposition can be performed with fewer parameters to represent the symmetric tensor. This means that symmetric tensor decomposition reduces computational complexity, speeds up computation and is

particularly beneficial when working with large datasets. 3) Greater representational capability: the special structure of the symmetric tensor allows for greater representational capability. The symmetric tensor decomposition can provide a more compact and accurate representation than the general tensor decomposition. 4) Better coping with noise: the symmetric nature can limit the degrees of freedom of the factors, thus reducing the effect of noise and resulting in more reliable decomposition results. 5) Greater generalisation capability: since symmetric tensor decomposition considers symmetry in the generation of factors, it can be better adapted to unknown symmetric data. This makes symmetric tensor decomposition more advantageous when processing datasets with diversity. 6) Ability to conduct low-rank approximations: symmetric tensor decomposition can conduct low-rank approximations, thus reducing storage and computational costs while retaining important information. It is important when dealing with large scale data. 7) Ability to process non-linear data: since symmetric tensor decomposition can represent data as a product of low-dimensional factors, it can be utilised when dealing with non-linear data, thus improving the fit of the data. 8) Ability to enhance model robustness: since the symmetric tensor decomposition can be constrained for optimisation, the risk of overfitting the model can be reduced, thus improving the robustness and generalisation of the model.

5.2 Regularisation terms for tensor multi-task learning regression

5.2.1 Generalised temporal correlation term

In AD progression prediction, the goal is to predict cognitive scores at multiple future time points (e.g., ADAS-Cog and MMSE). In the real-world diagnosis of AD, experts must not only rely on the patients' current symptoms, but also consider their previous symptoms. Therefore, we assume that the *i*-th progression of a single AD patient is related to all previous progressions. Inspired by the concept of convex combination, we propose a new word, true progression, denoted as Δy. Which describe as follows:

$$\begin{cases} \Delta y_1 = \delta y_1 \\ \Delta y_2 = \alpha_1 \Delta y_1 + (1-\alpha_1)\delta y_2 \\ \Delta y_3 = \alpha_2 \Delta y_2 + (1-\alpha_2)\delta y_3 \\ \cdots \\ \Delta y_{t-1} = \alpha_{t-2} \Delta y_{t-2} + (1-\alpha_{t-2})\delta y_{t-1} \end{cases} \tag{5.2}$$

where the parameter α represents the relational degree of *i*-th progression and all previous progressions, its value ranges [0, 1]. In addition, the impact of each stage for disease progression on the following stage cannot be consistent, and therefore the relational degree parameters differ for each disease progression stage. The definition of *i*-th progression δy_i for one patient is:

$$\delta y_i = y_i - y_{i+1}, i = 1,2,\cdots,t-1. \tag{5.3}$$

where y_i represents prediction for the cognitive score at *i*-th time point. Due to the utilise of linear regression model, $y_i - y_{i+1} = x^T w_i - x^T w_{i+1} = x^T(w_i - w_{i+1})$, where x represents input features, w_i is *i*-th column of model parameter matrix W corresponding to the time point *i*. Therefore, (5.3) is equivalent to: $\delta w_i = w_i - w_{i+1}$.

According to (5.2), we have the following statement:

$$\begin{cases} \Delta w_1 = \delta w_1 \\ \Delta w_2 = \alpha_1 \Delta w_1 + (1-\alpha_1)\delta w_2 \\ \Delta w_3 = \alpha_2 \Delta w_2 + (1-\alpha_2)\delta w_3 \\ \cdots \\ \Delta w_{t-1} = \alpha_{t-2} \Delta w_{t-2} + (1-\alpha_{t-2})\delta w_{t-1} \end{cases} \tag{5.4}$$

where Δw denoted the progression with preceding progression information. w_i is the *i*-th column of W. where the parameter α represents relational degree of *i*-th progression and all preceding progressions.

The definition of *i*-th progression δw_i for one patient is:

$$\delta w_i = w_i - w_{i+1}, i = 1,2,\cdots,t-1. \tag{5.5}$$

As a result, we can describe the more realistic temporal smoothness assumption with matrix multiplication:

$$\mathrm{WP}(\alpha) = \mathrm{WHD}_1(\alpha_1)\mathrm{D}_2(\alpha_2)\cdots\mathrm{D}_{t-2}(\alpha_{t-2}) \tag{5.6}$$

where $\mathrm{H} \in \mathbb{R}^{t\times(t-1)}$ has the following definition: $\mathrm{H}_{ij} = 1$ if $i = j, \mathrm{H}_{ij} = -1$ if $i = j+1$ and $\mathrm{H}_{ij} = 0$ otherwise. $\mathrm{P}(\alpha)$ denotes the correlation between progress, it comprises the hyperparameters α, which depends on the result 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 substituted by α_i if $m = i, n = i+1$, the value of $\mathrm{D}_{i_{m,n}}(\alpha_i)$ is substituted by $1-\alpha_i$ if $m = n = i+1$.

We define this generalised temporal smoothness term as:

$$\|\mathrm{WP}(\alpha)\|_{\mathrm{F}}^2 \tag{5.7}$$

where $\|\cdot\|_{\mathrm{F}}^2$ refers to the Frobenius norm.

5.2.2 Lasso for biomarker latent factors and model parameters

Due to the large number of biomarker latent factors and model parameters, we incorporated lasso regularisation into the algorithm to perform feature selection at arithmetic time and remove biomarker latent factors and features that were not critical for the prediction task. We define the lasso term for biomarker latent factors and model parameters as follows:

$$\|\mathrm{W}_t, \mathrm{A}_t, \mathrm{B}_t, \mathrm{C}_t\|_1 \tag{5.8}$$

where $\|\mathrm{W}_t, \mathrm{A}_t, \mathrm{B}_t, \mathrm{C}_t\|_1$ applying an $\ell 1$-norm for W_t, A_t, B_t and C_t matrices individually for *t*-th time point.

In addition to feature selection, the application of lasso regularisation to biomarker latent factors and model parameters simultaneously in the algorithm has the following advantages: 1) Solving the covariance problem: when there are multiple highly correlated features, applying lasso regularisation allows the weights between them to be evenly distributed to each of the relevant features. This reduces the variance of the model and improves the stability of model. 2) Enhance model generalisation: lasso

regularisation can effectively control the complexity of model, thus reducing overfitting and improving the generalisation ability of the model. 3) Enhance the robustness of model: lasso regularisation can assist in enhancing the robustness of model in the event of abnormal data and noise, thus preventing overfitting. 4) Improve the speed and efficiency of the model: lasso regularisation can reduce the complexity of model, thus reducing the amount of computation and improving the speed and efficiency of the model. 5) Enhance model stability: in machine learning, it is often necessary to optimise the stability of a model to avoid it being overly sensitive to small changes in the data. Lasso regularisation can help improve the stability of a model by regularising multiple parameters, reducing the sensitivity of the model to noise and outliers, and enhancing the reliability and robustness of the model.

5.3 Tensor multi-task learning regression

This section describes our main algorithm designs, including 5.3.1 Major algorithm design and construction, which presents the best performing algorithms from our design and construction process. 5.3.2 Diverse algorithm design and construction, which presents the remaining algorithms designed and constructed during our research. The algorithm design of 5.3.1 was superior to that of 5.3.2 in this research on disease progression prediction algorithms, but the other algorithm designs of 5.3.2 can be valid for other future researches.

5.3.1 Major algorithm design and construction

To predict various cognitive scores (e.g., ADAS-Cog and MMSE) at future time points. Consider a multi-dimensional tensor multi-task regression problem for *t* time points, *n* training samples with d_1 and d_2 biomarker features. Let $X \in \mathbb{R}^{d_1 \times d_2 \times n}$ be the input third-order tensor constructed from two consecutive MRI records and it is the combination of calculated correlation matrix of 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 (clinical scores) and $\mathrm{y}_t = [y_1, \cdots, y_n] \in \mathbb{R}^n$ is the corresponding target at different time points.

For *t*-th prediction time point, the objective function of the proposed approach can be stated as follows:

$$L_t(X, \mathrm{y}_t) = \min_{\mathrm{W}_t,\, \mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t} \frac{1}{2} \|\hat{\mathrm{y}}_t - \mathrm{y}_t\|_2^2 + \frac{\lambda}{2} \|X - [\![\mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t]\!]_{\mathrm{S}}\|_{\mathrm{F}}^2 + \beta \|\mathrm{W}_t,\, \mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t\|_1$$

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

$$\text{where } \mathrm{U} = (\mathrm{A}_t \mathrm{B}_t^{\mathrm{T}}) \odot \mathrm{K} \odot \mathrm{W}_t \odot \mathrm{X}_n \ , \ \mathrm{U} \in \mathbb{R}^{d_1 \times d_2}. \tag{5.9}$$

where the first term calculates the empirical error with 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 first biomarker dimension, $\mathrm{B}_t \in \mathbb{R}^{d_2 \times r}$ is the latent factor matrix for second biomarker dimension and $\mathrm{C}_t \in \mathbb{R}^{n \times r}$ is the latent factor matrix for patient sample 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 regularisation parameters. Acquiring biomarker latent factors by optimising the symmetric CP tensor decomposition objective function $\|X - [\![\mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t]\!]_{\mathrm{S}}\|_{\mathrm{F}}^2$, where $X = [\![\mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t]\!]_{\mathrm{S}} = \sum_{i=1}^{r} \frac{1}{2} (\mathrm{a}_i^t \circ \mathrm{b}_i^t \circ \mathrm{c}_i^t + \mathrm{b}_i^t \circ \mathrm{a}_i^t \circ \mathrm{c}_i^t)$ where $\circ$ denote the outer product operation between two vectors for *t*-th prediction time point. The A_t and B_t biomarker latent factor matrices are utilised in the algorithm as multi-task relationships to share knowledge and enhance model generalisation, the C_t patient sample latent factor matrix is only utilised as a component in the symmetric CP tensor decomposition objective function and it has no effect in the multi-task relationships. $\|\mathrm{W}_t,\, \mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t\|_1$ applying the ℓ 1-norm for W_t, A_t, B_t and C_t matrices individually. 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*. The matrix $\mathrm{K} \in \mathbb{R}^{d_1 \times d_2}$ is the duplicate data correction matrix which utilised to solve the duplicate data problem since the correlation tensor for brain structural variation constructed by the proposed quantification approach is a symmetric

tensor, meaning that correlations between biomarkers are calculated in pairs and therefore half of the data are duplicates. It has the following description:

$$\mathrm{K} = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ \vdots & & \ddots & \vdots \\ & & & 1 \\ 0 & & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{d_1 \times d_2} \tag{5.10}$$

For all prediction time points, the objective function can be stated as follows:

$$L(X, \mathrm{Y}) = \min_{\mathrm{W}_f} \sum_1^t L_t(X, \mathrm{y}_t) + \theta \left\| \mathrm{W}_f \mathrm{P}(\alpha) \right\|_{\mathrm{F}}^2 \tag{5.11}$$

where $\left\| \mathrm{W}_f \mathrm{P}(\alpha) \right\|_{\mathrm{F}}^2$ is the generalised temporal correlation term, the model parameter matrix $\mathrm{W}_f \in \mathbb{R}^{(d_1 \times d_2) \times t}$ is temporal dimension unfolding for the model parameter tensor $W \in \mathbb{R}^{d_1 \times d_2 \times t}$, θ is the regularisation parameter. Generalised temporal correlation term detailed in Section 5.2.1.

It is worth mentioning that the nature of the algorithm (multi-dimensional tensor multi-task regression) proposed in this research is a linear model in order to maintain a high level of interpretability, but the quantification approach (AMDQ) proposed in this research can capture the non-linear information and knowledge in MRI biomarkers. Non-linear algorithms are based on mapping the input data to an interval value (e.g., [0, 1] or [-1, 1]), while the proposed AMDQ approach can also map the input MRI biomarker feature values to the brain biomarker structural variation correlation values with interval between -1 and 1. Therefore, the overall approach proposed in this research can be considered as a semi-linear model.

5.3.2 Diverse algorithm design and construction

Supervised tucker tensor decomposition term

Acquiring biomarker latent factors by optimising the tucker tensor decomposition objective function:

$$\|X - [\![\mathcal{G};\ \mathrm{A},\ \mathrm{B},\ \mathrm{C}]\!]\|_{\mathrm{F}}^2 \tag{5.12}$$

Where $X \approx [\![\mathcal{G};\ \mathrm{A},\ \mathrm{B},\ \mathrm{C}]\!] = \sum_{k=1}^{p} \sum_{j=1}^{q} \sum_{i=1}^{r} g_{kji} \mathrm{a}_k \circ \mathrm{b}_j \circ \mathrm{c}_i$. where $\circ$ denote the outer product operation between two vectors for *t*-th prediction time point. Tucker decomposition technique decomposes the tensor into the result of the core tensor ($\mathcal{G}$) and factor matrix ($\mathrm{A},\ \mathrm{B},\ \mathrm{C}$) for each mode. Although it presents a more comprehensive statement, it is difficult to interpret the latent factors since the number of latent factors can be different for each mode. The algorithm design of this research requires maintaining a high level of interpretability, therefore tucker decomposition technique is not applicable to this research.

The adjusted iterative convex combination of progression (aiccp) term

Based on generalised temporal correlation term in 5.2.1, the adjusted iterative convex combination of progression term states as follow:

$$\begin{aligned} \mathrm{WP}_{\mathrm{aiccp}}(\alpha) &= \mathrm{WHD}_1(\alpha)\mathrm{D}_2(\alpha) \cdots \mathrm{D}_{t-2}(\alpha)\mathrm{E}(\alpha) \\ \mathrm{P}_{\mathrm{aiccp}}(\alpha) &= \mathrm{HD}_1(\alpha)\mathrm{D}_2(\alpha) \cdots \mathrm{D}_{t-2}(\alpha)\mathrm{E}(\alpha) \end{aligned} \tag{5.13}$$

where $\mathrm{P}_{\mathrm{aiccp}}(\alpha)$ is the relation matrix for disease progression. $\mathrm{E}(\alpha) \in \mathbb{R}^{(t-1)\times(t-1)}$ is an identity matrix and the value of $\mathrm{E}_{m,n}(\alpha)$ is replaced by $(1-\alpha)^{-1}$ if $m = n = 2,3,\ldots,(t-1)$.

We define the aiccp term as:

$$\left\|\mathrm{WP}_{\mathrm{aiccp}}(\alpha)\right\|_{\mathrm{F}}^{2} \tag{5.14}$$

We now present the intuitive understanding of the aiccp term. The *i*-th true progression $\Delta y_i = [y_1, y_2, \cdots, y_i] \times \boldsymbol{l_i} - y_{i+1}$, here $\boldsymbol{l_i} \in \mathbb{R}^i$ is the coefficient of convex combination of the sates from y_1 to y_i. For a vector $\boldsymbol{l_i}$, $i = 1,2,\cdots,t-1$, $\boldsymbol{l}_i^j$ represents the *j*-th element of $\boldsymbol{l_i}$, we have

$$l_i^k \geqslant 0, k = 1,2,\cdots,i \tag{5.15}$$

$$\sum_{k=1}^{i} l_i^k = 1 \tag{5.16}$$

$$[y_1, y_2, \cdots, y_i] \times \boldsymbol{l_i} \text{ is a convex hull of } y_1, y_2, \cdots, y_i. \tag{5.17}$$

Now, we prove (5.15), (5.16) and (5.17). It is easy to notice that

$$\boldsymbol{l}_{\boldsymbol{i+1}}^{\mathrm{T}} = \left[\alpha \boldsymbol{l}_{\boldsymbol{i}}^{\mathrm{T}}, 1-\alpha\right]^{\mathrm{T}} \tag{5.18}$$

where $\alpha \in [0,1]$ and $\boldsymbol{l}_1 = [1]$ is a real number. Therefore, the vector $\boldsymbol{l_i}$ is non-negative with $i = 1,2,\cdots,(t-1)$, that means all the element value of $\boldsymbol{l_i}$, $l_i^k \geqslant 0, k = 1,2,\cdots,i$, so (5.15) is satisfied. We prove (5.16) by mathematical induction. It is obvious that (5.16) is satisfied when $i = 1$ and $\boldsymbol{l}_1 = [1]$. Therefore, we can assume $\boldsymbol{l_i}$, $i = 2,3,\cdots,t-2$, satisfies (5.16). According to (5.18), $\sum_{k=1}^{i+1} l_{i+1}^k = 1$, so (5.16) is satisfied for $i = k+1$. Based on (5.15) and (5.16), $[y_1, y_2, \cdots, y_i] \times \boldsymbol{l_i}$ is a convex hull of $\{y_1, y_2, \cdots, y_i\}$, so (5.17) is satisfied.

5.4 Model optimisation strategy

Latent factors $A \in \mathbb{R}^{d_1 \times r \times t}$, $B \in \mathbb{R}^{d_2 \times r \times t}$, $C \in \mathbb{R}^{n \times r \times t}$ 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. We apply proximal gradient descent approach to deal with each subproblem since not all components of the objective function are differentiable. The parts in our objective function relating Frobenius norms are differentiable, but the

parts involving the sparsity $\ell 1$-norms are not differentiable. The proximal approach is frequently applied to build proximal problem for non-smooth objective functions [138][302][303]. Consider a non-differentiable objective function f(x), which can be factorized into a smooth differentiable function d(x) and a non-smooth function n(x), i.e., f(x) = d(x) + n(x). The proximal gradient descent approach can be utilised to iteratively update model parameters as follows:

$$\mathrm{x}^{(s)} = \mathbf{prox}_{z_s,\mathrm{n}}\left(\mathrm{x}^{(s-1)} - z_s \nabla \mathrm{d}\left(\mathrm{x}^{(s-1)}\right)\right) \tag{5.19}$$

where $\mathrm{x}^{(s)}$ is the parameter to be estimated at step s. $\mathbf{prox}_{z_s,\mathrm{n}}$ is the proximal operator for non-differentiable function n, $\nabla \mathrm{d}\left(\mathrm{x}^{(s-1)}\right)$ is the gradient for the smooth function d with regard to $\mathrm{x}^{(s-1)}$, z_s is the step size for gradient descent update. The proximal operator for $\ell 1$-norm function is the soft-thresholding operator [304] as follows:

$$\mathbf{prox}_{\xi,\mathrm{n}}(\mathrm{v}) = (\mathrm{v} - \xi)_+ - (-\mathrm{v} - \xi)_+ \tag{5.20}$$

where ξ is the threshold parameter. Parameters are iteratively updated by calculating the gradient on smooth part of the objective function then applying the soft-thresholding operator (proximal mapping function for $\ell 1$-norm) to establish its next value. The step size can be acquired utilising a line search algorithm. The approach can accelerate optimization convergence and simplify the design of distributed optimization algorithms.

5.5 Experiments, results analysis and discussion procedures

The following sections represent the experimental configurations and procedures required for the multi-dimensional tensor multi-task learning regression algorithm and model, as well as tests and analysis of the results for the algorithm and model. The dataset utilised in this research and its sources are detailed in Section 3.4.1. The evaluation metrics utilised in the experiments of the research are described in Section

3.4.3. The range of values for each parameter attempted and the computational infrastructures were utilised to conduct experiments during the research development are detailed in Section 3.4.4. The selection of comparison methods is described in Section 3.4.5. Section 5.6 provides details of the pre-processing procedures performed on the dataset, and the specifics of the pre-processed dataset. Section 5.7 compares the proposed algorithms and models with benchmarks and state-of-the-art approaches to validate performance and effectiveness of the proposed algorithm and model. Section 5.8 elaborates on the application of the proposed algorithm and model on clinical scenarios.

5.6 Pre-processing procedures for MRI data

The protocol for the research can be found at http://adni.loni.ucla.edu/research/mri-post-processing/. The ADNI official website (http://adni.loni.ucla.edu) has more information on ADNI MRI procedures and imaging equipment. We obtained the MRI data from the ADNI website and proceeded to conduct the following pre-processing procedures:

- Removal of image records with failed quality control;
- Individuals who lacked BL and M06 MRIs were eliminated;
- Removal of features with more than half of the samples having missing values;
- Fill in the remaining missing data by taking the average of the features.

There are 313 MRI features (Appendix B) after the pre-processing stages. which are divided into five categories: the volumes of cortical parcellations, the total surface area of the cortex, the volumes of specific white matter parcellations, the average cortical thickness and the standard deviation in cortical thickness. Table 5.1 and 5.2 illustrates the demographic features of the ADNI MRI data utilised in this research.

Table 5.1: Demographic characteristic of the studied subjects (MMSE and ADAS-Cog) valued are specified as mean±standard deviation.

Time point	Attribute	MMSE	ADAS-Cog
M12	Sample size (CN, MCI, AD)	1334 (359, 726, 249)	1321 (354, 722, 245)
	Gender(f/m)	580/754	575/746
	Age	74.9±7.2	74.9±7.1
M24	Sample size (CN, MCI, AD)	1127 (335, 620, 172)	1105 (332, 613, 160)
	Gender(f/m)	493/634	481/624
	Age	75.8±7.1	75.8±7.1
M36	Sample size (CN, MCI, AD)	745 (206, 528, 11)	730 (203, 518, 9)
	Gender(f/m)	324/421	318/412
	Age	76.4±7.0	76.4±7.1
M48	Sample size (CN, MCI, AD)	585 (218, 360, 7)	579 (215, 357, 7)
	Gender(f/m)	259/326	261/318
	Age	76.9±6.8	77.0±6.8
M60	Sample size (CN, MCI, AD)	333 (115, 216, 2)	330 (115, 213, 2)
	Gender(f/m)	144/189	143/187
	Age	78.2±6.7	78.3±6.7
M72	Sample size (CN, MCI, AD)	299 (142, 155, 2)	299 (142, 156, 1)
	Gender(f/m)	128/171	130/169
	Age	79.8±6.6	79.7±6.6
M84	Sample size (CN, MCI, AD)	229 (107, 121, 1)	225 (107, 117, 1)
	Gender(f/m)	102/127	100/125
	Age	80.5±6.3	80.4±6.3
M96	Sample size (CN, MCI, AD)	178 (87, 89, 2)	175 (87, 86, 2)
	Gender(f/m)	73/105	72/103
	Age	81.1±6.6	81.1±6.6
M108	Sample size (CN, MCI, AD)	121 (59, 62, 0)	118 (61, 57, 0)
	Gender(f/m)	55/66	52/66
	Age	81.7±6.3	81.7±6.1
M120	Sample size (CN, MCI, AD)	81 (49, 32, 0)	80 (49, 31, 0)
	Gender(f/m)	38/43	37/43
	Age	83.3±6.1	83.5±5.8

Table 5.2: Continuation of Table 5.1. Demographic characteristic of the studied subjects (RAVLT TOTAL and FLU ANIM) valued are specified as mean±standard deviation.

Time point	Attribute	RAVLT TOTAL	FLU ANIM
M12	Sample size (CN, MCI, AD)	960 (265, 530, 165)	1329 (359, 726, 244)
	Gender(f/m)	458/502	579/750
	Age	74.4±7.0	74.9±7.1
M24	Sample size (CN, MCI, AD)	816 (249, 450, 117)	1118 (334, 617, 167)
	Gender(f/m)	389/427	490/628
	Age	75.3±7.0	75.8±7.1
M36	Sample size (CN, MCI, AD)	548 (155, 387, 6)	740 (206, 524, 10)
	Gender(f/m)	252/296	321/419
	Age	76.0±6.9	76.4±7.1
M48	Sample size (CN, MCI, AD)	435 (163, 266, 6)	578 (217, 354, 7)
	Gender(f/m)	209/226	260/318
	Age	76.7±6.7	77.0±6.8
M60	Sample size (CN, MCI, AD)	248 (90, 156, 2)	328 (115, 211, 2)
	Gender(f/m)	109/139	141/187
	Age	78.0±6.6	78.2±6.7
M72	Sample size (CN, MCI, AD)	230 (113, 116, 1)	296 (142, 153, 1)
	Gender(f/m)	98/132	128/168
	Age	79.8±6.4	79.7±6.6
M84	Sample size (CN, MCI, AD)	184 (89, 94, 1)	228 (108, 119, 1)
	Gender(f/m)	83/101	104/124
	Age	80.6±6.0	80.4±6.3
M96	Sample size (CN, MCI, AD)	137 (69, 66, 2)	178 (87, 89, 2)
	Gender(f/m)	56/81	74/104
	Age	81.3±6.6	81.1±6.6
M108	Sample size (CN, MCI, AD)	95 (51, 44, 0)	125 (63, 61, 1)
	Gender(f/m)	41/54	55/70
	Age	81.9±5.9	81.6±6.3
M120	Sample size (CN, MCI, AD)	59 (35, 24, 0)	103 (60, 43, 0)
	Gender(f/m)	28/31	48/55
	Age	83.2±6.3	82.3±6.2

5.7 Experimental results

5.7.1 Comparison with the benchmarks and state-of-the-arts

We utilised the proposed amalgamated magnitude-direction quantification combined with the tensor multi-task learning regression algorithm (AMDQ-TMTL) to compare with the following single-task learning, benchmarks and state-of-the-art multi-task learning approaches, which were selected as competing methods in the research of predicting clinical deterioration. Including Ridge regression (Ridge) [55], Lasso regression (Lasso) [56], Temporal Group Lasso (TGL) [11], Non-convex Fused Sparse Group Lasso (nFSGL1) [127], Convex Fused Sparse Group Lasso (cFSGL) [11], Fused Laplacian Sparse Group Lasso (FL-SGL) [133], Non-Convex Calibrated Multi-Task Learning (NC-CMTL) [123], Joint feature and task aware multi-task feature learning (FTS-MTFL) [134], Group Asymmetric Multi-Task Learning (GAMTL) [122] and Dual feature correlation guided multi-task feature learning (dMTLc) [114]. The experimental results for MMSE prediction are shown in Table 5.3, for ADAS-Cog prediction are shown in Table 5.4, for RAVLT TOTAL are shown in Table 5.5 and for FLU ANIM are shown in Table 5.6.

The proposed approach has a lower rMSE than other comparison methods for most (95%) of the individual time points. In terms of overall regression performance, our proposed approach exceeds comparison methods in terms of nMSE and wR for MMSE, ADAS-Cog, RAVLT TOTAL and FLU ANIM, demonstrating that our method outperforms competitors. Our observations are as follows: 1) The proposed AMDQ-TMTL approach outperforms single-task learning, benchmarks 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 approach significantly improves prediction stability. The results obtained through 20 iterations of the experiment had a lower standard deviation than the comparison

approaches. 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.

Table 5.3: Comparison of the results from our proposed approaches with the single-task learning, benchmarks and state-of-the-art MTL methods for MMSE at time points M12 to M120. The best results are bolded.

Target: MMSE	nMSE	wR	M12 rMSE	M24 rMSE	M36 rMSE	M48 rMSE
Ridge	2.6314±1.2957	0.2110±0.0632	4.0605±1.2386	5.0048±1.2526	4.0442±1.4497	5.5393±1.4989
Lasso	0.8472±0.1195	0.4966±0.0757	1.9052±1.1061	2.4768±1.5718	2.5940±0.7297	2.2096±0.6772
TGL	0.4967±0.1441	0.7537±0.0732	1.7395±0.1282	1.7217±0.1640	1.8433±0.1870	1.7016±0.4988
nCFGL1	0.4248±0.1842	0.8129±0.0296	1.3466±0.1764	**1.3788±0.2848**	1.9872±0.1796	1.6300±0.3682
cFSGL	0.4018±0.2664	0.8410±0.0329	1.5864±0.6592	1.6055±0.7507	1.6029±0.4876	1.8139±0.9093
FL-SGL	0.4930±0.1557	0.7673±0.1676	1.7686±0.1391	2.2371±0.4471	2.0123±0.2851	1.7532±0.5990
NC-CMTL	0.4409±0.2107	0.8020±0.0637	1.4302±0.4354	1.5425±0.3381	1.7830±0.5100	2.0724±0.5488
FTS-MTFL	0.4598±0.1321	0.7807±0.0727	1.6373±0.3992	1.7305±0.4099	1.5596±0.6720	1.6122±0.5982
GAMTL	0.5099±0.2507	0.7030±0.1436	1.8805±0.1909	1.9699±0.2790	2.1551±0.1280	1.8221±0.6720
dMTLc	0.5018±0.1035	0.7082±0.1271	1.8372±0.1873	1.9649±0.6950	1.8731±0.5722	1.7281±0.5057
AMDQ-TMTL	**0.2917±0.0852**	**0.8617±0.0124**	**1.3194±0.1091**	1.4468±0.1247	**1.5221±0.1048**	**1.5720±0.1066**

Target: MMSE	M60 rMSE	M72 rMSE	M84 rMSE	M96 rMSE	M108 rMSE	M120 rMSE
Ridge	6.3593±1.4969	7.9274±0.9130	7.5438±1.0703	7.1825±1.1658	7.3953±0.9400	8.6403±1.3850
Lasso	2.2814±0.9163	3.5367±1.8656	4.4608±0.8752	5.2756±2.1243	5.1665±1.0716	5.3829±1.9340
TGL	1.8078±0.7116	1.9643±0.9502	2.1360±0.8765	2.6708±0.7499	3.0527±0.6373	3.1320±1.1055
nCFGL1	1.6967±0.6013	2.1037±0.6930	2.5159±0.3190	2.8206±0.5456	2.8242±0.5211	2.9035±0.6056
cFSGL	1.8209±0.8150	2.2522±0.7898	2.3981±0.3207	2.6192±1.3076	2.2614±0.5827	3.5795±0.3449
FL-SGL	1.6258±0.2276	1.9267±0.6614	2.0164±0.6971	2.5265±0.2901	3.0563±0.4682	3.7890±0.3501
NC-CMTL	1.9627±0.8186	1.9513±0.8896	2.1118±0.3061	1.9114±0.4402	2.1879±0.7359	3.7274±0.8463
FTS-MTFL	2.3715±0.7826	1.8304±0.2191	1.8528±0.7210	2.4228±0.8804	1.8972±0.2475	3.7141±1.0977
GAMTL	1.9757±0.5196	2.1530±0.6055	2.2004±0.7192	1.9492±0.7142	2.3220±1.1041	3.4550±0.6403
dMTLc	1.5245±0.8603	1.7714±0.4606	1.8294±0.3123	2.1448±0.8492	2.5025±0.2857	3.4188±0.6761
AMDQ-TMTL	**1.4833±0.1649**	**1.5661±0.1684**	**1.6105±0.2253**	**1.6302±0.2874**	**1.6375±0.2206**	**2.1358±0.3521**

Table 5.4: Comparison of the results from our proposed approaches with the single-task learning, benchmarks and state-of-the-art MTL methods for ADAS-Cog at time points M12 to M120. The best results are bolded.

Target: ADAS-Cog	nMSE	wR	M12 rMSE	M24 rMSE	M36 rMSE	M48 rMSE
Ridge	2.6042±0.5035	0.2668±0.1008	7.8783±1.2673	8.6661±1.4368	7.4033±1.5643	8.7409±1.8191
Lasso	0.8908±0.4853	0.4091±0.0237	5.7337±1.3240	6.7600±1.7528	6.5083±1.4925	7.4186±1.2113
TGL	0.4794±0.2107	0.7267±0.0646	2.6734±0.8648	2.3945±1.0648	3.6386±1.1138	3.4775±0.5774
nCFGL1	0.3746±0.1589	0.8207±0.2134	2.7870±0.5103	2.8953±0.7292	2.5707±0.6152	3.2667±0.3667
cFSGL	0.3490±0.1942	0.8536±0.0724	2.1080±0.2942	3.0577±0.2916	2.8119±0.2969	3.1791±0.3089
FL-SGL	0.4269±0.1011	0.7917±0.1235	2.6934±0.6158	2.4190±0.3710	3.0659±1.1312	3.4480±0.4345
NC-CMTL	0.4087±0.1030	0.8057±0.0370	2.2452±0.6398	2.7504±0.3242	3.1455±0.2992	3.1756±0.3682
FTS-MTFL	0.4074±0.1665	0.8066±0.0211	2.7443±0.8547	2.4485±0.4728	3.0464±0.3155	2.2464±0.4095
GAMTL	0.3381±0.1075	0.8660±0.0310	2.9950±0.5020	3.0906±0.4274	2.2789±0.4378	2.3314±0.3827
dMTLc	0.3309±0.0861	0.8730±0.0472	2.3471±0.3243	2.4841±0.4506	2.9834±0.6368	2.4808±0.6388
AMDQ-TMTL	**0.2436±0.0571**	**0.8843±0.0418**	**1.5298±0.2621**	**1.4607±0.2294**	**1.5192±0.1306**	**1.4679±0.2186**

Target: ADAS-Cog	M60 rMSE	M72 rMSE	M84 rMSE	M96 rMSE	M108 rMSE	M120 rMSE
Ridge	8.6757±1.6906	9.0341±1.9737	9.9267±1.0749	9.1536±2.8575	9.1306±2.7906	10.0144±2.7013
Lasso	6.7316±1.9468	7.0183±1.8743	7.6153±1.1450	7.6934±1.4234	8.4926±1.3660	8.7862±1.8170
TGL	3.3258±0.9778	3.1367±0.9183	4.9036±0.9516	4.6745±0.7821	6.5995±0.9961	8.6548±0.4650
nCFGL1	2.4117±1.0509	3.0193±0.4929	3.8003±1.1919	3.8142±0.7059	5.1221±0.4251	6.0504±0.4185
cFSGL	2.8553±0.7009	3.1638±0.7193	3.1942±0.9396	3.5658±0.5458	5.0652±0.5768	5.6949±0.8991
FL-SGL	3.3393±0.2468	3.3664±0.5003	3.5970±0.2306	4.0665±0.9310	5.6875±0.6848	6.2026±0.6328
NC-CMTL	3.6033±0.9507	3.0725±0.4072	4.3786±0.3920	3.5976±0.6149	3.2001±0.6762	7.6423±1.3861
FTS-MTFL	3.1216±0.3429	3.5349±0.2918	4.1791±0.4363	3.2672±0.4010	3.5568±0.4028	7.8608±1.2826
GAMTL	3.6593±0.7849	2.3207±0.5119	3.8476±0.7138	2.6636±0.1940	3.8485±0.2107	5.8844±1.1164
dMTLc	3.4192±0.2921	2.5766±0.3138	3.1199±0.3819	3.1522±0.5621	4.4009±0.7587	5.9746±1.5326
AMDQ-TMTL	**1.6805±0.2404**	**1.4750±0.1545**	**1.5658±0.1810**	**1.6959±0.1087**	**1.6982±0.1979**	**2.1631±0.3595**

Table 5.5: Comparison of the results from our proposed approaches with the single-task learning, benchmarks and state-of-the-art MTL methods for RAVLT TOTAL at time points M12 to M120. The best results are bolded.

Target: RAVLT TOTAL	nMSE	wR	M12 rMSE	M24 rMSE	M36 rMSE	M48 rMSE
Ridge	2.4552±0.4233	0.2970±0.0879	7.9376±1.9086	8.2124±1.4203	8.8234±1.0348	9.1893±0.5844
Lasso	0.8544±0.5287	0.4086±0.0524	5.3266±1.1728	5.7378±0.8868	6.3218±0.5375	6.4050±0.5636
TGL	0.4951±0.1742	0.7698±0.1060	2.2520±0.6345	2.7132±0.7384	3.2253±0.8426	3.3035±0.6113
nCFGL1	0.3726±0.1417	0.8062±0.0614	2.4384±0.6609	2.3252±0.6353	2.7824±0.6245	2.8225±0.4127
cFSGL	0.3474±0.1648	0.8355±0.0484	2.3591±0.8820	2.4297±0.5859	2.6760±0.3778	3.0636±0.3502
FL-SGL	0.3372±0.1158	0.8485±0.0702	2.2719±0.8909	2.4105±1.1230	3.2312±0.1693	2.6380±0.3065
NC-CMTL	0.3631±0.1015	0.8240±0.0685	2.7621±0.2550	2.3219±0.3455	2.7690±0.1420	3.1030±1.0612
FTS-MTFL	0.3554±0.1973	0.8280±0.0139	2.5827±0.6453	2.6332±0.7640	2.6991±0.7002	3.1862±0.6679
GAMTL	0.3599±0.2934	0.8150±0.0939	2.7978±0.4072	2.7376±0.7520	3.2829±0.2948	2.7709±0.7505
dMTLc	0.3650±0.1569	0.8106±0.0742	2.8858±0.1441	2.3956±0.4084	3.4184±0.4235	2.7441±0.7612
AMDQ-TMTL	**0.2546±0.0862**	**0.8722±0.0478**	**1.4314±0.1056**	**1.6062±0.1550**	**1.6425±0.2317**	**1.5017±0.2467**

Target: RAVLT TOTAL	M60 rMSE	M72 rMSE	M84 rMSE	M96 rMSE	M108 rMSE	M120 rMSE
Ridge	8.8236±0.8598	9.2771±1.2987	9.4471±0.9660	9.6810±1.1429	9.0462±1.4533	10.5317±1.7693
Lasso	7.1748±1.3071	7.6141±0.9423	8.1238±0.7045	7.9922±1.0334	8.0109±1.3453	8.2628±1.6440
TGL	2.8406±0.6077	3.4649±0.8560	4.6258±0.6558	6.7818±0.3483	6.9691±1.0105	7.5915±1.3104
nCFGL1	2.7932±0.6532	3.2405±0.6746	4.0655±0.5738	4.1362±1.2326	5.3567±0.8017	6.1909±0.7436
cFSGL	2.5154±0.4834	2.9631±0.9677	3.7082±0.5020	3.8204±0.7361	5.2445±1.1207	6.1520±0.5217
FL-SGL	2.7104±0.3276	3.4238±0.4533	3.6328±0.3646	3.5606±1.0927	5.0397±0.7480	5.7075±0.4890
NC-CMTL	3.1478±0.2801	2.9095±0.5860	3.8791±0.8183	4.1731±0.6744	5.1556±0.4280	6.4508±0.9031
FTS-MTFL	2.9759±0.5230	3.4540±0.8699	3.4254±0.3827	3.4863±0.7479	5.0220±0.4201	6.0470±1.0560
GAMTL	2.6841±0.9265	3.5007±0.3095	3.1905±0.1981	3.9752±0.7219	5.2273±0.2456	6.5358±0.6512
dMTLc	3.1043±0.6300	2.8967±0.5255	3.5972±0.6105	3.2667±0.4401	5.7278±0.4308	6.9348±0.4579
AMDQ-TMTL	**1.6537±0.1928**	**1.6270±0.2594**	**1.8475±0.2760**	**1.8646±0.2633**	**1.6571±0.2034**	**2.2641±0.4324**

Table 5.6: Comparison of the results from our proposed approaches with the single-task learning, benchmarks and state-of-the-art MTL methods for FLU ANIM at time points M12 to M120. The best results are bolded.

Target: ***FLU ANIM***	nMSE	wR	M12 rMSE	M24 rMSE	M36 rMSE	M48 rMSE
Ridge	2.2734±1.6694	0.2320±0.0913	5.4472±1.8837	6.1197±1.3662	4.2137±1.7492	6.4510±1.1308
Lasso	0.8544±0.5287	0.4868±0.0524	2.5727±0.8031	1.8910±0.6070	2.4185±1.2328	1.9587±0.9794
TGL	0.4316±0.1673	0.8272±0.0102	1.3364±0.7055	1.7315±0.6941	1.5969±0.6927	1.7610±0.6014
nCFGL1	0.4717±0.1426	0.8160±0.0614	1.3493±0.7175	1.9113±0.3184	2.2461±0.6539	2.2189±0.7169
cFSGL	0.4236±0.1204	0.8304±0.0461	1.4223±0.4127	1.6939±0.2266	1.8820±0.7462	2.0451±0.4260
FL-SGL	0.4812±0.0892	0.8016±0.0224	1.4973±0.2302	1.7054±0.2693	2.0819±0.2348	2.3585±0.2543
NC-CMTL	0.5290±0.1447	0.7531±0.0691	1.7739±0.8168	2.1327±0.6380	2.3885±0.2849	2.1918±0.3549
FTS-MTFL	0.4966±0.1727	0.7840±0.0423	1.3370±0.5862	2.0966±0.7896	2.2042±0.2374	1.9100±0.3930
GAMTL	0.5079±0.0496	0.7607±0.0236	1.8799±0.6390	2.1003±0.8619	2.3044±0.8389	**1.6065±0.7484**
dMTLc	0.4529±0.1132	0.8261±0.0519	1.6090±0.3121	1.7672±0.2789	1.7853±0.5852	2.0958±0.7601
AMDQ-TMTL	**0.3017±0.0325**	**0.8572±0.0267**	**1.3153±0.2125**	**1.4408±0.2174**	**1.4380±0.2559**	1.6463±0.1761

Target: ***FLU ANIM***	M60 rMSE	M72 rMSE	M84 rMSE	M96 rMSE	M108 rMSE	M120 rMSE
Ridge	5.3020±1.2161	5.9618±1.6148	4.3180±2.1455	5.2603±2.2362	4.6195±1.0933	6.1033±1.1361
Lasso	1.9725±0.9758	2.3282±0.5055	3.3004±1.6036	3.0561±1.3526	4.2697±1.0291	5.5735±1.1790
TGL	2.2004±0.8175	2.1354±0.6304	1.9228±0.4608	2.7740±1.2036	3.6916±0.8542	4.2636±1.1491
nCFGL1	1.8437±0.5516	1.8129±1.1824	1.9417±1.2021	2.9784±1.2004	3.5532±0.4738	4.0436±0.5043
cFSGL	2.0344±0.3400	2.3855±0.7762	1.6734±0.9634	2.0328±0.4551	3.3721±0.7598	3.4367±0.9471
FL-SGL	2.3354±0.6026	2.2094±0.3552	2.0910±0.2739	1.7509±0.3284	2.9450±0.5829	3.7098±0.4743
NC-CMTL	2.1895±0.5107	1.9796±0.9002	2.3260±0.6398	2.7319±0.2494	2.6456±0.4240	3.7135±0.2451
FTS-MTFL	1.9178±0.4717	2.0398±1.0532	2.0742±0.7679	1.8550±0.4091	2.6720±0.2375	3.6095±0.8001
GAMTL	1.7183±0.7254	2.3639±0.4500	2.4301±0.4398	2.3459±0.4785	2.4059±0.1056	3.5854±0.9655
dMTLc	1.7940±0.3447	2.4372±0.6470	2.1386±0.8401	2.2573±0.4613	2.4334±0.4058	3.3250±0.5132
AMDQ-TMTL	**1.4964±0.2937**	**1.4092±0.1424**	**1.5367±0.1827**	**1.4207±0.1808**	**1.6149±0.2787**	**2.1355±0.2624**

5.7.2 Ablation studies

The proposed AMDQ-TMTL approach contains various parameters and regularisation terms. The experimental results in Table 5.3, 5.4, 5.5 and 5.6 are sufficient to demonstrate the effectiveness of supervised symmetric tensor decomposition and biomarker latent factors. In addition, the main parts include 1) The duplicate data correction matrix K, 2) The $\ell 1$-norm and 3) The generalised temporal correlation term. In this section, we present the effectiveness of these parameters and regularisation terms in the proposed AMDQ-TMTL approach.

Figure 5.1, 5.2, 5.3 and 5.4 show the results of ablation experiments for different time points with different MRI inputs. The settings of the experiments are the same as the previous AMDQ-TMTL experiments. All reported rMSE values are based on the average of 20 iterations of different randomly split of data.

From the experimental results, it can be observed that the indicated parameters and regularisation terms have improved effects on the approach. Among them, the duplicate data correction matrix K has the greatest impact on the prediction effect, which proves that the repeated data generated by the symmetry problem of the proposed similarity tensor quantization method does have an enormous influence on the prediction effect. The second most influential is the $\ell 1$-norm, the number of features is huge (each sample has 48828 features) since the features are paired into a similarity matrix, and $\ell 1$-norm can make unimportant feature weights zero to assist select important features, enhance generalisability, robustness and stability for model prediction. The third influential is the generalised temporal correlation term, which demonstrated that combining the patient's current and earlier symptoms could improve the model's performance and conform to prior knowledge of real-world applications.

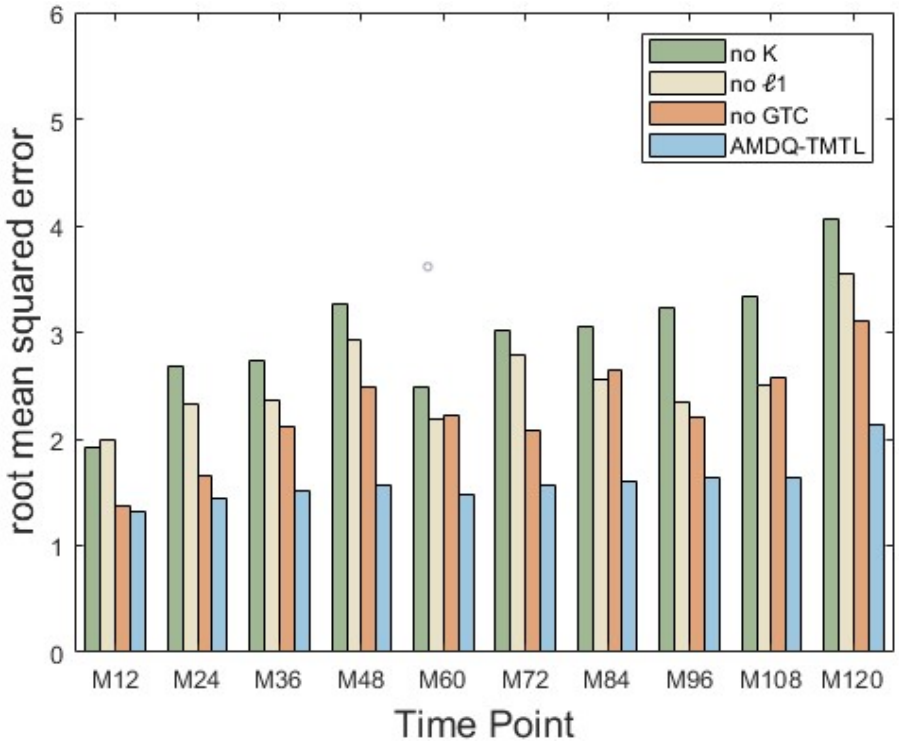


Figure 5.1: Ablation studies for proposed AMDQ-TMTL approach for MMSE prediction in various time points with BL and M06 MRI data.

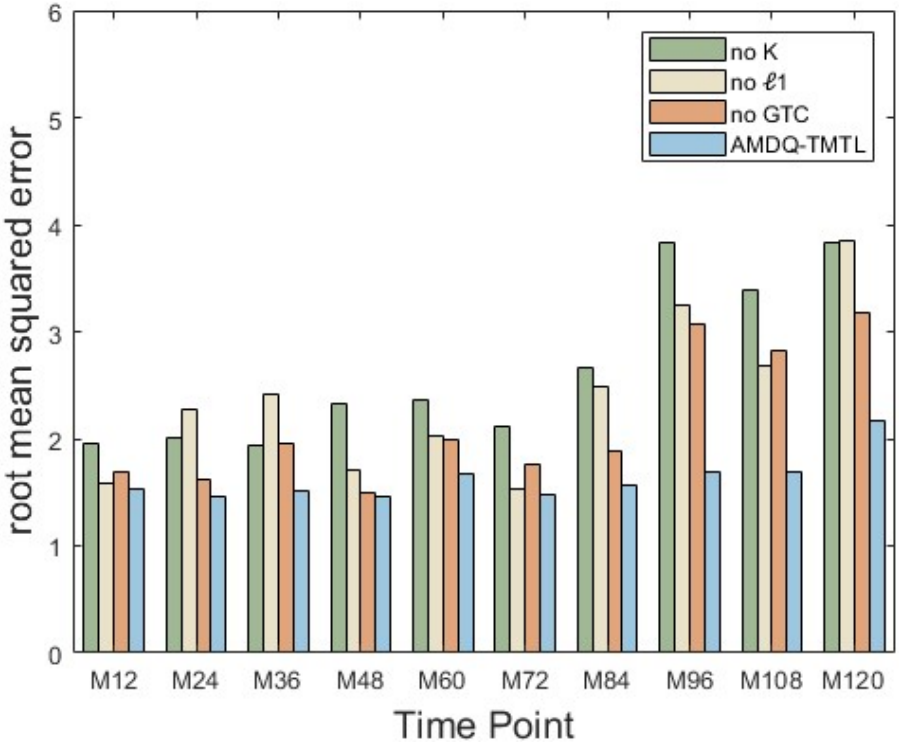


Figure 5.2: Ablation studies for proposed AMDQ-TMTL approach for ADAS-Cog prediction in various time points with BL and M06 MRI data.

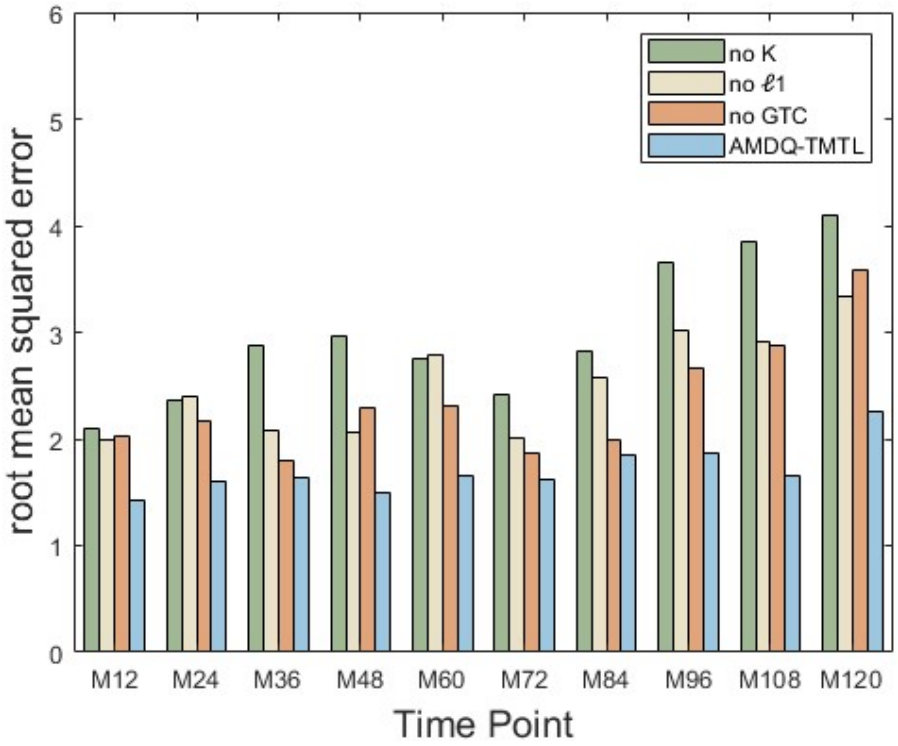


Figure 5.3: Ablation studies for proposed AMDQ-TMTL approach for RAVLT TOTAL prediction in various time points with BL and M06 MRI data.

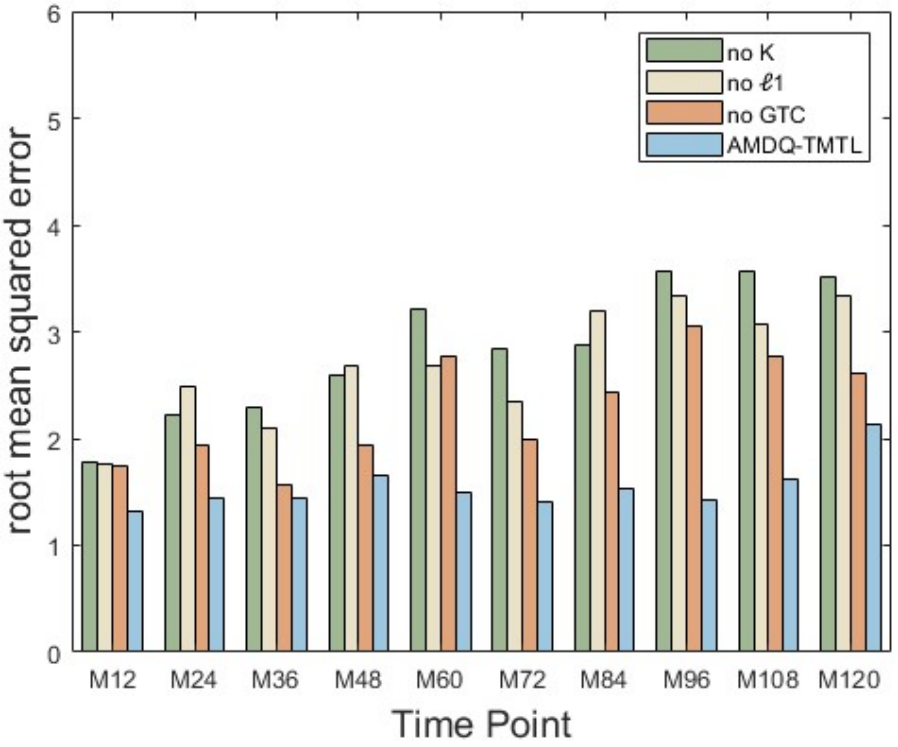


Figure 5.4: Ablation studies for proposed AMDQ-TMTL approach for FLU ANIM prediction in various time points with BL and M06 MRI data.

To compare the performance of different correlation calculation approaches (Cosine similarity, Euclidean distance and Mahalanobis distance) with our proposed AMDQ approach to the proposed multi-dimensional tensor multi-task learning regression model for disease progression prediction. Figures 5.5, 5.6, 5.7 and 5.8 demonstrate the average results of 20 independent experiments for the cognitive scores MMSE, ADAS-Cog, RAVLT TOTAL and FLU ANIM respectively, which utilise BL and M06 MRI data to predict cognitive scores at future time points.

The results demonstrate that the multi-dimensional tensor multi-task learning regression model with AMDQ approach has a lower rMSE than the mainstream correlation calculations for all individual time points. This may be due to the fact that the proposed AMDQ approach contains information on both the magnitude and direction of structural variations in brain biomarkers, which allows for the inclusion of more comprehensive information on brain structural variation in the disease progression prediction process.

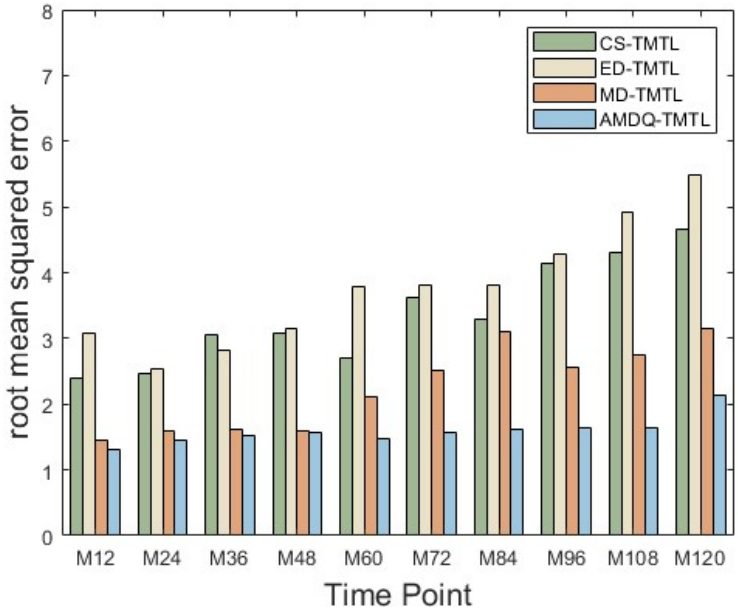


Figure 5.5: Comparison of different similarity calculation approaches on the MMSE prediction performance for multi-dimensional tensor multi-task learning.

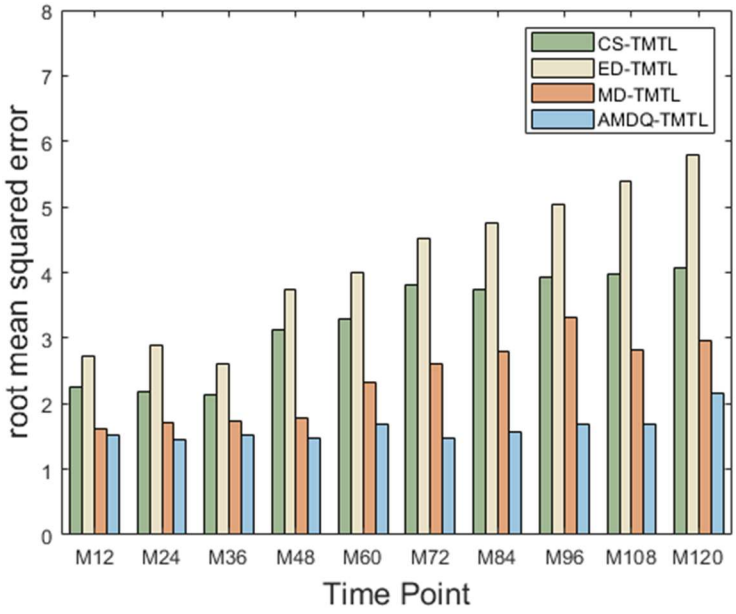


Figure 5.6: Comparison of different similarity calculation approaches on the ADAS-Cog prediction performance for multi-dimensional tensor multi-task learning.

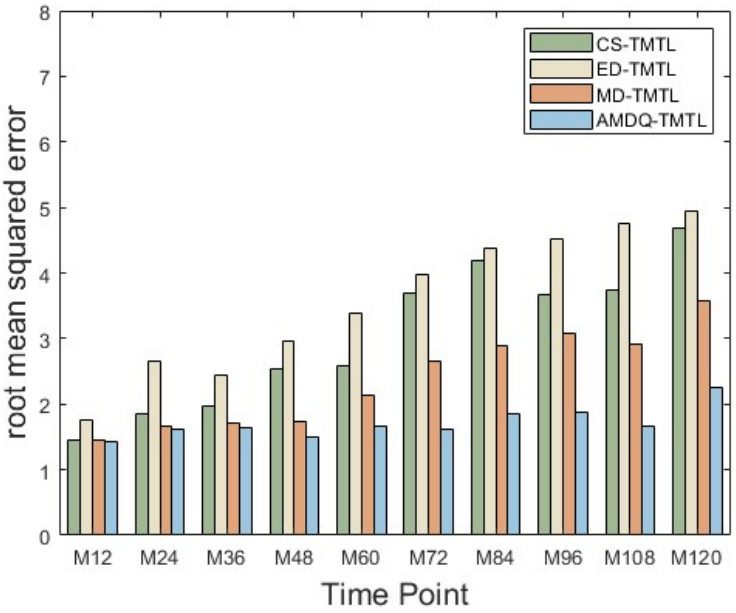


Figure 5.7: Comparison of different similarity calculation approaches on the RAVLT TOTAL prediction performance for multi-dimensional tensor multi-task learning.

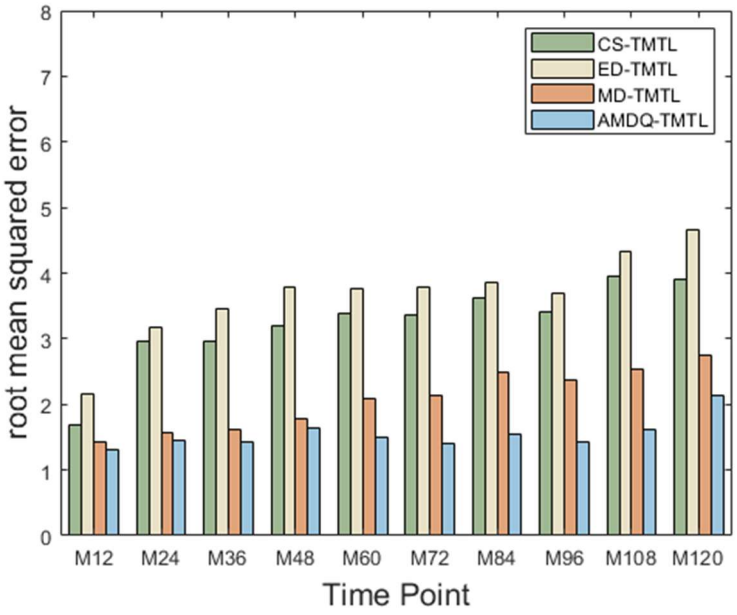


Figure 5.8: Comparison of different similarity calculation approaches on the FUN ANIM prediction performance for multi-dimensional tensor multi-task learning.

5.8 Clinical application

In clinical scenarios, the proposed AD progression prediction algorithms and models can be utilised to assist clinicians in making precise diagnostic and treatment decisions. The following are the applications of the proposed AD progression prediction algorithms and models in clinical scenarios: 1) Early diagnosis: the AD progression prediction model can make early diagnosis of patients based on their clinical characteristics, such as symptoms and imaging manifestations. Early diagnosis helps patients to receive timely treatment, mitigate disease progression and improve quality of life. 2) Disease prediction: AD progression prediction algorithms and models can predict future disease progression based on patients' clinical data, such as predicting the survival and degree of dementia of AD patients. This can assist clinicians in adjusting treatment plans and slowing disease progression in a timely manner. 3) Monitoring the efficacy of treatment: AD progression prediction model can track the progress of patients' disease and assist clinicians in assessing the efficacy of treatment. This can

help clinicians adjust the treatment plan in time to improve the treatment effect. 4) Individualized treatment: AD progression prediction algorithms and models can provide individualized treatment plans based on patients' clinical data. For example, the model can predict the patient's response to a certain medicine, thus assisting clinicians in selecting the optimal treatment plan. 5) Health management: the analysis of health data can predict the risk of AD for a patient. This can assist patient to take preventive actions to reduce the risk of AD. 6) Pathology analysis: AD progression prediction model can analysis patients' pathology data, including brain images, blood and cerebrospinal fluid biomarkers, etc. It assists clinicians in gaining a comprehensive understanding of a patient's condition, making accurate diagnostic and treatment decisions. 7) Translational medicine research: it can be utilised to investigate the effects and impact of different therapeutic intervention programs, providing reliable evidence for clinical practice. 8) Data mining: medical databases can utilise the proposed algorithms and models to mine and analyse large amounts of clinical data to discover potential disease risk factors, treatment effects, etc., providing accurate and effective support for medical researches and clinical practices.

The proposed algorithm and model for AD progression prediction in clinical scenarios has the following advantages: 1) Improved early diagnosis: it can help clinicians to diagnose AD early, including MCI and early AD lesions. It is important for the treatment and management of patients. Traditionally, the diagnosis of AD is typically diagnosed when clinical symptoms appear and progress to later stages, which often has caused irreversible brain damage and is relatively poorly treated. Therefore, being able to accurately diagnose and predict AD at an early stage of disease progression assists clinicians in taking early interventions to improve treatment outcomes and quality of life. 2) Improve the accuracy and efficiency of clinical decision-making: it can assist clinicians to diagnose and treat patients in more precise manner, improve the accuracy and efficiency of clinical decision-making, and provide better medical services to patients. 3) Promote the development of precision medicine: it can build personalized

treatment plans by analysing a large number of medical records and genomics data, improve treatment effects, and promote the development of precision medicine. 4) Improve the efficiency of clinical research: it can be utilised in clinical research to facilitate researchers to identify potential medicine targets and treatment strategies, as well as to enhance the efficiency and precision of research. 5) Reduce medical costs: it can improve the efficiency of medical resources utilisation, reduce non-essential tests and treatments, thus reducing medical costs and providing affordable medical services for patients. 6) Saving medical resources: it can assist clinicians to detect diseases at an early stage and intervene in time, thus reducing the cost of later treatment and management to save medical resources. 7) Improving the life experience of patients: it can provide patients with intelligent and convenient medical services, which helps patients obtain a better life experience and quality. 8) Promote the development of medical intelligence and digitalization: it is an important part of the development of medical intelligence and digitalization, which can enhance the efficiency of medical resources utilisation, reduce medical costs and improve medical quality. 9) Improve self-management and self-monitoring ability of patients: it can provide patients with greater self-management and self-monitoring tools to assist patients to have better control of their conditions and health status, thus better managing and controlling diseases. 10) Provide better care and support for families: it can help families better understand patients' conditions and health status and provide scientific, systematic care and support, thus reducing the burden and pressure on families. 11) Provide a data foundation for scientific research: it can contribute to scientific research by providing an important data foundation, assisting researchers to understand more deeply the mechanisms of disease occurrence and treatment principles, thus promoting the development of scientific research.

5.9 Summary

In this chapter, the multi-dimensional tensor multi-task learning regression algorithm and model are developed to describe the utilisation of multi-task learning algorithms combined with tensor-based data for AD progression prediction. The design goal of the algorithm and model is to enhance the accuracy, stability and interpretability of AD progression prediction in medical small dataset scenarios utilising multi-task learning ideas combined with quantitative multi-dimensional tensor of spatio-temporal structural variation information of MRI brain biomarkers from the Chapter 4. The algorithm and model are detailed in three modules: task and task relationship definition for tensor multi-task learning, algorithmic regularisation term and tensor multi-task learning regression algorithms. The components of the algorithm and model are analysed and defined. Then, the section presents the experimental configurations and processes required for the multi-dimensional tensor multi-task learning regression algorithm and model, as well as the testing and analysis of the algorithm and model results. Specifically, our quantitative approach (amalgamated magnitude-direction quantification) considers the magnitude and directional correlations of structural variation between brain biomarkers and quantifies them as third-order tensors to address the monotonic data form problem, and tensor multi-task learning regression utilises tensor latent factors as multi-task relationships to share knowledge and enhance model generalisation to address small dataset problem. Experimental results demonstrate that the proposed approach can be utilised to recognize brain structural variation differences in AD, CN and MCI individuals, with the ability to predict and diagnose AD progression, and requires only MRI data from patients to achieve superior prediction performance.

Chapter 6

Interpretability of algorithm results

This chapter presents the interpretability analysis from various aspects for the experimental results of the proposed multi-dimensional tensor multi-task learning regression algorithm. Interpretability refers to the readability and transparency of the algorithm, results and decision-making process. In this research, the readability and transparency of the algorithm and the decision-making process are ensured in the methodological design, on the basis of which this chapter interprets in detail the various parts of the algorithmic experimental results from both computer science and medical research perspectives. Section 6.1 investigates the latent factors of brain biomarkers derived from multi-dimensional tensor multi-task regression (AMDQ-TMTL) to recognize brain regions affected by AD progression. Section 6.2 explores and analyses the important relative structural variation correlations between brain biomarkers derived from experiments, which can be utilised to predict AD progression and can be exploited as potential indicators for early AD identification. Section 6.3 presents detailed analysis of potential indicators for AD early identification in terms of information on early relative structural variation between brain biomarkers corresponding to various specific disease progression time points.

6.1 Analysis of biomarker latent factors

We investigated the brain biomarker latent factors derived from multi-dimensional tensor multi-task regression (AMDQ-TMTL). The latent factor of each biomarker is a vector, and its elements represent the membership for each brain biomarker position to the given latent factor. Figure 6.1-6.3 shows the spatial distribution of top 20 rank of brain biomarker latent factors when the model predicts MMSE for time point M12, M24, M36, M48, M60, M72, M84, M96, M108 and M120. And Figure 6.4-6.6 shows the spatial distribution of top 20 rank of brain biomarker latent factors when the model predicts ADAS-Cog. And Figure 6.7-6.9 shows the spatial distribution of top 20 rank of brain biomarker latent factors when the model predicts RAVLT TOTAL. And Figure 6.10-6.12 shows the spatial distribution of top 20 rank of brain biomarker latent factors when the model predicts FLU ANIM. It can be seen from the figure that for different time points, the latent factors have different spatial distributions, which indicates that they capture different aspects of the spatial variability in the data. Visualisation was performed by the toolkit of BrainNet Viewer [305]. BrainNet Viewer can automatically arrange the value of the input latent factors to an appropriate range (radius: 2-7 mm), therefore we can more intuitively observe the difference in the size of the latent factors. The color represents the brain biomarker categories of latent factors.

6.1.1 For MMSE prediction targets

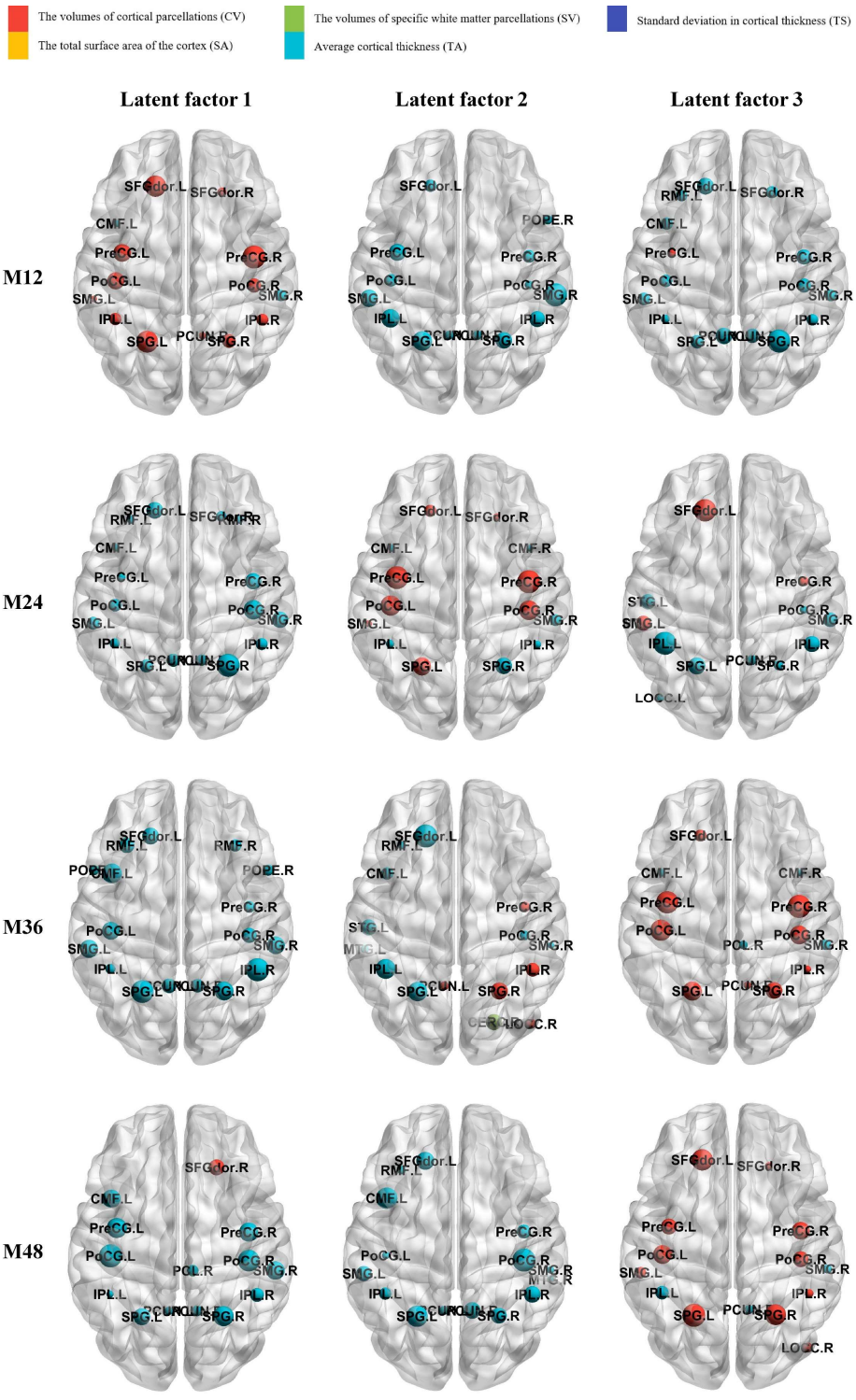


Figure 6.1: Spatial distribution of the biomarker latent factors for AD time points M12, M24, M36 and M48 with MMSE prediction targets. The abbreviations for the brain regions are in Appendix C.

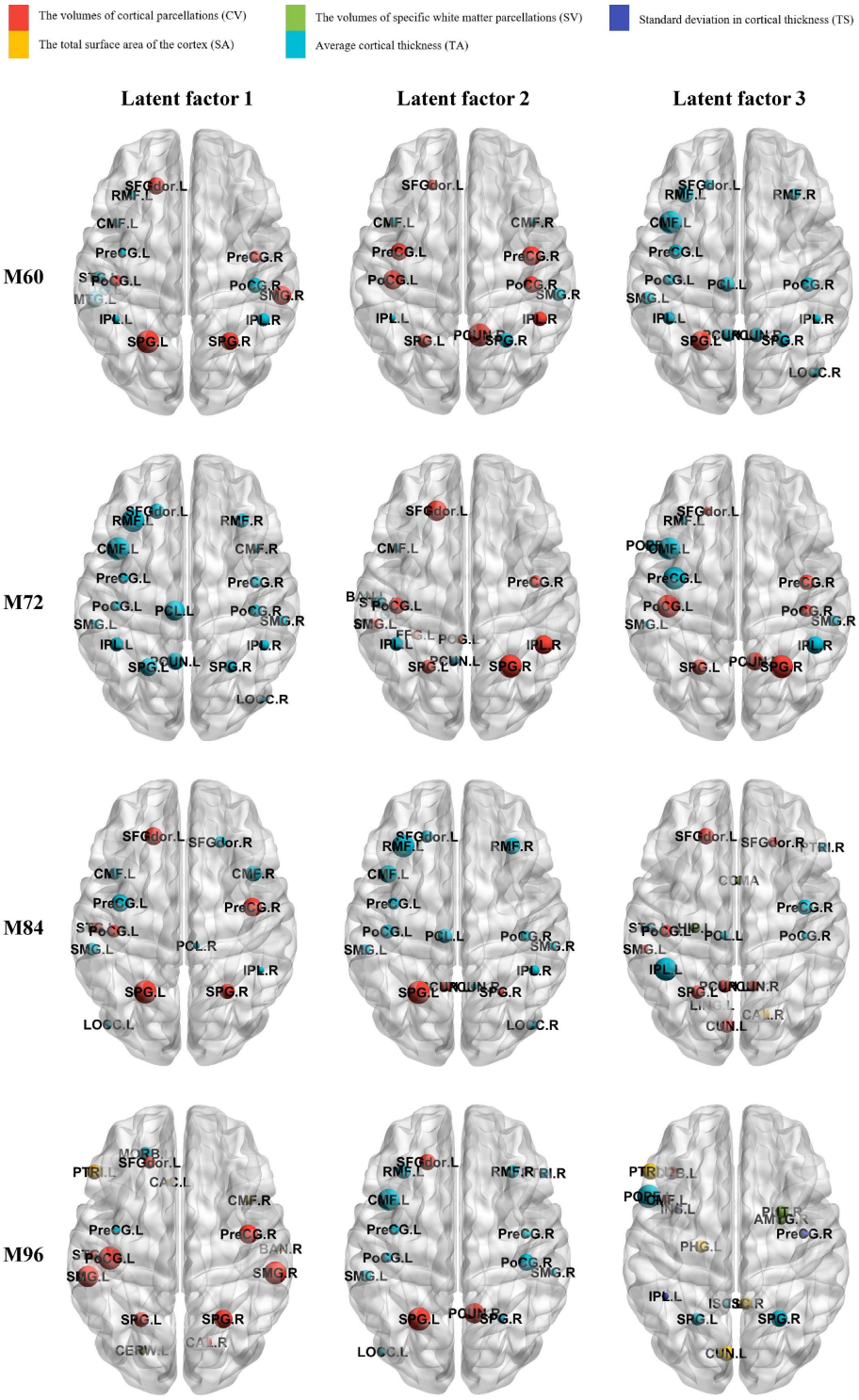


Figure 6.2: Spatial distribution of the biomarker latent factors for AD time points M60, M72, M84 and M96 with MMSE prediction targets. The abbreviations for the brain regions are in Appendix C.

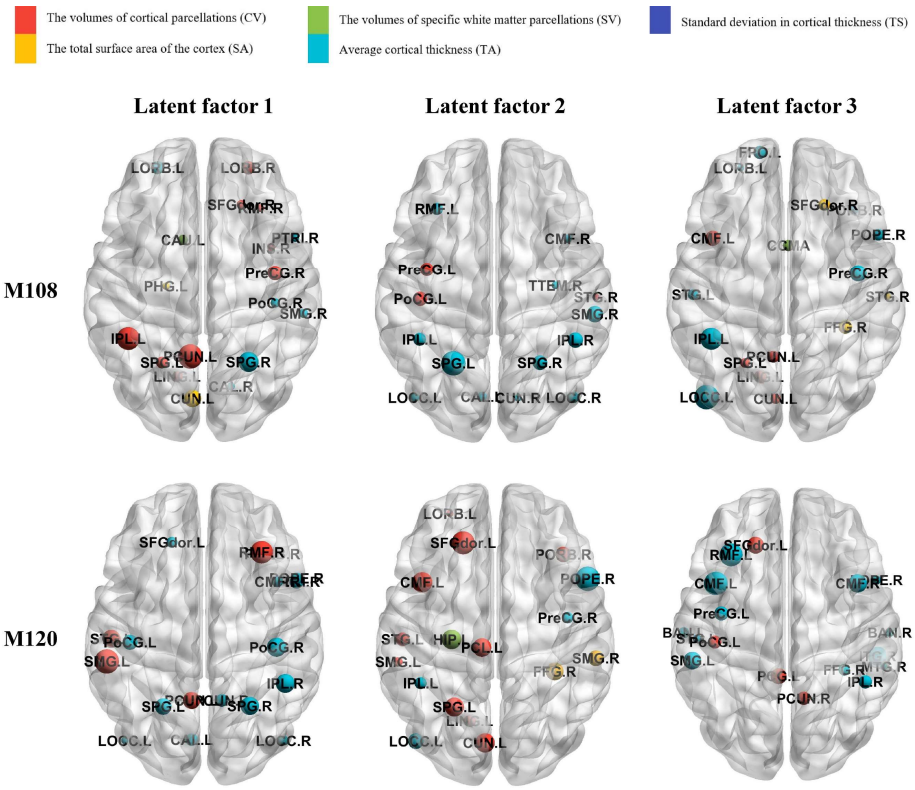


Figure 6.3: Spatial distribution of the biomarker latent factors for AD time points M108 and M120 with MMSE prediction targets. The abbreviations for the brain regions are in Appendix C.

For MMSE prediction targets, we observed that the three latent factors were dominated by cortical partition volume and mean cortical thickness at all time points, while total cortical surface area and specific white matter partition volume were also captured partially at later stages of disease (M96, M108 and M120).

6.1.2 For ADAS-Cog prediction targets

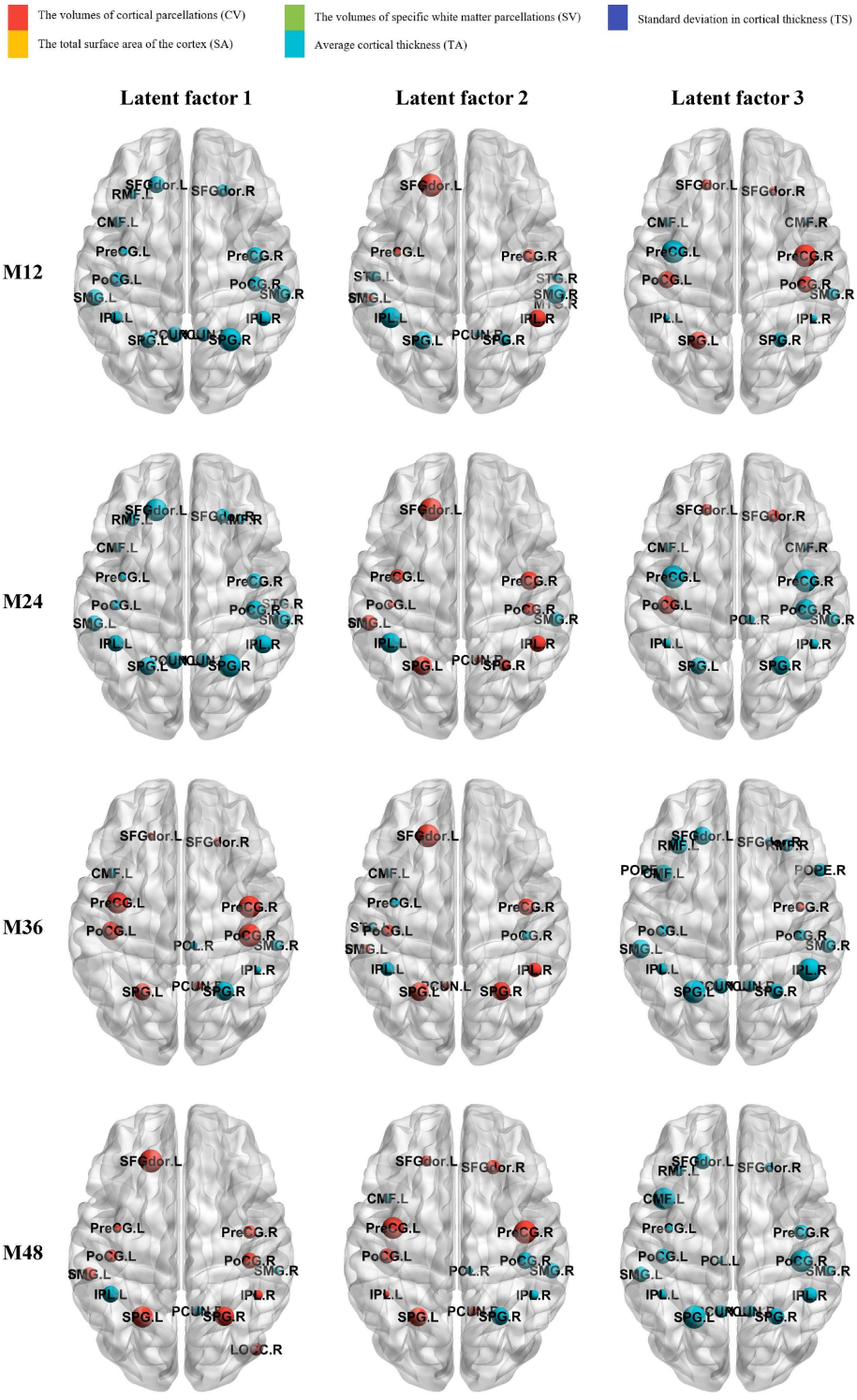


Figure 6.4: Spatial distribution of the biomarker latent factors for AD time points M12, M24, M36 and M48 with ADAS-Cog prediction targets. The abbreviations for the brain regions are in Appendix C.

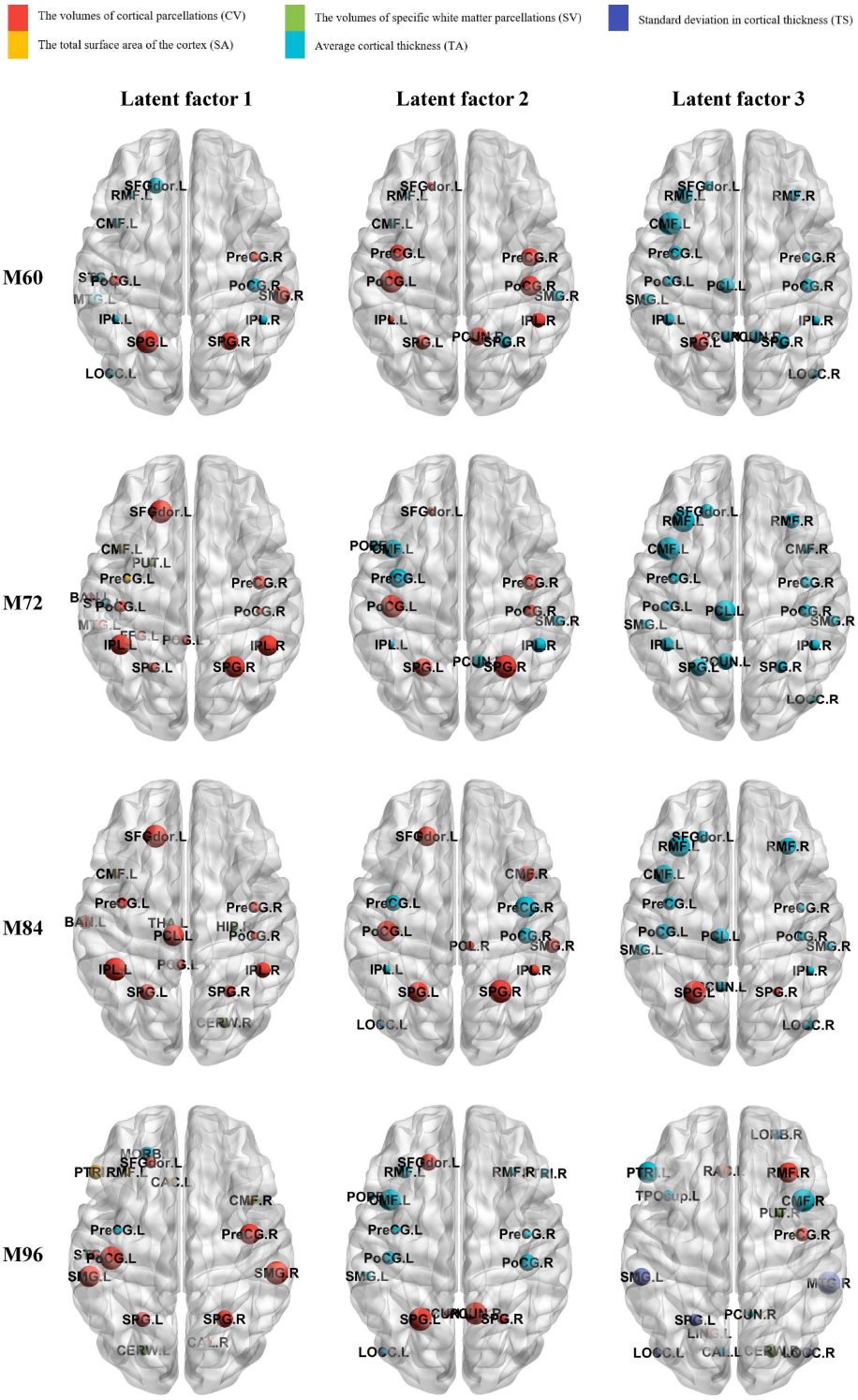


Figure 6.5: Spatial distribution of the biomarker latent factors for AD time points M60, M72, M84 and M96 with ADAS-Cog prediction targets. The abbreviations for the brain regions are in Appendix C.

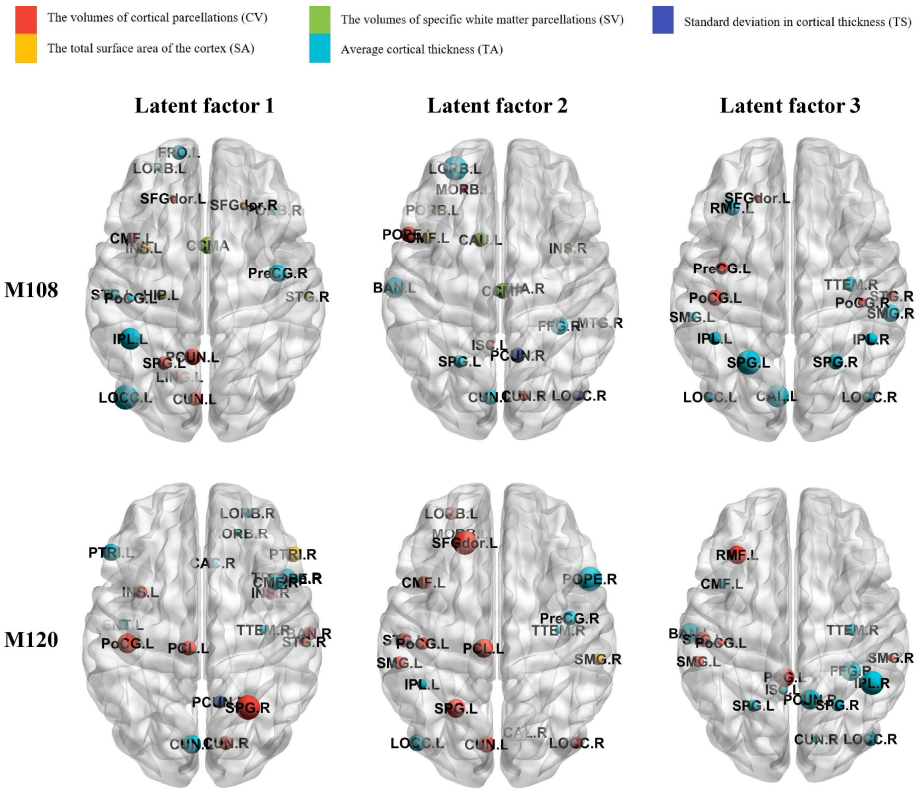


Figure 6.6: Spatial distribution of the biomarker latent factors for AD time points M108 and M120 with ADAS-Cog prediction targets. The abbreviations for the brain regions are in Appendix C.

For ADAS-Cog prediction targets, the distribution of brain biomarker latent factors is similar to that of the MMSE prediction target, which can be attributed to the fact that the ADAS-Cog test covers all aspects of the MMSE test and includes additional tests of executive function, visuospatial ability and abstract thinking. Specifically, we observed that the three latent factors were dominated by cortical partition volume and mean cortical thickness at all time points, while standard deviations of cortical thickness, total cortical surface area and specific white matter partition volume were also captured partially at later stages of the disease (M96, M108 and M120).

6.1.3 For RAVLT TOTAL prediction targets

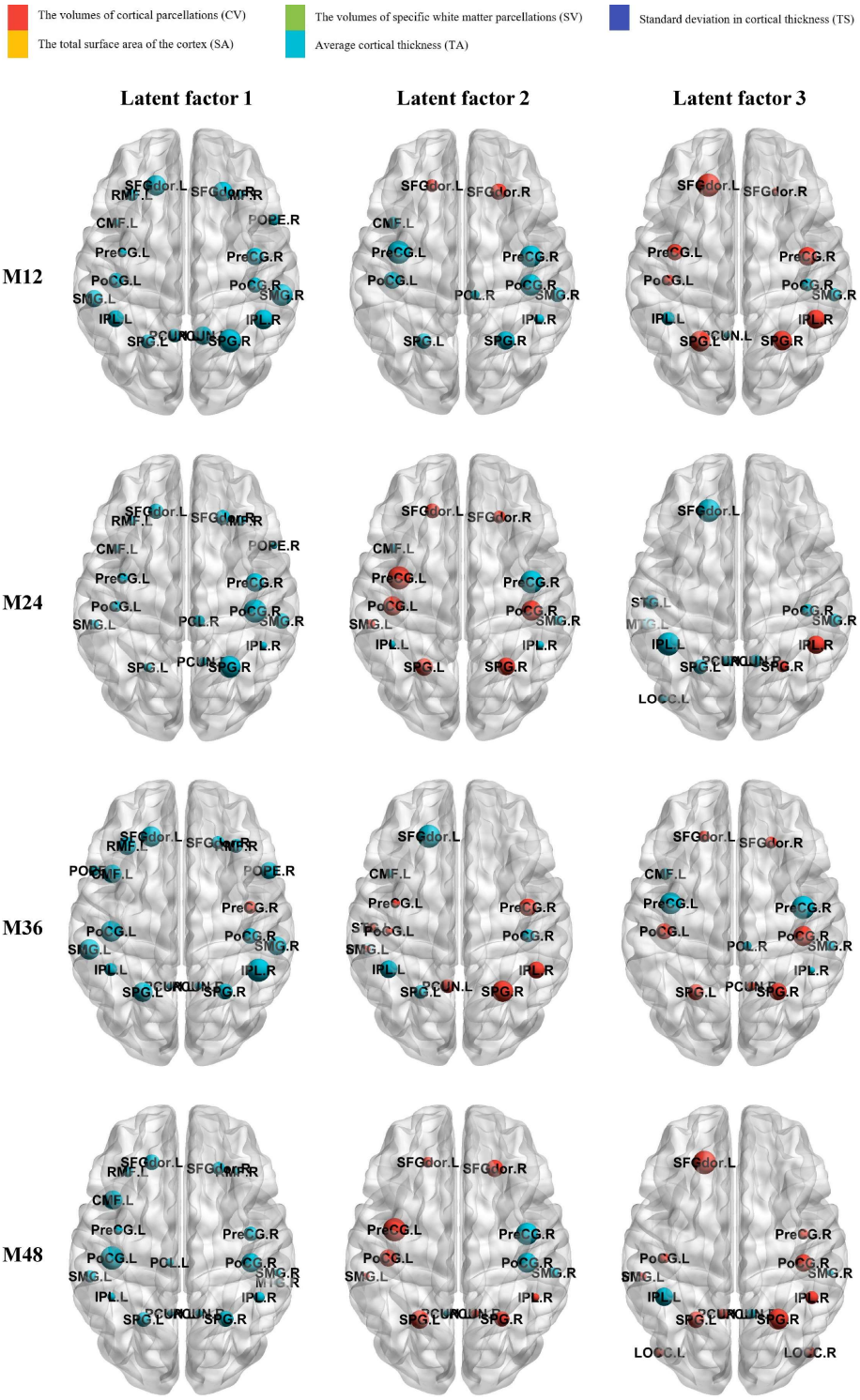


Figure 6.7: Spatial distribution of the biomarker latent factors for AD time points M12, M24, M36 and M48 with RAVLT TOTAL prediction targets. The abbreviations for the brain regions are in Appendix C.

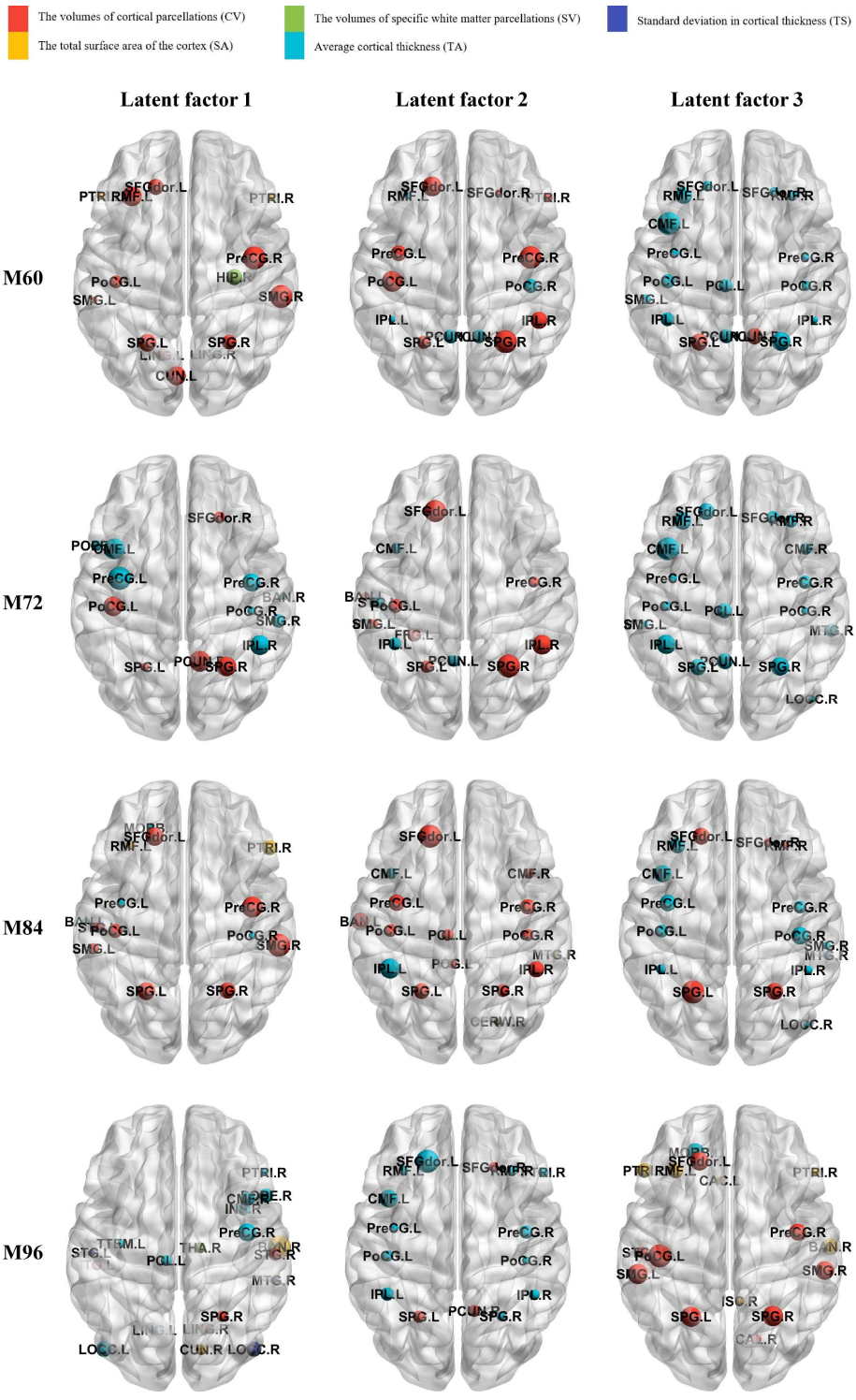


Figure 6.8: Spatial distribution of the biomarker latent factors for AD time points M60, M72, M84 and M96 with RAVLT TOTAL prediction targets. The abbreviations for the brain regions are in Appendix C.

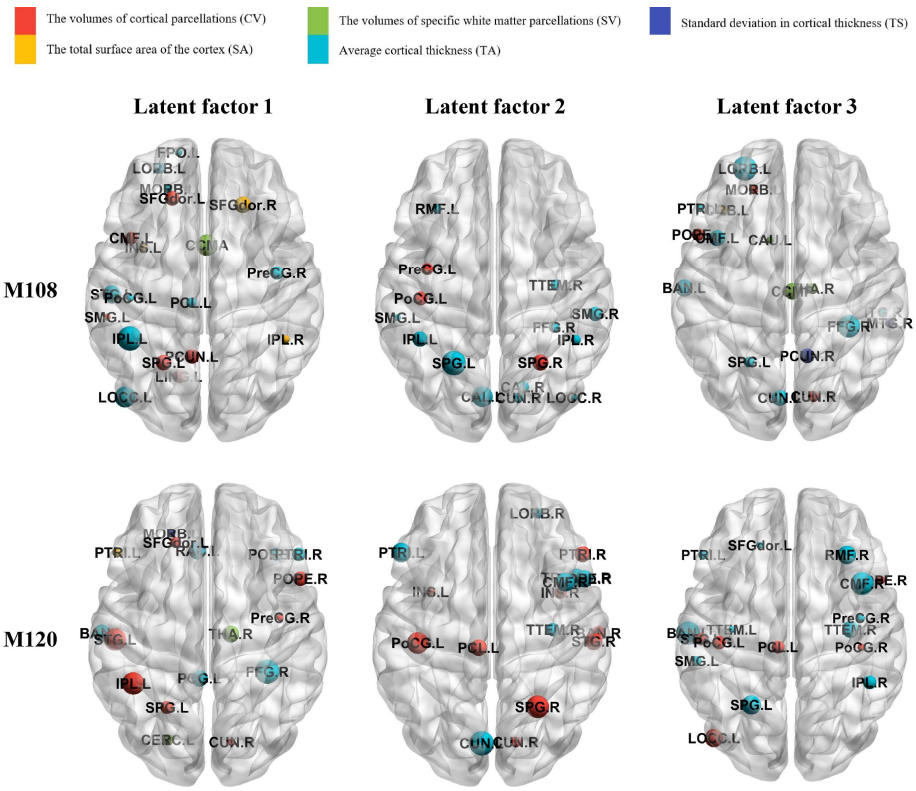


Figure 6.9: Spatial distribution of the biomarker latent factors for AD time points M108 and M120 with RAVLT TOTAL prediction targets. The abbreviations for the brain regions are in Appendix C.

For RAVLT TOTAL prediction targets, We observed that the three latent factors were dominated by cortical partition volume and mean cortical thickness at all time points, with mean cortical thickness being captured substantially in the early stages of disease (M12, M24 and M36), and cortical partition volume being captured substantially in the middle stages of disease (M48, M60, M72 and M84), and from middle to late stages of disease (M84, M96, M108 and M120), the total cortical surface area and the specific white matter compartment volumes are captured partially in addition to cortical compartment volumes and mean cortical thickness.

6.1.4 For FLU ANIM prediction targets

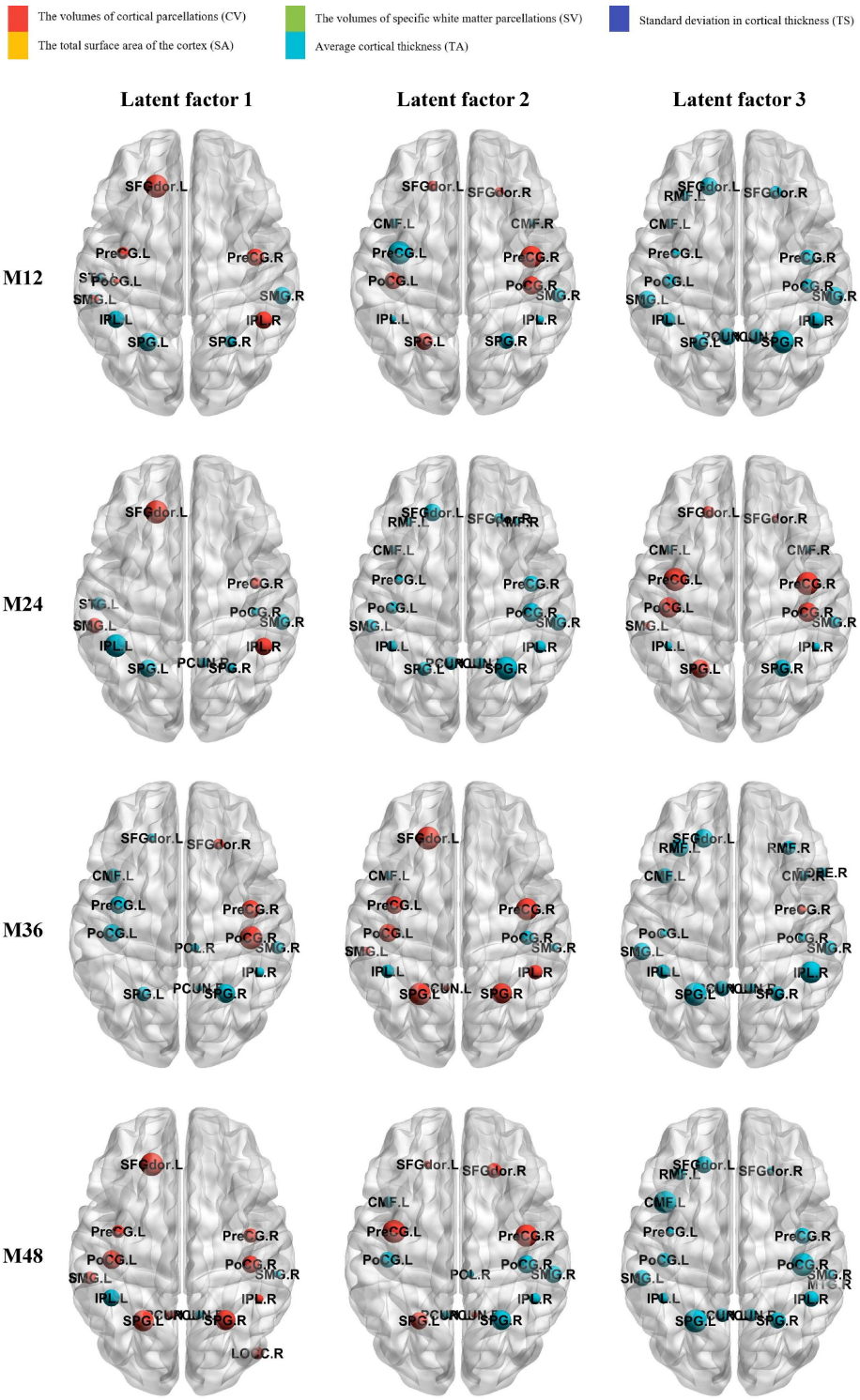


Figure 6.10: Spatial distribution of the biomarker latent factors for AD time points M12, M24, M36 and M48 with FLU ANIM prediction targets. The abbreviations for the brain regions are in Appendix C.

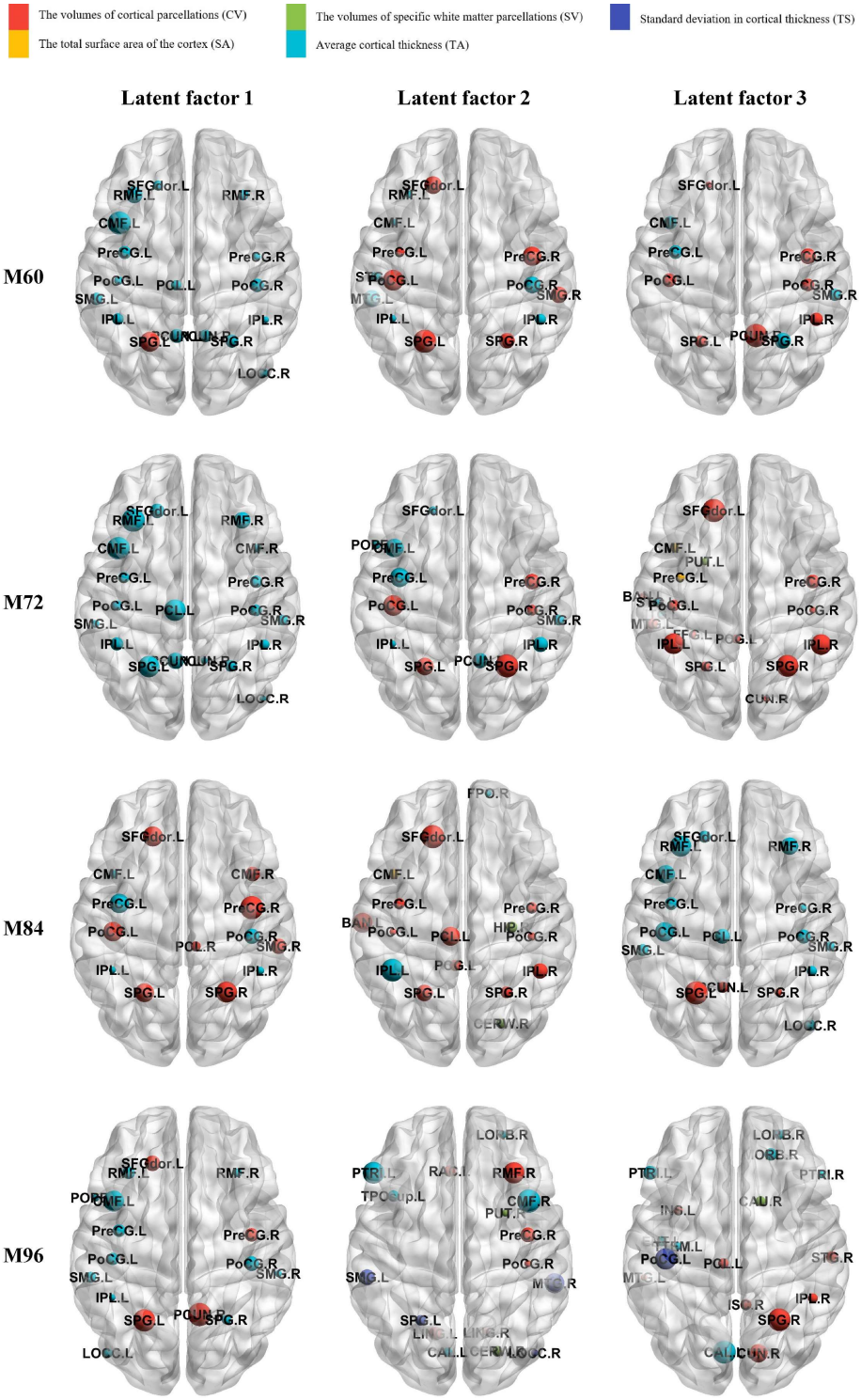


Figure 6.11: Spatial distribution of the biomarker latent factors for AD time points M60, M72, M84 and M96 with FLU ANIM prediction targets. The abbreviations for the brain regions are in Appendix C.

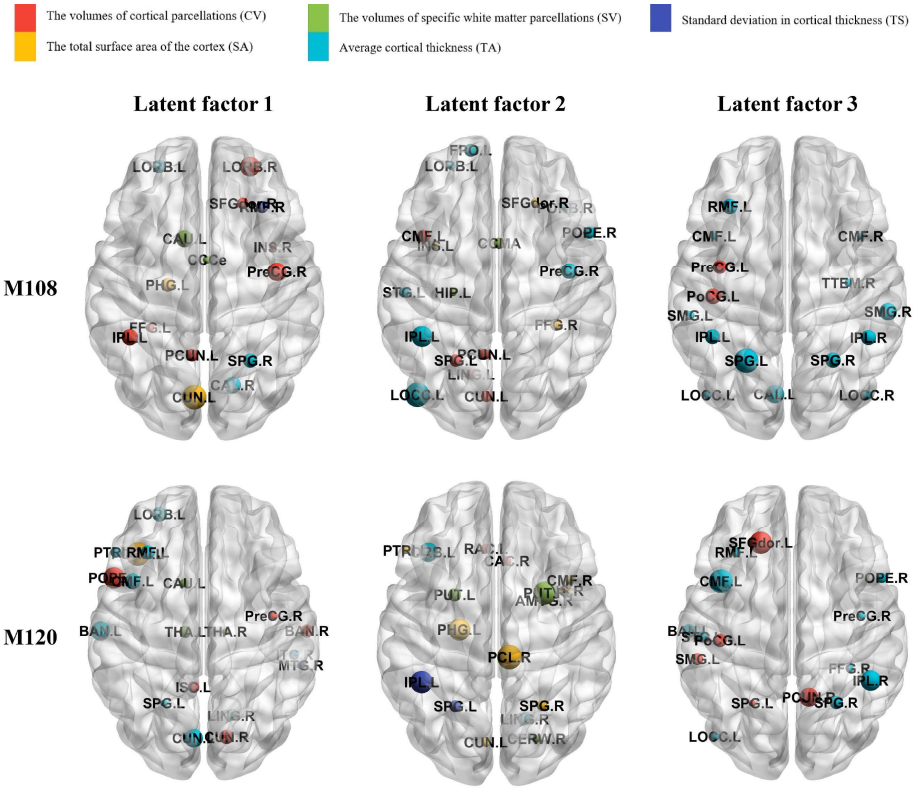


Figure 6.12: Spatial distribution of the biomarker latent factors for AD time points M108 and M120 with FLU ANIM prediction targets. The abbreviations for the brain regions are in Appendix C.

For FLU ANIM prediction targets, we observed that all three latent factors were dominated by cortical partition volume and mean cortical thickness at all time points, but differed from the three previous cognitive score tests in that cortical partition volume was captured at a higher proportion than average cortical thickness during early and mid-stage disease (M12, M24, M36, M48, M60, M72 and M84). In the later stages of the disease (M96, M108 and M120) the standard deviation of cortical thickness, the total cortical surface area and the volume of specific white matter subdivisions are captured partially, similarly to the three previous cognitive scores.

6.2 Interpretability of structural variation correlations between brain biomarkers

The interpretability of approach and results is as crucial in medical research as model performance. Because there is currently no cure for AD, the key to present treatment is early detection and prevention of the disease. Therefore, identifying significant brain biomarker structural variation correlations in early MRI data can assist clinicians recognize individuals with suspected AD for early prevention. To facilitate readability, all the top 10 brain biomarker structural variation correlations of the proposed AMDQ-TMTL approach for all time points and prediction targets are included in Appendix D. For instance, the top 10 brain biomarker structural variation correlations of the proposed AMDQ-TMTL approach are shown in descending order of the weighted parameter values by MMSE prediction at time point M12 in Table 6.1. Higher values indicate a greater impact on the final prediction. The important brain biomarker correlations identified can be utilised as potential indicators for early identification of AD.

6.2.1 For MMSE prediction targets

Table 6.1: The top-10 rank brain biomarker correlations in time point M12 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.5301
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.5244
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.5145
CTA. of R.Precuneus - Vol(C). of R.Supramarginal	0.4916
Vol(C). of R.Precentral - CTA. of R.Supramarginal	0.4889
CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal	0.4679
CTA. of R.SuperiorTemporal - CTA. of L.SuperiorFrontal	0.4582
CTA. of L.InferiorParietal - CTA. of L.Precentral	0.4479
Vol(C). of L.SuperiorFrontal - Vol(C). of L.SuperiorParietal	0.4289
Vol(C). of R.Supramarginal - Vol(C). of L.Precuneus	0.4261

We discovered that brain structural variation correlations between several brain biomarkers were significant at all time points or most time points (nine out of ten time points). Specifically, they are Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal, CTA. of R.Precuneus - Vol(C). of R.Supramarginal, CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal and CTA. of L.InferiorParietal - CTA. of L.Precentral.

For Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal, the right inferior parietal lobule is involved in the perception of emotion and interpretation of sensory information in facial stimuli [306]. The right inferior parietal lobule is associated with language, mathematical operations and bodily imagery. The structural variation correlation between cortical volume and mean cortical thickness in the right inferior parietal lobule of the brain can be a factor in the low verbal vocabulary, naming difficulties and impaired numeracy symptoms of AD patients.

For CTA. of R.Precuneus - Vol(C). of R.Supramarginal, the anterior part of the right precuneus is associated with self-awareness (self-personality perception, introspection) and the posterior part with episodic memory, while the other part of the right precuneus is associated with visuospatial processing [307]. The right supramarginal gyrus contains part of Wernicke's area and is closely related to language function. It is a part of the somatosensory cortex, which interprets tactile signals and is also associated with the perception of space and limb position. It is involved in the perception of other people's gestures and postures and thus is a part of the mirror neuron system [308]. This correlation can be a factor in the appearance of temporal orientation deficits, ability to orientate to places and people, difficulty orienting to geographical locations, poor visuospatial ability of complex structures in AD patients.

For CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal, the left postcentral gyrus is the site of primary somatosensory cortex, which is the nerve centre of somatosensory system, and this area perceives a variety of bodily sensations from the body, including

touch, pressure, temperature and pain [309]. The left superior parietal gyrus is associated with the brain's spatial orientation function, which remembers the location of objects in space. It plays an important role in manipulation and resetting of information in working memory. In addition, the left superior parietal gyrus receives a large amount of visual information and sensory information from hand, which can form a sense of object shape, roughness, size, material, etc [310]. The left postcentral gyrus processes sensory information mainly on the right side of body, while the left superior parietal gyrus is typically associated with mathematical and computational abilities, they perform important functions in spatial sensation and cognition. This correlation can be a factor in behavioural disturbances, progressive decline in daily living skills and limb stiffness in AD patients.

For CTA. of L.InferiorParietal - CTA. of L.Precentral, the left precentral gyrus is located on the lateral surface of each frontal lobe, it runs parallel to the central sulcus and extends to the precentral sulcus [311]. The primary motor cortex is located in the left precentral gyrus and is responsible for controlling voluntary motor movements. Both left Inferior parietal and left precentral are parts of the human cerebral cortex that work in tandem to control the movement and sensation of the body's muscles. They perform major functions in muscle movements, language processing, attention and spatial perception. This correlation can be a factor in the occurrence of spatial memory, working memory, word and semantic memory, difficulties in comprehending and utilising language, task performance skills and cognitive control problems in AD patients.

In addition to brain structural variation correlations between several brain biomarkers that were significant at all or most time points (nine out of ten), we identified two brain structural variation correlations that are significant in the early stages of AD progression and that can be utilised to identify whether a patient is in the AD early stages. Specifically, they are Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal and

Vol(C). of R.Precentral - CTA. of R.Supramarginal.

For Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal, we have mentioned above that the right postcentral gyrus and the left superior parietal gyrus are both located in the parieto-occipital region of the brain and they have important functions in sensory, motor and cognitive functions. Although the right postcentral gyrus and left superior parietal gyrus have distinct sensory and cognitive functions, they are often required to collaborate with each other in processing complex perceptual information and cognitive tasks. Researches have demonstrated that there is a large number of neural connections between the two regions and that they have similar activation patterns in cognitive tasks, thus it can be hypothesized that they are interdependent in sensory and cognitive aspects [312][313]. This correlation can be a factor in the early stages of AD patients in terms of cognitive impairment, behavioural problems and motor control disorders.

For Vol(C). of R. Precentral - CTA. of R.Supramarginal, the right precentral gyrus and the right supramarginal gyrus interact in 1) The control and perception of bodily movements. 2) Language and reading: the right supramarginal gyrus have an important function in language processing, including reading and writing. The right precentral gyrus can be involved in reading and language production. 3) Attention and cognitive control: the right supramarginal gyrus have an essential contribution to cognitive control and attention, while the right precentral gyrus requires attentional regulation in the control of body movements. Therefore, there can be an association between these two areas in terms of cognitive control and attention. 4) Motor imagery: the right precentral gyrus is activated when people imagine performing a certain movement. The right supramarginal gyrus can be involved in the execution and control of motor imagery. 5) Social emotion: the right supramarginal gyrus can have an essential function in the perception and expression of social emotion. The right precentral gyrus is implicated in the expression of social emotions, such as facial expressions and gestures. 6) Cognitive flexibility: it refers to the ability of people to adapt and adjust

their thinking and behaviour in response to different situations. This correlation can be a factor in social cognitive impairment and behavioural changes, attention and executive function impairment, diminished working memory and learning ability, language and communication impairment, spatial cognition and orientation perception impairment, visual motion perception and visual-spatial processing impairment in AD patients.

6.2.2 For ADAS-Cog prediction targets

We discovered that one brain structural variation correlation between brain biomarkers were significant at most time points (nine out of ten time points). Specifically, it is Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal. The left inferior parietal lobule and left superior parietal gyrus on the left side of brain are two areas of the parietal lobe of brain that are associated with each other in the following aspects [306][310][314]: 1) Spatial cognition and visuospatial processing: the left inferior parietal lobule is responsible for processing information on the shape, size, orientation and position of objects, while the left superior parietal gyrus is responsible for processing visuospatial information and the position of the body in space. The two areas work in tandem to enable accurate perception and processing of objects and space in the surrounding environment. 2) Hand-eye coordination: these two regions operate in concert when performing fine hand manipulation, such as writing, playing the piano or playing ball, etc. The left inferior parietal lobule collects visual information from the eyes and works in concert with the motor control region in the left superior parietal gyrus to regulate the movements of the arm and hand in order to achieve accurate hand-eye coordination. 3) Memory and attentional control: the left inferior parietal lobule is associated with working memory and attentional control, particularly in the areas of spatial cognition and object recognition. The left superior parietal gyrus is related to attention and inhibition control, assisting to filter and ignore irrelevant information in order to better focus on and process information of interests. 4) Mathematical ability: the left inferior parietal lobule is one of the key domains of mathematical ability, while

the left superior parietal gyrus is involved in aspects relating to spatial cognition and mathematical computation. These two areas cooperate to support complex mathematical operations such as computation, geometric reasoning and algebraic operations. 5) Language processing: the left inferior parietal lobule is involved in language processing, particularly in the understanding of the syntax and semantics of language. The left superior parietal gyrus is involved in phonological and intonation processing. These two areas contribute to better understanding and expression of language by working in tandem. 6) Attention and awareness: the left inferior parietal lobule is responsible for maintaining the stability and regulation of attention, while the left superior parietal gyrus is concerned with body sensation and body orientation. These two areas work in synergy to support the regulation of consciousness and attention. This correlation can be a factor in the memory loss, difficulty concentrating, reduced decision-making ability, language impairment, diminished ability to perform activities of daily living, and increased difficulty in self-care and social activities experienced by AD patients.

In addition to the brain structural variation correlation that was significant at most time points (nine out of ten), we identified two brain structural variation correlations that are significant in the early stages of AD progression and that can be utilised to identify whether a patient is in the AD early stages. Specifically, they are Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal and Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal.

For Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal, this correlation was similarly seen in MMSE, demonstrating its significant representation in the early stages of AD progression.

For Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal, this correlation appears similarly in the MMSE, but unlike the MMSE which indicates importance at all time

points, it only indicates significance in the early stages of AD for ADAS-Cog predictions. The possible reason for this is that due to the different detection methods and focus of MMSE and ADAS-Cog, the correlation highlights the importance of different aspects. The MMSE is a simple test that is widely utilised to assess brain function and cognitive ability. The test consists of a series of simple questions on memory, attention, orientation, language and visuospatial ability. The ADAS-Cog is a more detailed and comprehensive assessment tool for cognitive function, covering all aspects covered in the MMSE test and with the addition of more test items such as executive function, visuospatial ability and abstract thinking. Therefore, the MMSE is mainly intended to assess the degree of cognitive ability, especially in early diagnosis of diseases such as MCI and AD. On the other hand, the ADAS-Cog is a more comprehensive assessment tool for evaluating various aspects of cognitive function and is utilised to monitor disease progression and assess the treatment effectiveness. Therefore, this correlation has a prescient predictive ability for additive test components of the ADAS-Cog such as executive function, visuospatial ability and abstract thinking, and a global predictive ability for aspects of memory, attention, orientation, language and visuospatial ability.

6.2.3 For RAVLT TOTAL prediction targets

We discovered that one brain structural variation correlation between brain biomarkers were significant at most time points (nine out of ten time points). Specifically, it is CTA. of R.ParsTriangularis - CTA. of L.Postcentral. The right pars triangularis involves semantic processing of language [315]. The right pars triangularis and the left postcentral gyrus are both areas of the human cerebral cortex. Connections between the right pars triangularis and the left postcentral gyrus are primarily concerned with language processing and the transmission of tactile information from the face and oral cavity [316][317][318]. Language processing involves the activity of oral and facial muscles, and sensory information from these muscles comes from tactile receptors on the face and mouth, and information from these receptors is passed to the left

postcentral gyrus for processing. The information is then converted into language signals and processed via neural pathways connected to the right pars triangularis. In the production and processing of language, the right pars triangularis is responsible for the phonological production and grammatical processing of language, while the left postcentral gyrus is responsible for the processing of sensory information from the body, which includes tactile sensations and temperature, etc. Therefore, during speech production, the activity of the facial muscles produces tactile signals that are passed to the left postcentral gyrus for processing and then through the neural pathways to the right pars triangularis. Right pars triangularis translates this information into speech signals and performs grammatical analysis and processing. This correlation can be a factor in causing AD patients to have problems in understanding phonology, grammar and vocabulary.

In addition to the brain structural variation correlation that was significant at most time points (nine out of ten), we identified two brain structural variation correlations that are significant in the early stages of AD progression and that can be utilised to identify whether a patient is in the AD early stages. Specifically, they are Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal and Vol(C). of L.Precentral - Vol(C). of L.Precuneus.

For Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal, the right postcentral and left superior parietal areas of the brain are two important neural regions in the human brain that are critical in the processing and transmission of sensory information [312][319]. These two areas work synergistically in a number of cognitive tasks, for instance, simultaneous activation in the execution and perception of hand movements, visual-spatial attention, and language comprehension. This correlation can be a factor in the clinical manifestations of reduced sensory and spatial cognition, inattention and behavioural abnormalities in AD patients. For instance, patients experience a significant decline in spatial memory and spatial navigation in the early stages of AD,

along with a loss of control over bodily sensations. The right postcentral is the area of the brain responsible for the processing and integration of sensory information, while the left superior parietal is involved in the cognition of bodily sensations and spatial orientation. Damage to these regions in AD patients can result in loss of control of body sensation, manifested by symptoms such as numbness in the limbs, loss of sensation and balance.

For Vol(C). of L.Precentral - Vol(C). of L.Precuneus, The left precentral gyrus and the left precuneus are two distinct brain regions, and in researches of cognitive neuroscience field, the left precentral gyrus and left precuneus of brain are considered to potentially collaborate with each other when performing complex cognitive tasks [320][321]. For instance, the activity of both the precentral gyrus and precuneus increases when performing action observation tasks. This can reflect the synergistic interaction of these two regions when observing the actions of others. And when people perform a spatial memory task, both structural and functional connectivity between the precentral gyrus and precuneus are enhanced, suggesting that there can be mutual collaboration between these two brain regions in spatial memory. This correlation can be a factor in the motor dysfunction, progressive decline in daily living skills, severe impairment of near and distant memory, and reduced visuospatial ability of simple structures in AD patients.

6.2.4 For FLU ANIM prediction targets

We discovered that brain structural variation correlations between several brain biomarkers were significant at all time points or most time points (nine out of ten time points). Specifically, they are Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal, Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal and Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal.

For Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal, the superior frontal

gyrus is implicated in higher-order cognitive functions of the brain, particularly working memory [322][323]. The superior frontal gyrus on one side of the cerebral hemisphere is responsible for planning certain complex movements on the other side of body. The right superior frontal gyrus and the left superior frontal gyrus are both prefrontal areas of the cerebral cortex that perform important functions in cognitive and emotional functions. The left superior frontal gyrus is involved in language processing and executive control functions, while the right superior frontal gyrus is more associated with attention and working memory. This correlation can be a factor in the cognitive and behavioural deficits, attention defects, mood disorders and impaired memory function in AD patients.

For Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal, The left superior frontal gyrus is associated with speech and language-related cognitive functions, such as language production, comprehension and processing [324]. At the same time, this area is implicated in a number of other cognitive processes, such as spatial cognition and attention. This correlation can be a factor in the impaired higher cognitive functions, reduced decision-making, executive control and planning ability, reduced language and memory skills, and working memory impairment in AD patients.

For Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal, the right middle frontal gyrus is associated with literacy and numeracy, and is involved in cognitive processes such as decision-making, planning and executive control. Activation of both regions frequently occurs simultaneously during the performance of cognitive control tasks [312][325]. Both regions have a major impact on self-awareness and spatial perception in humans. Specifically, the right precuneus is regarded as the centre of the brain's self-awareness, assisting in building perceptions of the self by participating in the processing of emotion regulation and emotional experience, as well as integrating external and internal sensory information. The right middle frontal gyrus has a significant role in spatial cognition and attentional control, assisting in understanding information such as

spatial orientation and distance, as well as in attentional shifts and task performance. These two regions are closely linked in visual image processing and language processing as well. Specifically, the right precuneus is involved in the brain's higher-level processing of visual images and the integration of visuospatial information, while the right middle frontal gyrus functions in expression and comprehension of language, assisting in language processing and the construction of meaning. This correlation can be a factor to the cognitive control and executive function deficits, inability to calculate, reduced judgement, inability to analyse, think and judge events, difficulty in dealing with complex problems, impaired spatial orientation, reduced visuospatial ability of complex structures and severe impairment in processing problems and identifying similarities and differences in things experienced by AD patients.

In addition to brain structural variation correlations between several brain biomarkers that were significant at all or most time points (nine out of ten), we identified a number of brain structural variation correlations that are significant in the early stages of AD progression and that can be utilised to identify whether a patient is in the AD early stages. Specifically, they are Vol(C). of L.Precentral - CTA. of L.Supramarginal, Vol(C). of L.Precuneus - CTA. of L.SuperiorParietal, Vol(C). of R.SuperiorFrontal - Vol(C). of L.Precentral and Vol(C). of R.Precentral - CTA. of R.CaudalMiddleFrontal.

For Vol(C). of L.Precentral - CTA. of L.Supramarginal, the left precentral gyrus and left supramarginal gyrus are two vital regions of the human brain, the neural connections between these two regions are essential for speech and motor control, helping to translate the information heard into spoken language during comprehension, and enabling muscle movements in the mouth and throat through motor control [326][327][328]. This correlation can be a factor in the impaired motor function, reduced body-motor sensory, impairments in language and reading, and impaired social interaction in AD patients.

For Vol(C). of L.Precuneus - CTA. of L.SuperiorParietal, both left precuneus and left superior parietal gyrus are involved in the processing of bodily sensory and motor control, and they are associated with higher cognitive functions such as attention and working memory [329][330][331]. In addition, both regions are implicated in the processing of spatial attention and motor planning in response to task stimuli. For instance, in terms of the processing of spatial perception and cognition, the left precuneus processes the perception, memory and orientation of spatial relations, while the left superior parietal gyrus processes motion and position sensation, vision and spatial cognition. In terms of attention and working memory, the left precuneus is associated with visuospatial attention, while the left superior parietal gyrus deals with motor and spatial attention. Both areas are associated with the processing of working memory, especially when information is required to be maintained and manipulated. In terms of cognitive control, the left precuneus is involved in the process of decision making and execution of plans, while the left superior parietal gyrus is involved in the process of planning and adjusting behaviour. This correlation can be a factor in the symptoms of orientation and navigation difficulties, diminished spatial memory and spatial planning abilities, attention deficits, impaired working memory, reduced cognitive flexibility and creativity in AD patients.

For Vol(C). of R.SuperiorFrontal - Vol(C). of L.Precentral, although the right superior frontal gyrus and the left precentral gyrus are situated in different regions of the brain, their functions overlap to certain extents. For instance, the coordination and execution of muscle movements is required when performing certain higher cognitive tasks, which involve the interaction of the two regions, these higher cognitive functions include attention, working memory, decision making and planning [332][333]. This correlation can be a factor in the occurrence of cognitive and motor decline, hand movement disorders, attention and working memory loss, behavioural disturbances and language impairment in AD patients.

For Vol(C). of R.Precentral - CTA. of R.CaudalMiddleFrontal, both the right precentral gyrus and the right middle frontal gyrus perform significant motor and cognitive functions, working in tandem to control the body's movements [332][334]. For instance, the right middle frontal gyrus can "pre-program" the right precentral gyrus to assist in the execution of a predetermined movement sequence before the activity is undertaken. This correlation can be a factor in symptoms of impaired motor function (including gait instability and limb stiffness), abnormal behaviour, diminished decision-making, emotional instability, language impairment and impaired comprehension in AD patients.

6.3 Potential indicators for AD early detection

This section provides a detailed explanation of the significant biomarker correlations identified in Section 6.2 in terms of differences in the distribution of AMDQ quantitative values of early stage (BL-M06) spatio-temporal structural variation correlations (e.g. Figure 6.13) and differences in the distribution of relative structural variation states between biomarkers (e.g. Figure 6.14) for cognitively impaired and non-cognitively impaired individuals at various time points. To facilitate readability, all the significant biomarker correlations information for all time points are included in Appendix E.

For Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal (Figure E.1-E.4), it was observed that non-cognitively impaired individuals would have a higher probability of experiencing structural variation in the opposite direction and a lower probability of experiencing structural variation in both growth between the two biomarkers at the early stage (BL-M06) than cognitively impaired individuals.

For CTA. of R.Precuneus - Vol(C). of R.Supramarginal (Figure E.5-E.8), it was observed that non-cognitively impaired individuals would have a greater probability of having structural variation in the opposite direction and a lower probability of having

structural variation in both grow between the two biomarkers at the early stage (BL-M06) than cognitively impaired individuals.

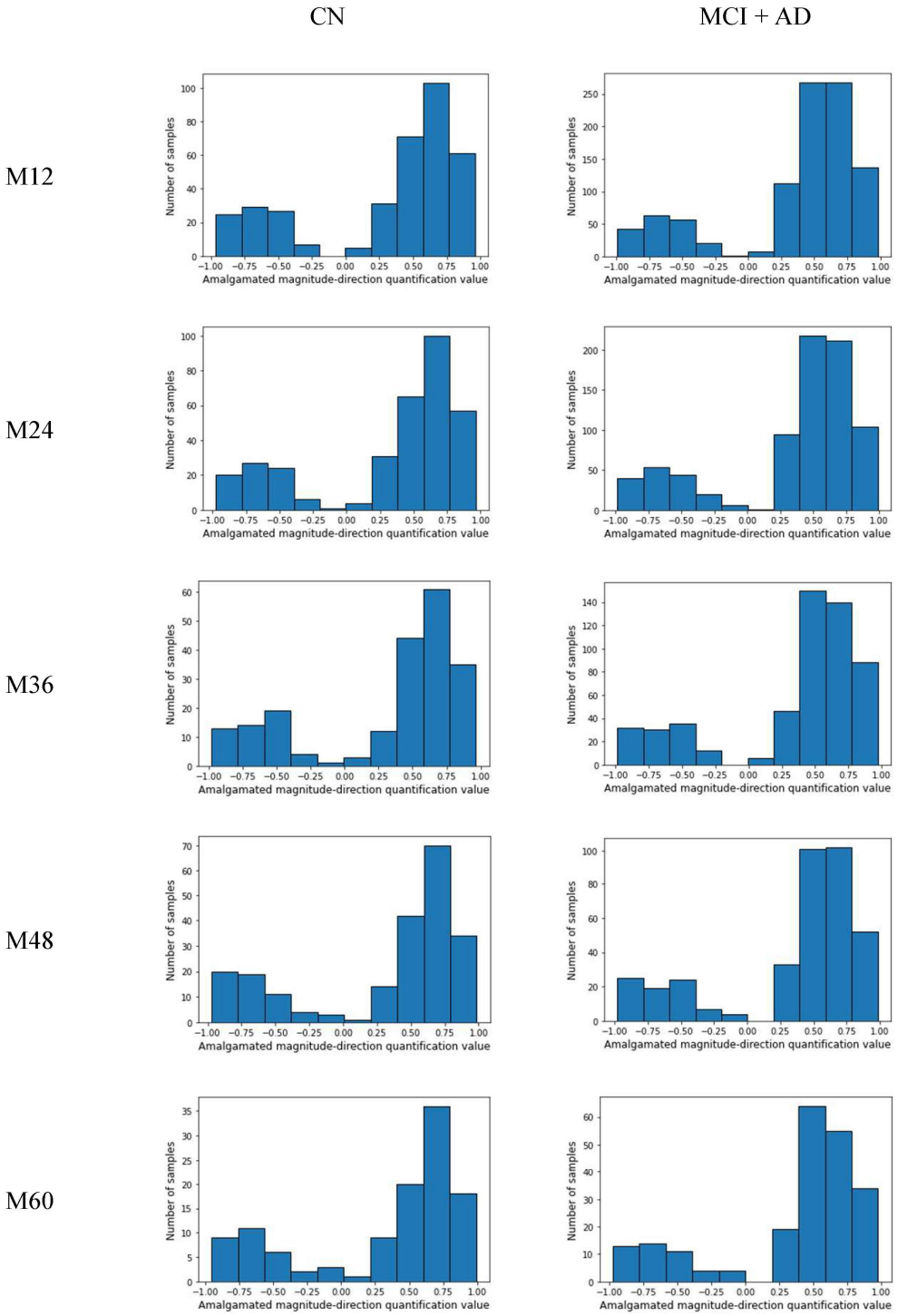


Figure 6.13: Differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

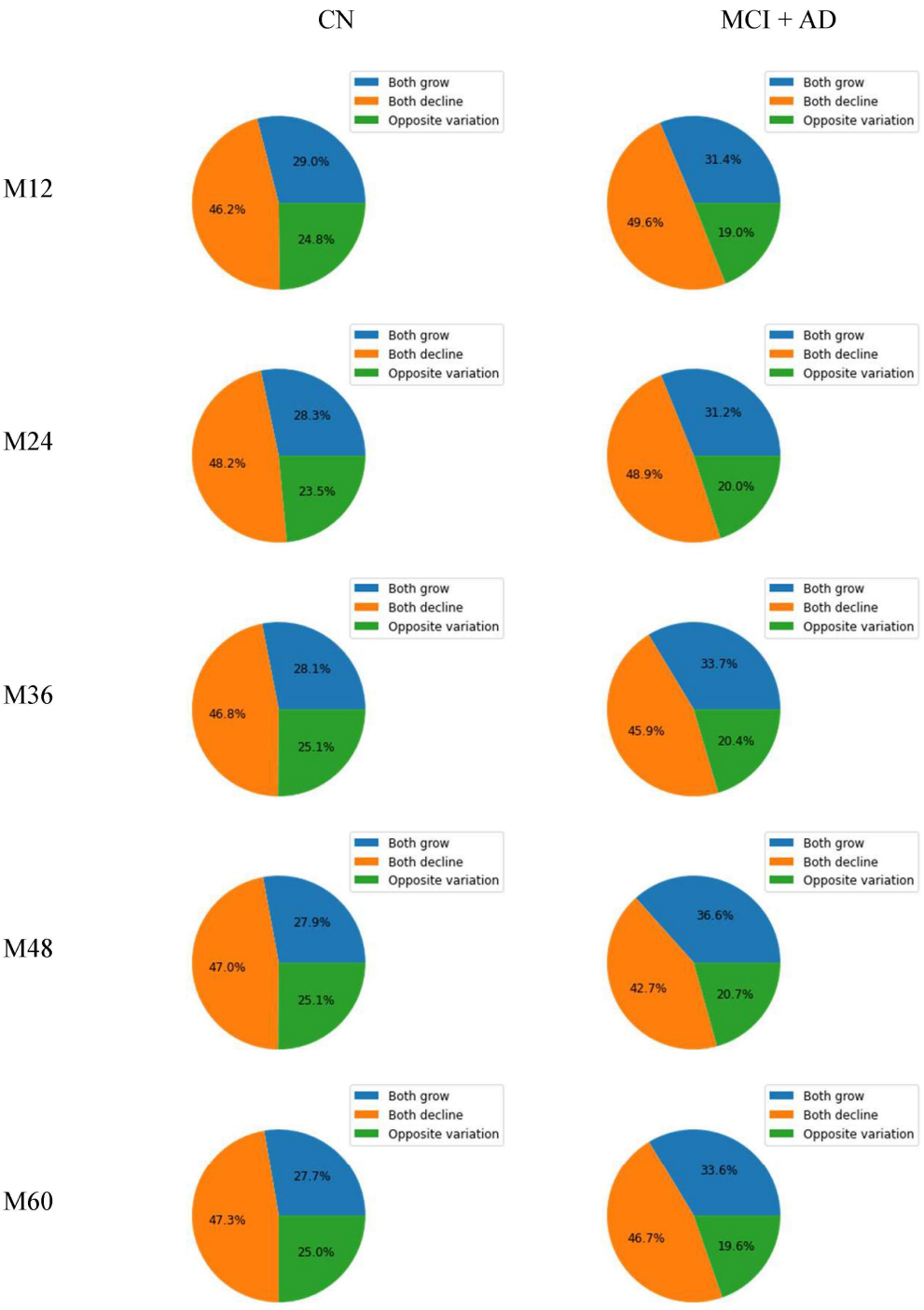


Figure 6.14: Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

For CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal (Figure E.9-E.12), it was observed that non-cognitively impaired individuals had a slightly greater probability of opposite direction structural variation between the two biomarkers than cognitively impaired individuals at the early stage (BL-M06), and that the AMDQ quantification of the opposite direction biomarker structural variation correlation was more concentrated between -0.25 and -0.75 for non-cognitively impaired individuals, whereas the corresponding AMDQ quantification was more evenly distributed in cognitively impaired individuals.

For CTA. of L.InferiorParietal - CTA. of L.Precentral (Figure E.13-E.16), it was observed that non-cognitively impaired individuals had a higher probability of opposite direction structural variation and a lower probability of both growth structural variation between the two biomarkers than cognitively impaired individuals at an early stage (BL-M06) and that the correlations of opposite direction biomarker structural variation in AMDQ were more concentrated between -0.5 and -1 in non-cognitively impaired individuals, whereas the corresponding AMDQ quantitative distribution was more concentrated between -0.4 and -0.75 for cognitively impaired individuals.

For Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal (Figure E.17-E.20), it was observed that non-cognitively impaired individuals experienced a greater probability of opposite directional structural variation between the two biomarkers than cognitively impaired individuals at early stages (BL-M06). At later time points (M72 to M120) samples, cognitively impaired individuals had a greater probability of both decline structural variation and a lower probability of both growth structural variation between the two biomarkers than non-cognitively impaired individuals at early stages (BL-M06).

For CTA. of R.ParsTriangularis - CTA. of L.Postcentral (Figure E.21-E.24), it was observed that cognitively impaired individuals experienced a higher probability of both growth structural variation and a slightly lower probability of opposite direction

structural variation between the two biomarkers than non-cognitively impaired individuals at the early stage (BL-M06).

For Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal (Figure E.25-E.28), it was observed that cognitively impaired individuals exhibited a greater probability of both growth structural variation and a lower probability of opposite direction structural variation between the two biomarkers than non-cognitively impaired individuals at the early stage (BL-M06).

For Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal (Figure E.29-E.32), it was observed that cognitively impaired individuals experienced a slightly greater probability of both decline structural variation and a slightly lower probability of both growth structural variation between the two biomarkers than non-cognitively impaired individuals at the early stage (BL-M06).

For Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal (Figure E.33-E.36), it was observed that non-cognitively impaired individuals experienced a greater probability of opposite direction structural variation, a lower probability of both decline structural variation and a lower probability of both growth structural variation between the two biomarkers than cognitively impaired individuals at the early stage (BL-M06).

6.4 Summary

This chapter presents interpretable analyses of the experimental results of the proposed multi-dimensional tensor multi-task learning regression algorithm from various aspects. Specifically, the latent factors of brain biomarkers derived from multi-dimensional tensor multi-task regression (AMDQ-TMTL) were first investigated to identify brain regions affected by AD progression. The analysis demonstrated that the latent factors were dominated by the volumes of cortical parcellations and average cortical thickness

at all time points for different cognitive score prediction tasks, but the latent factors were different in detail, maybe due to the different test orientations and methodologies of the different cognitive score tests. Significant relative structural variation correlations between brain biomarkers identified from the experiments were then explored and analysed, which could be utilised to predict AD progression and as potential indicators of early AD identification. The usage possibility of potential indicators for early AD identification is then analysed in detail based on the distribution of early stage quantitative AMDQ values and relative structural variation information of brain biomarkers for cognitively impaired and non-cognitively impaired individuals at various specific time points.

Chapter 7

4D tensor multi-task learning algorithm design for time-continuous data

The multi-dimensional tensor multi-task learning regression proposed in the previous sections can address both the monotonic data form and small dataset problems faced by disease progression prediction. In addition to the two critical problems mentioned above, disease progression prediction faces the problem of scarcity of time-continuous data, meaning that the number of available datasets further decreases as the disease progresses. In order to solve all three problems mentioned above at once, we increase the data dimensionality of the proposed multi-dimensional tensor multi-task learning regression from three to four dimensions and propose two diverse algorithms. In this chapter, we firstly present the concept of disease dynamics prediction utilised to address the above three problems, and then design and construct two time-continuous algorithms based on multi-dimensional tensor multi-task learning regression, namely 4D tensor multi-task ensemble learning and 4D tensor multi-task continual learning respectively, and this chapter details the design concepts, construction processes, experimental results and analysis of each algorithm.

7.1 Concept of disease dynamics prediction

Machine learning techniques for predicting AD progression can substantially assist researchers and clinicians establish strong AD preventive and treatment strategies.

However, current research on AD prediction algorithms encounters challenges with monotonic data form, small dataset and scarcity of time-continuous data. In order to solve all three problems at once, we add the concept of disease dynamics prediction to the algorithms.

The most important characteristic of disease dynamic prediction is its dynamic property, i.e., it can constantly update the prediction results based on the patient's historical data, current status and latest observation, and realize real-time monitoring and prediction. Since disease progression is a dynamic process, disease progression can be influenced by a variety of factors, including environmental factors, genetic factors, and lifestyle. And these factors can vary over time, thus affecting the disease progression. Therefore, it is not sufficient to make predictions based on the current state of the disease alone; it is necessary to dynamically assess and predict the future progression of an individual's disease in conjunction with its historical record and current state. This characteristic allows dynamic prediction of disease progression to have significant implications in clinical applications, providing more accurate and personalized treatment recommendations and helping clinicians to better formulate individualized treatment plans.

Traditional disease diagnosis and treatment protocols often consider only the patient's current symptoms and condition, lacking long-term monitoring and prediction. In contrast, dynamic prediction of disease progression can combine a variety of information such as a patient's historical medical records and biomarkers to predict and monitor the future progression of a patient's condition. The dynamic approach to disease prediction not only allows earlier detection of disease variations, but also provides clinicians with more timely and accurate treatment recommendations. By analyzing patients' historical data, more accurate and comprehensive prediction models can be built and provide significant support for the design and result analysis of clinical trials.

In addition to the above benefits, disease dynamic prediction has the following advantages: 1) Improve patients' quality of life: disease dynamic prediction can help patients better comprehend the disease progression and possible complications, thus taking corresponding preventive actions and enhancing patients' quality of life. 2) Promote the development of precision medicine: dynamic disease prediction can provide personalized treatment plans based on individual patient characteristics and historical records, combined with the dynamic variation of the disease, thus promoting the development of precision medicine and improving the survival rate of patients and the success rate of medical treatment. 3) Provide reference for medicine development: disease dynamics prediction can provide relevant data and trends of disease progression to facilitate medicine development and improve treatment effectiveness. 4) Reduce medical costs: disease dynamics prediction can reduce medical costs for medical institutions and individuals through early detection and prevention of diseases, thus enhancing the efficiency and sustainability of medical resources utilisation.

For the problem of the scarcity of time-continuous data. In real-world applications, people with suspected Alzheimer's disease will continue to be tested in hospitals, which is a waste of future incremental data if only a baseline model is utilised or if the patient's ongoing testing records cannot be adequately incorporated.

7.2 4D Tensor Multi-task Ensemble Learning

In order to solve all three problems (monotonic data form, small dataset and scarcity of time-continuous data) at once, we add the concept of disease dynamics prediction to the algorithms. And to address the problem of the scarcity of time-continuous data, we proposed the gradient boosting ensemble learning approach to integrate continuous test recordings of subjects to continuously improve prediction accuracy. Specifically, the proposed algorithm and model are based on the previously proposed multi-dimensional

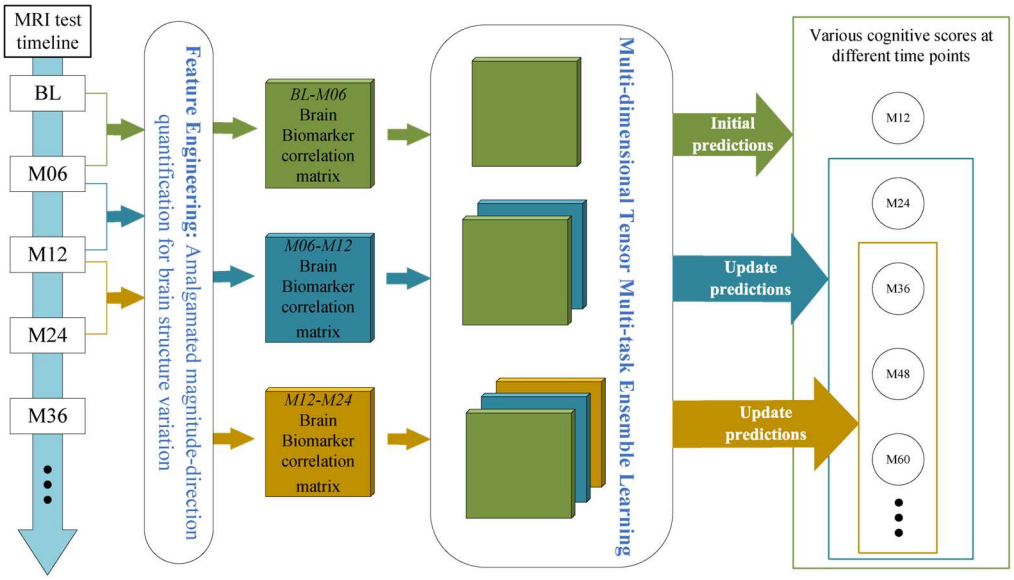


Figure 7.1: The graphical representation of the proposed approach to deal with disease dynamics prediction (continuous updating of disease prediction results) for new patient.

tensor multi-task learning regression to predict disease progression and utilise the gradient boosting ensemble learning approach to combine participants' current predictions with all previous predictions to further improve prediction accuracy and update prediction results when new MRI data are available to participants (Figure 7.1).

7.2.1 Gradient boosting

By combining a collection of weak learners to create stronger learners, ensemble learning has been proved to be efficient in a diversity of prediction tasks. Boosting is the dominating strategy in ensemble learning approaches, which generates a group of weak learners by training predictors sequentially rather than individually, with the goal of utilising the prior learner's errors to construct a more effective model for the next learner.

Gradient Boosting (GB) is a boosting method extension that utilise gradient descent optimization methods to recognize global or local minima of cost function. It uses a

sequence of weak learners to train the machine to fit the model on the input feature space, each of which enhances the prediction accuracy of learner before it. GB develops powerful learners by merging numerous weak learners in multiple iterations [335][336], The proposed approach improves prediction accuracy by gradually fitting a more precise model to the residuals of prior phase in the final stage of GB building framework. This procedure will remain on until a sufficiently precise model is produced.

7.2.2 Algorithm design and construction

Two successive MRI examinations were utilised to calculate the correlation of structural variation between different brain biomarkers. Preliminary versions of the quantitative technique (Amalgamated magnitude-direction quantification for brain structure variation) have been presented in Section 4.2, and this algorithm extends and implements it across number of subsequent time periods (BL to M06, M06 to M12, M12 to M24).

The algorithm design is the same as the algorithm in Section 5.3.1, but applies its algorithm to each two consecutive time points, and then integrate and optimize all the results utilising the gradient boosting ensemble learning approach.

7.3 4D Tensor Multi-task Continual Learning

In order to solve all three problems (monotonic data form, small dataset and scarcity of time-continuous data) at once, we propose a novel machine learning approach that implements the 4D tensor multi-task continual learning algorithm to predict AD progression by quantifying multi-dimensional information on brain structural variation and knowledge sharing between patients. To meet real-world application scenarios, the method can integrate knowledge from all available data as patient data increases to

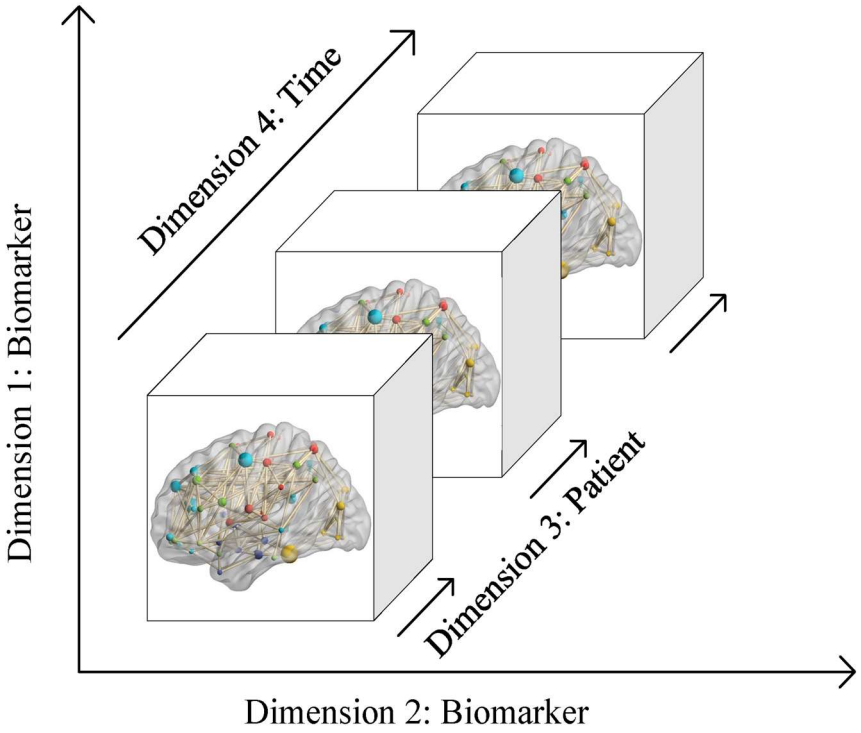


Figure 7.2: The 4D tensor data structure constructed and utilised in the research.

continuously update and optimise prediction results. Figure 7.2 illustrates the 4D tensor data structure constructed and exploited in this research.

7.3.1 Algorithm design and construction

In the real world, patients suspected of AD will continue to go to hospital for testing. Subsequent incremental data is wasted if only a baseline model is utilised or if consecutive test records of patients cannot be reasonably integrated. To solve the problem, we apply concept of continual learning to our approach, which can update the prediction results by allowing the model to receive new MRI data while receiving all the latent factors from all previous prediction models. Figure 7.3 depicts the architecture, learning process and real-world applications of the proposed approach.

The correlation of structural variance between different brain biomarkers was calculated utilising two consecutive MRI scans. Quantitative approach has been presented in preliminary versions (Amalgamated magnitude-direction quantification

for brain structure variation) in Section 4.2, and this work expands and executes it throughout a number of following time periods (BL to M06, M06 to M12, M12 to M24).

To predict various cognitive scores (e.g., ADAS-Cog and MMSE) for AD at future time points. Consider the multi-dimensional tensor multi-task continual regression problem for *t* time points, *n* training samples with d_1 and d_2 features. Let $X \in \mathbb{R}^{d_1 \times d_2 \times n}$ be the input three-dimensional tensor from two successive MRI records and it is the combination of correlation 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 (clinical scores) and $\mathrm{y}_t = [y_1, \cdots, y_n] \in \mathbb{R}^n$ is the corresponding target at various time points.

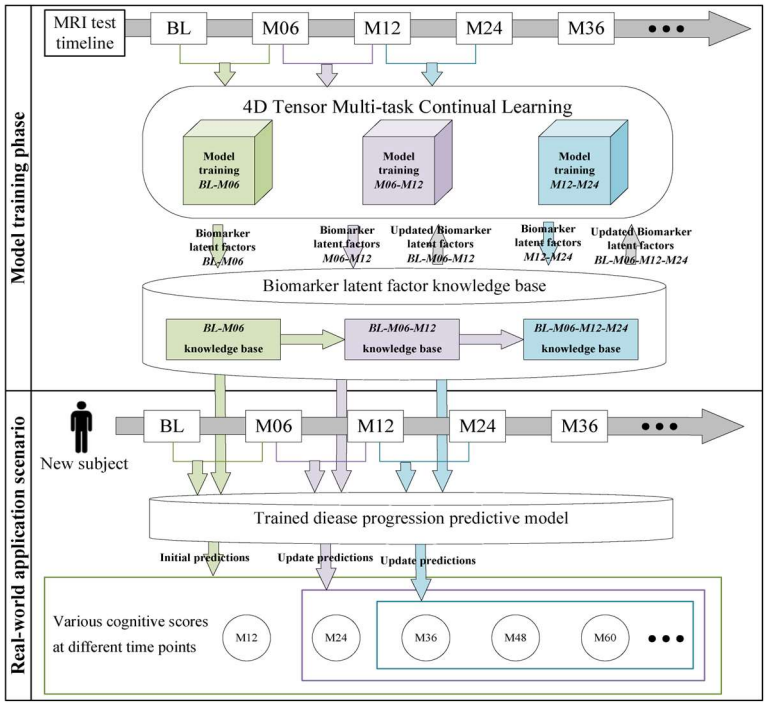


Figure 7.3: Architecture, learning procedure and real-world application for the proposed 4D tensor multi-task continual learning (4DTMTCL) approach. Specifically, from a single prediction perspective, we model the AD progression using a tensor multi-task learning algorithm based on the calculation of correlations of structural variation in brain biomarkers. Correlations are integrated into the algorithm to address the monotonic data form problem, while prediction for each patient sample is set as one task, with all tasks sharing a set of biomarker latent factors obtained through tensor decomposition to address the generalisation problem caused by small data sets. From a continuous prediction perspective, the new prediction model acquires all the latent factors from previous models and updates the predictions whenever the patient's MRI data is updated.

For *t*-th prediction time point, the objective function of proposed approach can be stated as follows:

$$L_t(X, \mathrm{y}_t) = \min_{\mathrm{W}_t,\, \mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t} \frac{1}{2}\|\hat{\mathrm{y}}_t - \mathrm{y}_t\|_2^2 + \frac{\lambda}{2}\|X - [\![\mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t]\!]_{\mathrm{S}}\|_{\mathrm{F}}^2 + \beta\|\mathrm{W}_t,\, \mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t\|_1$$

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

$$\text{where } \mathrm{U} = [\eta \odot \mathrm{V} + (1-\eta) \odot (\mathrm{A}_t \mathrm{B}_t^{\mathrm{T}})] \odot \mathrm{K} \odot \mathrm{W}_t \odot \mathrm{X}_n \ , \ \mathrm{U} \in \mathbb{R}^{d_1 \times d_2} \tag{7.1}$$

where the first term calculates the empirical error with training data, $\hat{\mathrm{y}}_t = [\hat{y}_1, \cdots, \hat{y}_n] \in \mathbb{R}^n$ are predicted values, $\mathrm{A}_t \in \mathbb{R}^{d_1 \times r}$ is the latent factor matrix for first biomarker dimension and $\mathrm{B}_t \in \mathbb{R}^{d_2 \times r}$ is the latent factor matrix for 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 regularisation parameters. $\mathrm{V} \in \mathbb{R}^{d_1 \times d_2}$ is knowledge base matrix which stores principal biomarker latent factors from all preceding model predictions. V is updated after each model prediction with following equation: $\mathrm{V}_{\mathrm{new}} = \eta \odot \mathrm{V}_{\mathrm{old}} + (1-\eta) \odot (\mathrm{A}_t \mathrm{B}_t^{\mathrm{T}})$. The hyperparameter η is utilised to control the proportion of preceding and present knowledge base that is employed. Acquiring latent factors by optimising the symmetric CP tensor decomposition objective function$\|X - [\![\mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t]\!]_{\mathrm{S}}\|_{\mathrm{F}}^2$, where $X = [\![\mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t]\!]_{\mathrm{S}} = \sum_{i=1}^{r} \frac{1}{2}(\mathrm{a}_i^t \circ \mathrm{b}_i^t \circ \mathrm{c}_i^t + \mathrm{b}_i^t \circ \mathrm{a}_i^t \circ \mathrm{c}_i^t)$ and $\circ$ denote outer product operation between two vectors, while $\mathrm{a}_i^t, \mathrm{b}_i^t$ and c_i^t correspond to vectors related with *i*-th latent factor for *t*-th prediction time point. $\|\mathrm{W}_t,\, \mathrm{A}_t,\, \mathrm{B}_t,\, \mathrm{C}_t\|_1$ applying the $\ell 1$-norm on W_t, A_t, B_t and C_t matrices individually. 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*. And $\mathrm{Z} = m \odot \mathrm{N}$ denotes $z_{ij} = m n_{ij}$, for all *i, j*. The matrix $\mathrm{K} \in \mathbb{R}^{d_1 \times d_2}$is the duplicate data correction matrix which was implemented to fix the duplicate data problem because the correlation tensor for brain structural variation created by the proposed quantification approach is a symmetric tensor, which

means that correlations between biomarkers are calculated in pairs, resulting in half of the data being duplicates.

The rest of the algorithm design and construction is the same as the algorithm in Section 5.3.1.

7.4 Experimental results

The experiments utilise the same experimental settings and sources (ADNI dataset) as in Section 3.4.1, with an additional pre-processing step added after the pre-processing procedure in Section 5.6, which is updated as follows:

- Removal of image records with failed quality control;
- Removal of features if more than half of the sample's values are missing;
- Excluding subjects who lacked BL and M06 MRI;
- The average of features was applied to fill in the missing data;
- Exclude subjects who had no further MRI detections for the research of AD dynamic prediction.

Tables 7.1 and 7.2, Figure 7.4 and 7.5 illustrate the experimental results of MMSE predictions. Tables 7.3 and 7.4, Figure 7.6 and 7.7 illustrate the experimental results of ADAS-Cog predictions. Tables 7.5 and 7.6, Figure 7.8 and 7.9 illustrate the experimental results of RAVLT TOTAL predictions. Tables 7.7 and 7.8, Figure 7.10 and 7.11 illustrate the experimental results of FLU ANIM predictions.

In terms of overall regression performance, our proposed approaches exceed single-task learning, benchmarks and state-of-the-art MTL approaches in terms of nMSE and wR for MMSE, ADAS-Cog, RAVLT TOTAL and FLU ANIM cognitive scores.

Furthermore, the proposed approaches achieve lower rMSE than comparison approaches for all individual time points. Our primary observations are as follows:

1) The proposed 4D Tensor Multi-task Ensemble Learning (4DTMTEL) and 4D Tensor Multi-task Continual Learning (4DTMTCL) approaches outperform single-task learning models, benchmarks and state-of-the-art MTL models, which validates the application of similarity calculations that incorporate both magnitude and direction information of brain structural variation and the utilisation of the tensor latent factor hypothesis in our MTL formulation.

2) Prediction stability is greatly increased by the suggested 4DTMTEL and 4DTMTCL approaches. The standard deviation of results for 20 iterations of the experiments were lower than that of the comparative approaches. This can be as a result of the fact that the proposed quantification approach combines knowledge on global brain variability and integrates latent factors from brain biomarkers to the prediction algorithm to increase stability.

3) The chronologically continuous MRI recordings of the participants can be efficiently aggregated by the proposed 4DTMTEL approach to enhance prediction accuracy, and as time-continuous MRI records were increased, the prediction accuracy at subsequent time points increased. In contrast, the addition of time-continuous MRI recordings had no significant effect on the benchmarks and state-of-the-art competing approaches.

4) In the scenario of AD dynamic prediction, the proposed 4DTMTCL approach can achieve outstanding results. Knowledge from the previous models and the present model are combined in the present prediction model. Prediction accuracy can be increased by time-continuous MRI recordings of the participants, and when additional time-continuous MRI recordings are provided, prediction accuracy increases over time.

Contrarily, the inclusion of time-continuous MRI recordings has no beneficial effect on the benchmarks and state-of-the-art competing approaches.

5) We have observed that although continuous-time MRI data improves the prediction performance, it does not lead to a significant performance gain. This may be due to AD patients do not experience major brain structural variations in a short term, i.e., the brain structural variations presented by MRI data of AD patients in the early stage (BL to M06) are not significantly different from that presented by the MRI data in the later short stages (M06 to M12, M12 to M24).

Table 7.1: Comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for MMSE at time points M12 to M48. The best results are bolded.

Target: MMSE	**Input MRI data**	M12 rMSE	M24 rMSE	M36 rMSE	M48 rMSE
Ridge	BL, M06	4.0605±1.2386	5.0048±1.2526	4.0442±1.4497	5.5393±1.4989
	BL, M06, M12	-	7.3996±1.5241	7.9911±1.6971	7.9596±1.8440
	BL, M06, M12, M24	-	-	9.5140±0.5463	9.7311±1.2082
Lasso	BL, M06	1.9052±1.1061	2.4768±1.5718	2.5940±0.7297	2.2096±0.6772
	BL, M06, M12	-	2.9061±0.2043	3.0122±0.6175	3.2217±0.5007
	BL, M06, M12, M24	-	-	3.1207±0.8536	2.3820±0.4220
TGL	BL, M06	1.7395±0.1282	1.7217±0.1640	1.8433±0.1870	1.7016±0.4988
	BL, M06, M12	-	2.2411±0.1879	2.1053±0.6062	1.9857±0.8790
	BL, M06, M12, M24	-	-	2.2048±0.5030	2.2445±0.1649
nCFGL1	BL, M06	1.3466±0.1764	**1.3788±0.2848**	1.9872±0.1796	1.6300±0.3682
	BL, M06, M12	-	1.7405±0.1550	2.2477±0.4332	2.3808±0.1430
	BL, M06, M12, M24	-	-	2.7400±0.3393	3.6115±0.2793
cFSGL	BL, M06	1.5864±0.6592	1.6055±0.7507	1.6029±0.4876	1.8139±0.9093
	BL, M06, M12	-	2.3473±0.2539	1.2499±0.5370	2.5298±1.1984
	BL, M06, M12, M24	-	-	1.5847±0.3165	1.8561±0.3992
FL-SGL	BL, M06	1.7686±0.1391	2.2371±0.4471	2.0123±0.2851	1.7532±0.5990
	BL, M06, M12	-	2.3175±0.4090	1.8827±0.1246	2.0954±0.8623
	BL, M06, M12, M24	-	-	1.8288±0.1448	1.8031±0.1364
NC-CMTL	BL, M06	1.4302±0.4354	1.5425±0.3381	1.7830±0.5100	2.0724±0.5488
	BL, M06, M12	-	1.6922±0.1010	1.6497±0.1516	2.4353±0.6481
	BL, M06, M12, M24	-	-	1.7187±0.2140	1.9310±0.7061
FTS-MTFL	BL, M06	1.6373±0.3992	1.7305±0.4099	1.5596±0.6720	1.6122±0.5982
	BL, M06, M12	-	1.5405±0.3122	1.9834±0.8174	1.8015±0.6018
	BL, M06, M12, M24	-	-	2.1884±0.5894	2.0873±0.1993
GAMTL	BL, M06	1.8805±0.1909	1.9699±0.2790	2.1551±0.1280	1.8221±0.6720
	BL, M06, M12	-	2.4891±0.6281	1.8632±0.7538	1.8624±0.7350
	BL, M06, M12, M24	-	-	1.7557±0.3493	1.8827±0.4350
dMTLc	BL, M06	1.8372±0.1873	1.9649±0.6950	1.8731±0.5722	1.7281±0.5057
	BL, M06, M12	-	1.7089±0.6240	2.2800±0.3541	2.0459±0.1398
	BL, M06, M12, M24	-	-	2.3536±0.8492	1.6002±0.6151
4DTMTEL	BL, M06	**1.3194±0.1091**	1.4468±0.1247	1.5221±0.1048	1.5720±0.1066
	BL, M06, M12	-	1.4352±0.1343	1.5952±0.2240	1.5457±0.1214
	BL, M06, M12, M24	-	-	1.5103±0.2104	1.5275±0.1396
4DTMTCL	BL, M06	**1.3194±0.1091**	1.4468±0.1247	1.5221±0.1048	1.5720±0.1066
	BL, M06, M12	-	1.4085±0.1382	1.5141±0.2379	1.5277±0.1147
	BL, M06, M12, M24	-	-	**1.4661±0.2166**	**1.5131±0.1281**

Table 7.2: Continuation of Table 7.1, comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for MMSE at time points M60 to M120. The best results are bolded.

***Target:* MMSE**	M60 rMSE	M72 rMSE	M84 rMSE	M96 rMSE	M108 rMSE	M120 rMSE
Ridge	6.3593±1.4969	7.9274±0.9130	7.5438±1.0703	7.1825±1.1658	7.3953±0.9400	8.6403±1.3850
	8.2289±0.6598	8.7497±1.5140	7.9796±1.0312	7.6519±1.5062	7.1341±0.5937	9.6025±0.5670
	9.6005±1.2726	9.1240±0.7957	8.9965±1.7003	9.4487±0.8986	7.9082±0.7490	10.3640±0.8991
Lasso	2.2814±0.9163	3.5367±1.8656	4.4608±0.8752	5.2756±2.1243	5.1665±1.0716	5.3829±1.9340
	3.2202±0.7442	3.0084±1.0496	4.7956±1.0323	4.2441±0.8997	5.8198±1.9195	6.8139±0.8465
	3.0229±0.7160	4.1972±0.6126	4.8492±1.1596	5.6150±1.8645	5.0347±1.1981	6.2462±0.5733
TGL	1.8078±0.7116	1.9643±0.9502	2.1360±0.8765	2.6708±0.7499	3.0527±0.6373	3.1320±1.1055
	1.6548±0.7567	2.3353±0.5955	2.2137±1.4488	1.9908±0.9140	2.7911±0.2629	3.5358±1.0471
	2.5419±0.1712	1.7823±0.9306	1.8185±0.3213	2.7418±0.1237	3.2831±0.7004	3.0445±0.2944
nCFGL1	1.6967±0.6013	2.1037±0.6930	2.5159±0.3190	2.8206±0.5456	2.8242±0.5211	2.9035±0.6056
	2.4251±0.4989	2.3726±0.4320	2.3622±1.0407	2.2978±1.1791	3.2648±0.7847	4.3086±1.3547
	2.4003±0.2102	2.5715±0.1857	2.9225±0.2710	3.0332±0.2173	2.8457±0.5165	2.6099±0.6046
cFSGL	1.8209±0.8150	2.2522±0.7898	2.3981±0.3207	2.6192±1.3076	2.2614±0.5827	3.5795±0.3449
	2.1973±0.1435	1.9645±0.3910	2.6420±0.3640	2.1932±1.0139	2.5769±0.7082	3.3759±0.4894
	2.1737±0.1826	1.8531±0.7090	2.6006±0.7572	3.2842±0.6721	2.7810±0.5580	3.5367±0.3265
FL-SGL	1.6258±0.2276	1.9267±0.6614	2.0164±0.6971	2.5265±0.2901	3.0563±0.4682	3.7890±0.3501
	1.6153±0.1950	1.9959±0.5304	2.0198±0.3995	2.7929±0.9338	3.9560±0.4159	4.0319±0.4056
	2.0860±0.5319	1.8906±0.3893	2.2538±0.2050	3.5946±1.0574	4.1514±0.9780	3.2839±0.2996
NC-CMTL	1.9627±0.8186	1.9513±0.8896	2.1118±0.3061	1.9114±0.4402	2.1879±0.7359	3.7274±0.8463
	2.4318±0.5330	1.9267±0.1926	2.1159±0.6690	2.4782±0.2540	2.5883±0.7630	3.8521±1.0784
	2.2790±0.7284	1.6552±0.5087	2.5465±0.7302	2.2351±0.3470	2.2149±0.3965	4.2324±0.1721
FTS-MTFL	2.3715±0.7826	1.8304±0.2191	1.8528±0.7210	2.4228±0.8804	1.8972±0.2475	3.7141±1.0977
	2.3476±0.1237	1.5421±0.4664	2.3142±0.5606	2.3964±0.1815	2.4296±0.3599	3.3442±0.3510
	2.4585±0.2939	2.1518±0.1055	2.6991±0.9247	2.0088±0.1687	2.9686±0.1747	3.8057±0.8186
GAMTL	1.9757±0.5196	2.1530±0.6055	2.2004±0.7192	1.9492±0.7142	2.3220±1.1041	3.4550±0.6403
	1.7496±0.8844	2.0029±0.4856	1.6487±0.2324	2.2326±0.3414	2.7316±0.6654	3.3714±0.6662
	1.9926±0.4640	2.5822±0.3401	2.0073±0.2722	2.1107±0.2498	2.7462±0.2259	3.4308±0.8263
dMTLc	1.5245±0.8603	1.7714±0.4606	1.8294±0.3123	2.1448±0.8492	2.5025±0.2857	3.4188±0.6761
	1.8893±0.6059	1.9354±0.5053	1.7133±0.3051	2.6840±0.1889	2.9853±0.7579	3.1045±0.9678
	2.0094±0.3696	1.9294±0.2160	2.5837±0.2923	2.1787±0.2055	2.7875±0.8355	3.5373±0.2367
4DTMTEL	1.4833±0.1649	1.5661±0.1684	1.6105±0.2253	1.6302±0.2874	1.6375±0.2206	2.1358±0.3521
	1.4137±0.1929	1.5745±0.1501	1.7903±0.2986	1.6384±0.2836	1.6783±0.3767	2.0227±0.2636
	1.4409±0.1776	1.4566±0.1983	1.6768±0.3934	**1.6121±0.1870**	**1.5578±0.2118**	1.9943±0.2503
4DTMTCL	1.4833±0.1649	1.5661±0.1684	1.6105±0.2253	1.6302±0.2874	1.6375±0.2206	2.1358±0.3521
	1.4166±0.1647	1.4973±0.1948	1.6258±0.2817	1.6169±0.2485	1.7467±0.3980	2.0793±0.4031
	1.4583±0.1580	**1.4047±0.1423**	**1.5457±0.3498**	1.6158±0.1769	1.5630±0.2641	**1.9048±0.2986**

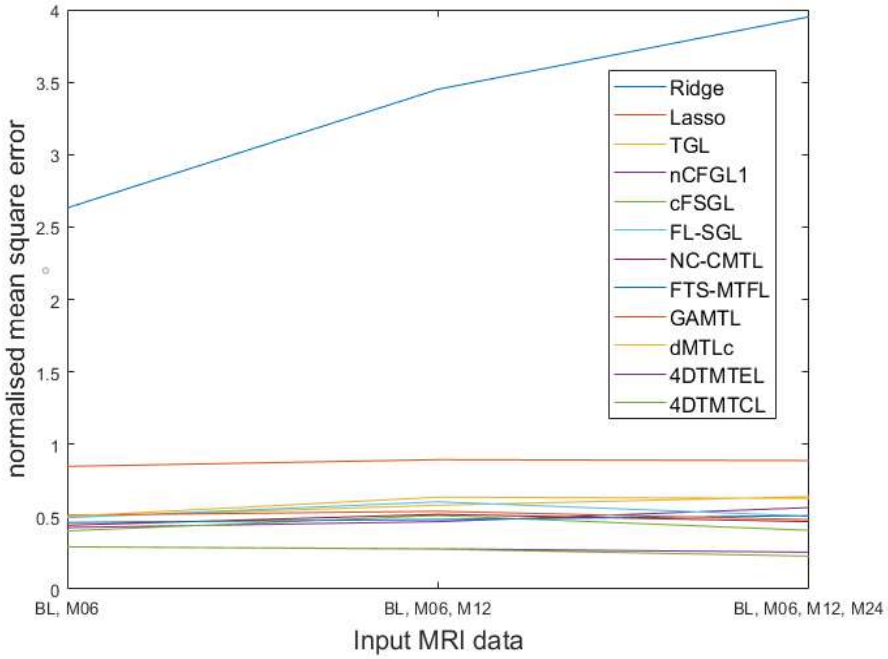


Figure 7.4: nMSE comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for MMSE prediction.

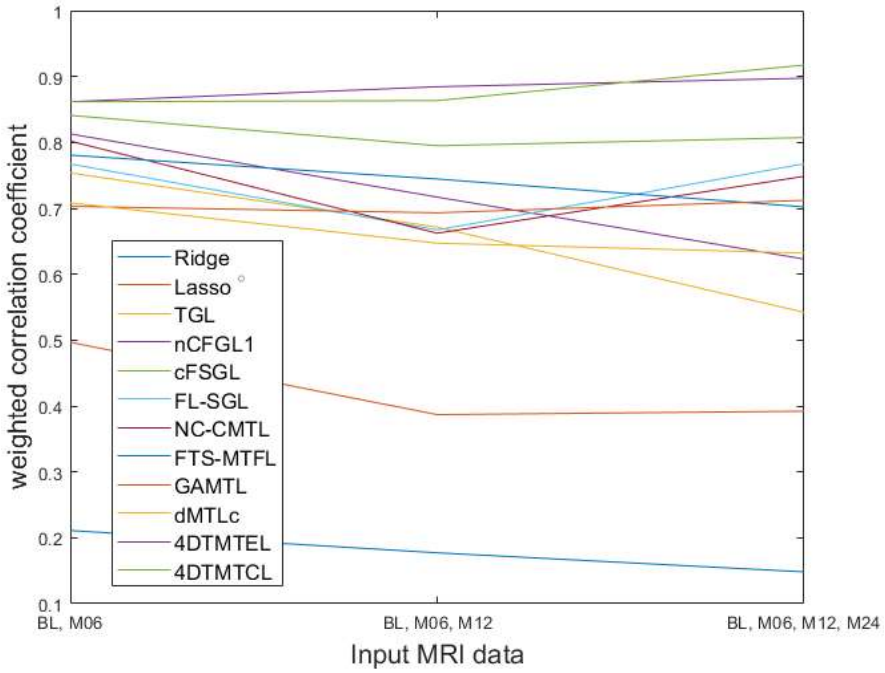


Figure 7.5: wR comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for MMSE prediction.

Table 7.3: Comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for ADAS-Cog at time points M12 to M48. The best results are bolded.

Target: ***ADAS-Cog***	**Input MRI data**	M12 rMSE	M24 rMSE	M36 rMSE	M48 rMSE
Ridge	BL, M06	7.8783±1.2673	8.6661±1.4368	7.4033±1.5643	8.7409±1.8191
	BL, M06, M12	-	8.8243±1.7906	8.8064±1.1930	9.9611±1.4785
	BL, M06, M12, M24	-	-	9.8293±1.3956	10.2182±2.0502
Lasso	BL, M06	5.7337±1.3240	6.7600±1.7528	6.5083±1.4925	7.4186±1.2113
	BL, M06, M12	-	7.4115±1.7434	6.9891±0.8733	7.4439±1.8853
	BL, M06, M12, M24	-	-	6.6223±0.9704	8.2291±1.4105
TGL	BL, M06	2.6734±0.8648	2.3945±1.0648	3.6386±1.1138	3.4775±0.5774
	BL, M06, M12	-	2.2042±0.4317	3.3266±0.6929	3.2301±0.9390
	BL, M06, M12, M24	-	-	3.4046±0.1384	3.3095±0.5528
nCFGL1	BL, M06	2.7870±0.5103	2.8953±0.7292	2.5707±0.6152	3.2667±0.3667
	BL, M06, M12	-	2.9747±0.9894	2.4972±0.2057	3.0209±0.3035
	BL, M06, M12, M24	-	-	2.3693±0.2131	3.3951±0.2052
cFSGL	BL, M06	2.1080±0.2942	3.0577±0.2916	2.8119±0.2969	3.1791±0.3089
	BL, M06, M12	-	3.1176±0.8786	2.9217±0.2783	3.3979±0.5687
	BL, M06, M12, M24	-	-	2.6926±0.8575	2.8180±0.5518
FL-SGL	BL, M06	2.6934±0.6158	2.4190±0.3710	3.0659±1.1312	3.4480±0.4345
	BL, M06, M12	-	2.3972±0.4220	3.2004±0.3470	3.6422±0.6449
	BL, M06, M12, M24	-	-	3.6064±0.3216	3.3209±0.9170
NC-CMTL	BL, M06	2.2452±0.6398	2.7504±0.3242	3.1455±0.2992	3.1756±0.3682
	BL, M06, M12	-	3.0751±0.7944	3.6671±0.5416	2.9686±0.7840
	BL, M06, M12, M24	-	-	2.8344±0.2553	3.4965±0.9807
FTS-MTFL	BL, M06	2.7443±0.8547	2.4485±0.4728	3.0464±0.3155	2.2464±0.4095
	BL, M06, M12	-	2.5766±0.5356	2.9794±0.6024	2.5600±0.9375
	BL, M06, M12, M24	-	-	3.3496±0.7875	2.4568±0.5509
GAMTL	BL, M06	2.9950±0.5020	3.0906±0.4274	2.2789±0.4378	2.3314±0.3827
	BL, M06, M12	-	2.5737±0.1773	2.9424±0.2239	2.2051±0.4937
	BL, M06, M12, M24	-	-	2.7214±0.2758	2.1606±0.3091
dMTLc	BL, M06	2.3471±0.3243	2.4841±0.4506	2.9834±0.6368	2.4808±0.6388
	BL, M06, M12	-	2.9829±0.1359	3.2502±0.7970	2.8080±0.7329
	BL, M06, M12, M24	-	-	3.3811±0.6955	3.2836±0.5295
4DTMTEL	BL, M06	**1.5298±0.2621**	1.4607±0.2294	1.5192±0.1306	1.4679±0.2186
	BL, M06, M12	-	**1.4540±0.2314**	1.4427±0.2473	1.4782±0.2067
	BL, M06, M12, M24	-	-	1.5508±0.1954	1.3830±0.2061
4DTMTCL	BL, M06	**1.5298±0.2621**	1.4607±0.2294	1.5192±0.1306	1.4679±0.2186
	BL, M06, M12	-	1.5040±0.2401	1.5249±0.2421	1.4574±0.1396
	BL, M06, M12, M24	-	-	**1.4615±0.1691**	**1.3449±0.1305**

Table 7.4: Continuation of Table 7.3, comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for ADAS-Cog at time points M60 to M120. The best results are bolded.

Target: ***ADAS-Cog***	M60 rMSE	M72 rMSE	M84 rMSE	M96 rMSE	M108 rMSE	M120 rMSE
Ridge	8.6757±1.6906	9.0341±1.9737	9.9267±1.0749	9.1536±2.8575	9.1306±2.7906	10.0144±2.7013
	9.7869±1.5471	9.1319±1.7265	9.6065±1.5761	9.4961±1.8502	9.3202±1.7947	10.4074±2.0445
	9.8216±1.9641	9.2911±1.6340	10.7204±1.7989	10.2439±1.3134	9.9873±1.5460	11.1823±1.6486
Lasso	6.7316±1.9468	7.0183±1.8743	7.6153±1.1450	7.6934±1.4234	8.4926±1.3660	8.7862±1.8170
	7.4301±1.4566	7.0473±1.7117	7.0180±0.9244	8.4846±1.2954	8.6891±0.9937	8.8344±1.7533
	7.2961±1.5863	7.5864±1.6891	7.6392±1.0676	8.6128±1.6273	8.6326±1.5070	9.9269±1.4078
TGL	3.3258±0.9778	3.1367±0.9183	4.9036±0.9516	4.6745±0.7821	6.5995±0.9961	8.6548±0.4650
	3.0159±0.5363	3.0309±0.1220	4.7138±0.4656	5.1646±0.4734	6.5410±0.6574	9.5285±0.2986
	3.5144±0.5259	3.5285±0.6067	4.6988±0.4720	4.8467±0.6282	6.8362±0.3460	9.0363±0.2487
nCFGL1	2.4117±1.0509	3.0193±0.4929	3.8003±1.1919	3.8142±0.7059	5.1221±0.4251	6.0504±0.4185
	2.8907±0.2034	3.0822±0.3067	3.7512±0.6412	3.4966±0.5142	5.1815±0.3721	6.9153±0.6533
	2.8912±0.5105	3.1992±0.1577	3.5886±0.2840	3.4199±0.3548	5.2403±0.2265	6.9320±0.8559
cFSGL	2.8553±0.7009	3.1638±0.7193	3.1942±0.9396	3.5658±0.5458	5.0652±0.5768	5.6949±0.8991
	2.7303±0.3574	3.1941±0.3956	3.5208±0.3989	3.9385±0.9832	5.4814±0.3208	6.3122±0.6872
	2.9001±0.2520	3.1635±0.2414	3.5194±0.4982	4.0421±0.3027	5.5745±0.4515	6.2812±0.8709
FL-SGL	3.3393±0.2468	3.3664±0.5003	3.5970±0.2306	4.0665±0.9310	5.6875±0.6848	6.2026±0.6328
	3.7001±0.2235	3.3402±0.2480	3.9060±0.7186	4.1261±0.9629	5.9868±0.3685	6.9738±0.3244
	3.5321±0.3502	3.6667±0.1622	4.1440±0.2350	4.4765±0.4200	6.6259±0.4258	7.0545±0.4010
NC-CMTL	3.6033±0.9507	3.0725±0.4072	4.3786±0.3920	3.5976±0.6149	3.2001±0.6762	7.6423±1.3861
	3.3945±0.6881	3.1020±0.2123	5.2297±0.7078	5.3846±0.2142	5.3040±0.8341	7.2417±1.3707
	3.7005±0.5167	3.2329±0.5174	5.3419±0.4138	5.2602±0.5021	4.7866±0.7095	7.6715±1.0704
FTS-MTFL	3.1216±0.3429	3.5349±0.2918	4.1791±0.4363	3.2672±0.4010	3.5568±0.4028	7.8608±1.2826
	3.1899±0.4851	3.3461±0.1768	4.2652±0.6417	4.2030±0.9975	4.7321±0.3665	8.4307±1.3024
	3.4971±0.4328	3.4414±0.4792	5.2403±0.2027	4.2788±0.4251	4.8755±0.2291	8.1201±1.1792
GAMTL	3.6593±0.7849	2.3207±0.5119	3.8476±0.7138	2.6636±0.1940	3.8485±0.2107	5.8844±1.1164
	4.0169±0.7510	3.1659±0.4720	4.0634±0.2164	3.4429±0.4732	3.7646±0.8413	6.4669±1.7327
	3.4841±0.8162	3.1396±0.5686	4.3963±0.3237	3.4970±0.2537	3.7177±0.6651	6.0876±1.1467
dMTLc	3.4192±0.2921	2.5766±0.3138	3.1199±0.3819	3.1522±0.5621	4.4009±0.7587	5.9746±1.5326
	3.0881±0.6078	2.4452±0.2432	3.2371±0.8014	3.6831±0.9082	5.3355±0.5443	6.3685±0.6062
	3.8958±0.4380	2.5043±0.4995	3.4825±0.2745	3.6807±0.3103	5.3240±0.8509	6.6902±0.3988
4DTMTEL	1.6805±0.2404	1.4750±0.1545	1.5658±0.1810	1.6959±0.1087	1.6982±0.1979	2.1631±0.3595
	1.6587±0.3974	1.5076±0.1225	**1.4859±0.2067**	1.7446±0.1991	1.7327±0.2129	2.0394±0.3792
	1.6182±0.3573	1.4807±0.1816	1.5098±0.2249	1.7115±0.3015	1.7138±0.1669	2.0587±0.4068
4DTMTCL	1.6805±0.2404	1.4750±0.1545	1.5658±0.1810	1.6959±0.1087	1.6982±0.1979	2.1631±0.3595
	1.5991±0.2052	1.4961±0.3782	1.5486±0.1613	**1.6220±0.1044**	1.6749±0.1746	2.0374±0.3798
	1.4035±0.1876	**1.4717±0.2796**	1.5315±0.1464	1.6376±0.1985	**1.6089±0.1538**	**1.9621±0.3717**

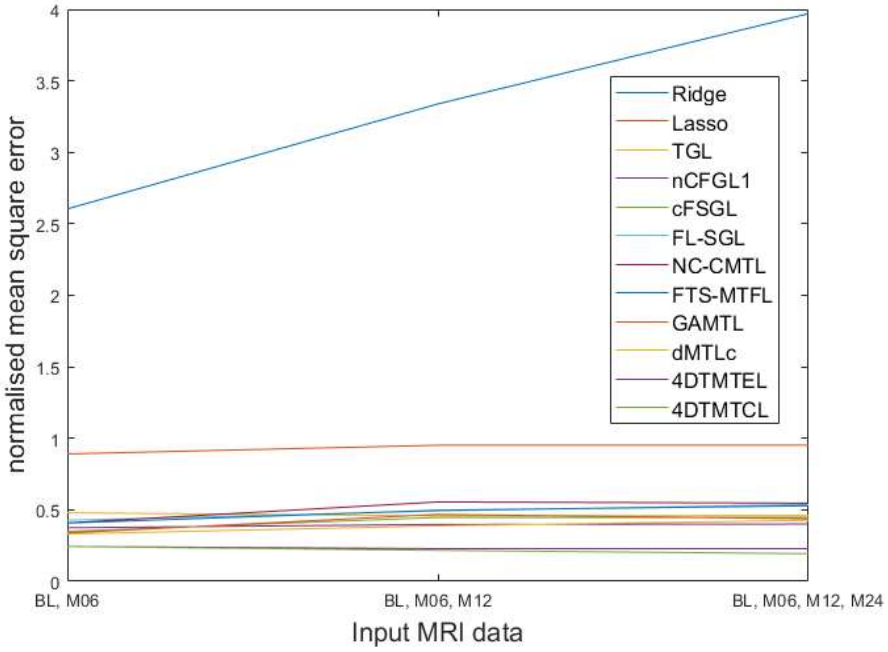


Figure 7.6: nMSE comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for ADAS-Cog prediction.

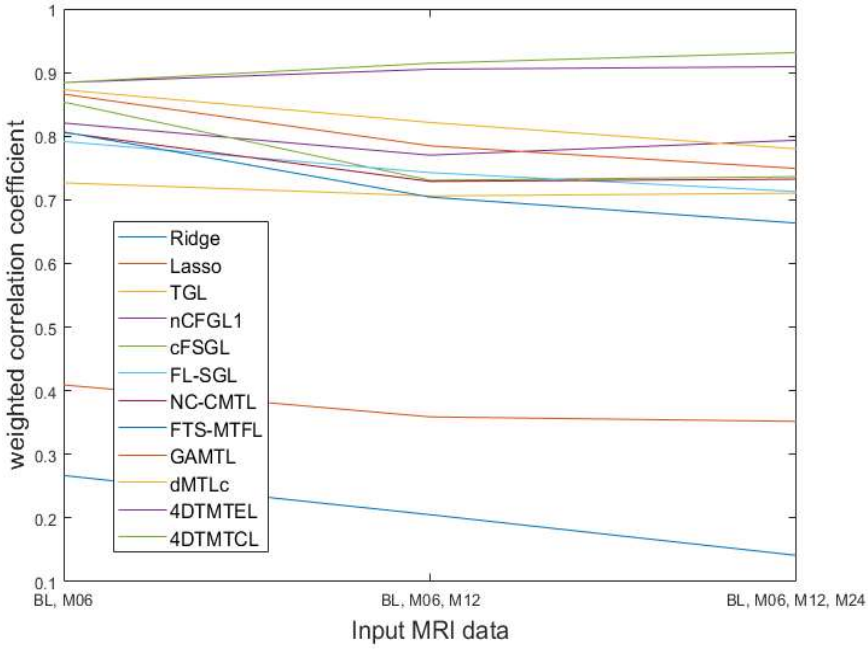


Figure 7.7: wR comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for ADAS-Cog prediction.

Table 7.5: Comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for RAVLT TOTAL at time points M12 to M48. The best results are bolded.

Target: ***RAVLT TOTAL***	**Input MRI data**	M12 rMSE	M24 rMSE	M36 rMSE	M48 rMSE
Ridge	BL, M06	7.9376±1.9086	8.2124±1.4203	8.8234±1.0348	9.1893±0.5844
	BL, M06, M12	-	8.8007±0.9214	9.6013±0.9954	9.3914±0.6852
	BL, M06, M12, M24	-	-	9.4451±0.5333	9.4749±0.8549
Lasso	BL, M06	5.3266±1.1728	5.7378±0.8868	6.3218±0.5375	6.4050±0.5636
	BL, M06, M12	-	5.2906±0.7313	6.0855±0.3243	6.6021±0.3645
	BL, M06, M12, M24	-	-	6.6543±0.8148	6.5371±0.2537
TGL	BL, M06	2.2520±0.6345	2.7132±0.7384	3.2253±0.8426	3.3035±0.6113
	BL, M06, M12	-	2.3232±0.8310	3.3428±0.3111	3.3346±0.1367
	BL, M06, M12, M24	-	-	2.9319±0.2928	2.9832±0.4650
nCFGL1	BL, M06	2.4384±0.6609	2.3252±0.6353	2.7824±0.6245	2.8225±0.4127
	BL, M06, M12	-	2.6925±0.2743	3.1080±0.6475	2.8572±0.4472
	BL, M06, M12, M24	-	-	2.5564±0.2099	2.7609±0.1828
cFSGL	BL, M06	2.3591±0.8820	2.4297±0.5859	2.6760±0.3778	3.0636±0.3502
	BL, M06, M12	-	2.2641±0.4062	2.9111±0.3892	3.5745±0.8169
	BL, M06, M12, M24	-	-	2.3136±0.2712	3.6462±0.8933
FL-SGL	BL, M06	2.2719±0.8909	2.4105±1.1230	3.2312±0.1693	2.6380±0.3065
	BL, M06, M12	-	2.1728±0.6094	3.3078±0.2724	3.7892±0.7611
	BL, M06, M12, M24	-	-	3.5879±0.5526	3.3378±0.5622
NC-CMTL	BL, M06	2.7621±0.2550	2.3219±0.3455	2.7690±0.1420	3.1030±1.0612
	BL, M06, M12	-	2.6269±0.8970	3.1795±0.2982	3.6578±0.3720
	BL, M06, M12, M24	-	-	3.1408±0.7582	3.8772±0.3989
FTS-MTFL	BL, M06	2.5827±0.6453	2.6332±0.7640	2.6991±0.7002	3.1862±0.6679
	BL, M06, M12	-	2.9855±0.8002	2.8321±0.8145	3.6179±0.1949
	BL, M06, M12, M24	-	-	2.8494±0.7425	3.2735±0.1476
GAMTL	BL, M06	2.7978±0.4072	2.7376±0.7520	3.2829±0.2948	2.7709±0.7505
	BL, M06, M12	-	2.9147±0.7963	3.3549±0.6479	2.6803±0.4160
	BL, M06, M12, M24	-	-	3.1316±0.6682	3.2177±0.3317
dMTLc	BL, M06	2.8858±0.1441	2.3956±0.4084	3.4184±0.4235	2.7441±0.7612
	BL, M06, M12	-	2.4132±0.2166	3.2582±0.6420	3.2824±0.3683
	BL, M06, M12, M24	-	-	3.2904±0.5136	3.1140±0.4140
4DTMTEL	BL, M06	**1.4314±0.1056**	1.6062±0.1550	1.6425±0.2317	1.5017±0.2467
	BL, M06, M12	-	1.5981±0.1639	1.5659±0.3415	1.5120±0.2408
	BL, M06, M12, M24	-	-	1.5741±0.1947	**1.4169±0.2357**
4DTMTCL	BL, M06	**1.4314±0.1056**	1.6062±0.1550	1.6425±0.2317	1.5017±0.2467
	BL, M06, M12	-	**1.5829±0.1405**	**1.5603±0.2057**	1.4287±0.2721
	BL, M06, M12, M24	-	-	1.5814±0.1748	1.4583±0.2265

Table 7.6: Continuation of Table 7.5, comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for RAVLT TOTAL at time points M60 to M120. The best results are bolded.

Target: ***RAVLT TOTAL***	M60 rMSE	M72 rMSE	M84 rMSE	M96 rMSE	M108 rMSE	M120 rMSE
Ridge	8.8236±0.8598	9.2771±1.2987	9.4471±0.9660	9.6810±1.1429	9.0462±1.4533	10.5317±1.7693
	9.6864±1.2575	9.2802±1.0148	9.5226±0.5733	10.1523±0.7347	9.4438±0.6998	10.5704±0.6914
	9.0315±0.9378	9.2327±0.9339	10.3465±1.0210	10.2515±0.8749	9.6306±0.5632	10.7428±1.1482
Lasso	7.1748±1.3071	7.6141±0.9423	8.1238±0.7045	7.9922±1.0334	8.0109±1.3453	8.2628±1.6440
	7.1012±1.1572	7.3180±0.9218	8.6879±0.7074	8.1670±0.8737	8.7347±0.8120	8.2623±0.3789
	7.5048±0.4498	7.4309±1.2445	9.1976±0.5128	7.9789±0.5191	8.7071±0.9851	9.3811±0.7010
TGL	2.8406±0.6077	3.4649±0.8560	4.6258±0.6558	6.7818±0.3483	6.9691±1.0105	7.5915±1.3104
	2.8831±0.8359	3.3458±0.2526	4.7725±0.2623	7.0538±0.3695	7.1591±0.6287	7.8032±1.2013
	2.9808±0.4199	3.3466±0.3480	4.7163±0.1068	7.1540±0.2570	6.9129±0.5762	8.9559±1.3608
nCFGL1	2.7932±0.6532	3.2405±0.6746	4.0655±0.5738	4.1362±1.2326	5.3567±0.8017	6.1909±0.7436
	2.6147±0.5162	2.8099±0.5101	3.9522±0.2536	4.6950±1.1791	5.1171±0.9857	6.9941±0.5808
	3.0624±0.6127	3.2116±0.8442	3.8652±0.5102	4.6956±1.3787	5.7927±0.6606	6.8852±0.7595
cFSGL	2.5154±0.4834	2.9631±0.9677	3.7082±0.5020	3.8204±0.7361	5.2445±1.1207	6.1520±0.5217
	3.1100±0.4778	2.7875±0.3285	4.2585±0.3313	4.1207±0.8173	5.7601±0.7476	6.3592±0.5610
	2.7076±0.8490	3.2241±0.9394	3.7763±0.3714	4.5403±0.3288	5.5547±0.3839	7.1099±0.4410
FL-SGL	2.7104±0.3276	3.4238±0.4533	3.6328±0.3646	3.5606±1.0927	5.0397±0.7480	5.7075±0.4890
	2.9273±0.5356	3.3863±0.1660	4.2130±0.5641	3.7728±0.1909	5.6941±0.3799	5.3804±0.3166
	3.4157±0.7649	3.3844±0.1217	4.4202±0.3905	4.4519±0.5589	5.0796±0.8646	5.8488±0.4255
NC-CMTL	3.1478±0.2801	2.9095±0.5860	3.8791±0.8183	4.1731±0.6744	5.1556±0.4280	6.4508±0.9031
	3.9082±0.1255	3.1074±0.5122	4.2884±0.4550	4.0514±0.5160	5.6976±0.2125	6.5843±0.7498
	3.6727±0.6546	3.5574±0.4481	4.1230±0.9414	4.9867±0.3840	5.8663±0.3233	6.7519±0.5930
FTS-MTFL	2.9759±0.5230	3.4540±0.8699	3.4254±0.3827	3.4863±0.7479	5.0220±0.4201	6.0470±1.0560
	3.0897±0.6313	4.0272±0.2445	4.3600±0.2904	4.0967±0.5048	5.1602±0.4048	7.2104±0.4693
	2.9469±0.6075	4.3877±0.1758	4.1270±0.2002	4.2450±0.2296	5.5538±0.2259	8.0735±0.5181
GAMTL	2.6841±0.9265	3.5007±0.3095	3.1905±0.1981	3.9752±0.7219	5.2273±0.2456	6.5358±0.6512
	3.1773±0.7391	4.3476±0.4101	3.7327±0.5021	4.2664±0.9268	5.3719±0.4385	7.2649±0.2507
	3.0584±0.9408	4.2743±0.6460	4.7784±0.4751	4.0618±1.0423	5.9255±0.8284	7.0283±0.7280
dMTLc	3.1043±0.6300	2.8967±0.5255	3.5972±0.6105	3.2667±0.4401	5.7278±0.4308	6.9348±0.4579
	3.3486±0.2651	2.9606±0.4286	4.0858±0.8040	3.7802±0.5004	5.9785±0.4390	7.2396±0.7095
	3.9273±0.3915	3.7343±0.5265	3.6884±0.3570	4.2501±0.2017	6.0975±0.6798	7.3750±0.6304
4DTMTEL	1.6537±0.1928	1.6270±0.2594	1.8475±0.2760	1.8646±0.2633	1.6571±0.2034	2.2641±0.4324
	1.6318±0.1559	1.6596±0.3249	1.7659±0.2717	1.9131±0.3287	1.6915±0.3036	2.2064±0.3271
	1.6190±0.1897	1.6327±0.2825	1.7905±0.3173	2.1802±0.3047	**1.6259±0.3837**	2.1294±0.3258
4DTMTCL	1.6537±0.1928	1.6270±0.2594	1.8475±0.2760	1.8646±0.2633	1.6571±0.2034	2.2641±0.4324
	1.6265±0.1398	1.6147±0.2981	1.7251±0.2862	**1.7912±0.2043**	1.8083±0.2994	2.1947±0.3446
	1.6073±0.1547	**1.6058±0.2513**	**1.6975±0.2576**	2.0503±0.3165	1.7482±0.3852	**2.0983±0.3104**

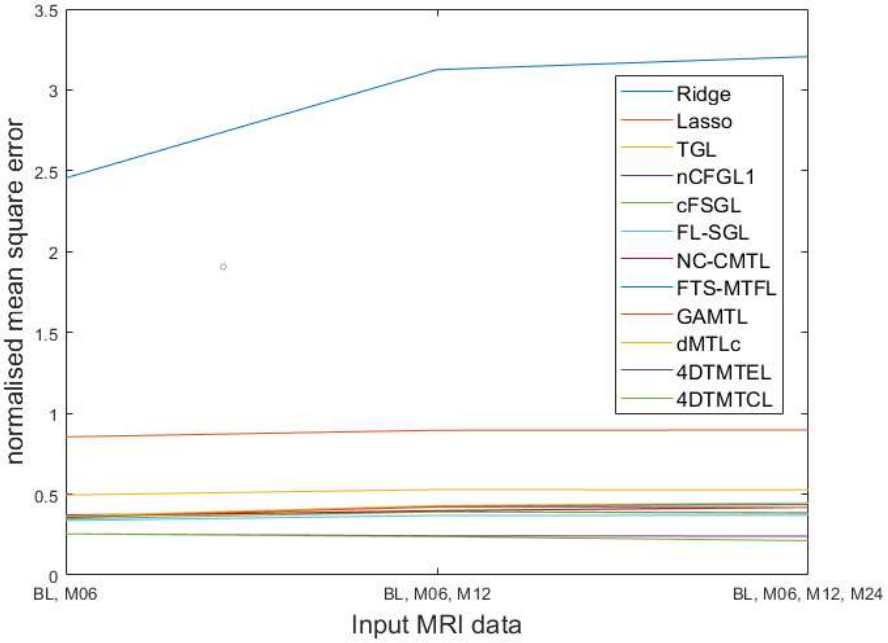


Figure 7.8: nMSE comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for RAVLT TOTAL prediction.

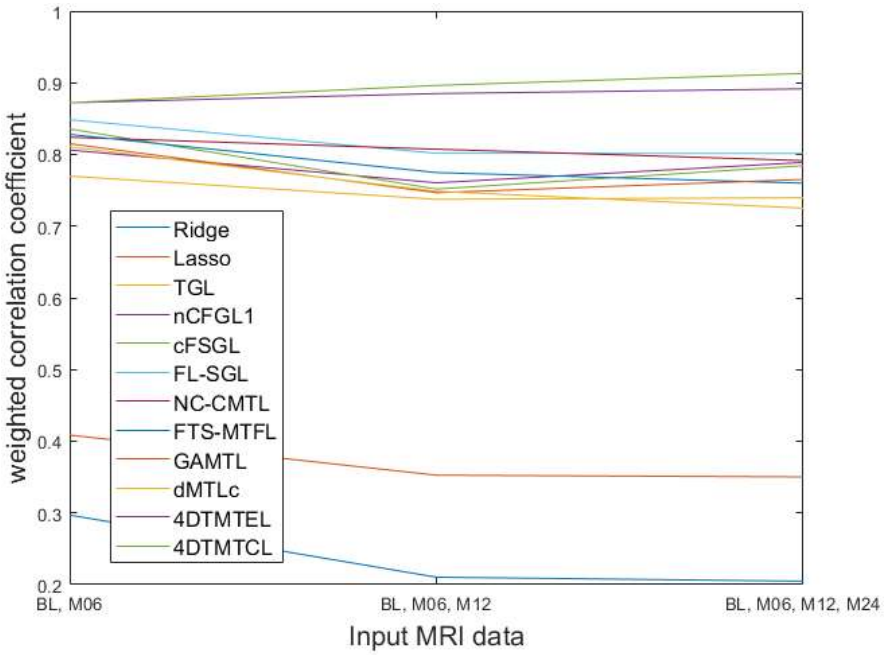


Figure 7.9: wR comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for RAVLT TOTAL prediction.

Table 7.7: Comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for FLU ANIM at time points M12 to M48. The best results are bolded.

Target: ***FLU ANIM***	**Input MRI data**	M12 rMSE	M24 rMSE	M36 rMSE	M48 rMSE
Ridge	BL, M06	5.4472±1.8837	6.1197±1.3662	4.2137±1.7492	6.4510±1.1308
	BL, M06, M12	-	6.5482±1.6059	5.5125±0.8588	6.8574±1.5865
	BL, M06, M12, M24	-	-	5.0159±1.3208	7.0810±0.9852
Lasso	BL, M06	2.5727±0.8031	1.8910±0.6070	2.4185±1.2328	1.9587±0.9794
	BL, M06, M12	-	2.4256±0.7036	2.7457±0.4112	2.0623±0.4206
	BL, M06, M12, M24	-	-	2.4505±0.6243	2.3042±0.4790
TGL	BL, M06	1.3364±0.7055	1.7315±0.6941	1.5969±0.6927	1.7610±0.6014
	BL, M06, M12	-	2.1237±0.8498	2.0413±0.4277	2.4378±0.2720
	BL, M06, M12, M24	-	-	1.9762±0.4580	1.8997±0.7941
nCFGL1	BL, M06	1.3493±0.7175	1.9113±0.3184	2.2461±0.6539	2.2189±0.7169
	BL, M06, M12	-	1.8783±0.3709	2.6614±0.4721	3.0950±0.5016
	BL, M06, M12, M24	-	-	3.1997±0.4066	3.0097±0.8738
cFSGL	BL, M06	1.4223±0.4127	1.6939±0.2266	1.8820±0.7462	2.0451±0.4260
	BL, M06, M12	-	2.0603±0.7598	1.9656±0.3674	2.8748±0.7519
	BL, M06, M12, M24	-	-	2.0217±0.3945	2.3547±0.3998
FL-SGL	BL, M06	1.4973±0.2302	1.7054±0.2693	2.0819±0.2348	2.3585±0.2543
	BL, M06, M12	-	1.9325±0.4688	2.9024±0.4167	3.3747±0.3591
	BL, M06, M12, M24	-	-	2.8441±0.6891	2.9753±0.4089
NC-CMTL	BL, M06	1.7739±0.8168	2.1327±0.6380	2.3885±0.2849	2.1918±0.3549
	BL, M06, M12	-	2.5618±0.2379	2.2544±0.6713	2.9353±0.6227
	BL, M06, M12, M24	-	-	2.7935±0.4557	3.3069±0.7748
FTS-MTFL	BL, M06	1.3370±0.5862	2.0966±0.7896	2.2042±0.2374	1.9100±0.3930
	BL, M06, M12	-	2.0998±0.7607	2.5010±0.5357	2.3996±0.6334
	BL, M06, M12, M24	-	-	2.1815±0.7246	2.3187±0.6297
GAMTL	BL, M06	1.8799±0.6390	2.1003±0.8619	2.3044±0.8389	1.6065±0.7484
	BL, M06, M12	-	2.5823±0.4788	2.3230±0.3129	1.9877±0.6910
	BL, M06, M12, M24	-	-	3.0064±0.4694	2.6007±0.7523
dMTLc	BL, M06	1.6090±0.3121	1.7672±0.2789	1.7853±0.5852	2.0958±0.7601
	BL, M06, M12	-	1.6253±0.1993	2.1077±0.7180	2.9203±0.2372
	BL, M06, M12, M24	-	-	2.4650±0.8052	2.6450±0.3055
4DTMTEL	BL, M06	**1.3153±0.2125**	1.4408±0.2174	1.4380±0.2559	1.6463±0.1761
	BL, M06, M12	-	1.3850±0.1853	1.4021±0.2336	1.5948±0.2039
	BL, M06, M12, M24	-	-	1.3172±0.1431	1.5765±0.2194
4DTMTCL	BL, M06	**1.3153±0.2125**	1.4408±0.2174	1.4380±0.2559	1.6463±0.1761
	BL, M06, M12	-	**1.3813±0.2107**	1.4702±0.1218	**1.5105±0.1845**
	BL, M06, M12, M24	-	-	**1.3116±0.1285**	1.5524±0.2091

Table 7.8: Continuation of Table 7.7, comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for FLU ANIM at time points M60 to M120. The best results are bolded.

Target: ***FLU ANIM***	M60 rMSE	M72 rMSE	M84 rMSE	M96 rMSE	M108 rMSE	M120 rMSE
Ridge	5.3020±1.2161	5.9618±1.6148	4.3180±2.1455	5.2603±2.2362	4.6195±1.0933	6.1033±1.1361
	5.5693±1.1430	6.5907±1.8627	6.4110±0.8108	6.6372±1.4649	5.0106±0.4828	7.6252±0.9186
	6.0497±0.9887	6.3894±1.2921	6.0122±0.5507	6.2665±1.4851	4.7195±0.3615	6.9532±1.5189
Lasso	1.9725±0.9758	2.3282±0.5055	3.3004±1.6036	3.0561±1.3526	4.2697±1.0291	5.5735±1.1790
	2.2068±0.4147	2.8199±0.9205	3.1723±0.3476	3.3834±0.5178	4.5455±0.8233	6.5209±0.6184
	2.2791±0.5415	3.2807±0.4456	3.1627±0.3656	3.7182±0.4234	4.4349±0.7695	6.5149±0.4812
TGL	2.2004±0.8175	2.1354±0.6304	1.9228±0.4608	2.7740±1.2036	3.6916±0.8542	4.2636±1.1491
	2.0988±1.0196	2.0820±0.9765	2.2707±0.9832	3.0330±0.9804	4.5487±0.6067	4.3870±0.5329
	3.0203±0.2544	2.3850±0.2158	2.4631±0.5045	2.8546±0.7618	5.0691±0.7155	5.3716±0.7334
nCFGL1	1.8437±0.5516	1.8129±1.1824	1.9417±1.2021	2.9784±1.2004	3.5532±0.4738	4.0436±0.5043
	2.2644±0.3923	2.2156±0.9846	2.0867±0.8407	3.5654±0.9716	4.0206±0.8272	4.3236±0.3788
	2.2346±0.9190	2.3067±0.7629	2.0179±0.2303	3.1751±0.4388	3.9389±0.2645	4.8759±0.8599
cFSGL	2.0344±0.3400	2.3855±0.7762	1.6734±0.9634	2.0328±0.4551	3.3721±0.7598	3.4367±0.9471
	2.3490±0.6695	2.3274±0.7137	2.0816±0.8770	2.5642±0.5770	2.8175±0.2571	3.4808±0.3352
	2.1410±0.7104	2.1426±0.5319	1.9803±0.5456	2.7469±0.4903	3.1952±0.5170	4.3147±0.8972
FL-SGL	2.3354±0.6026	2.2094±0.3552	2.0910±0.2739	1.7509±0.3284	2.9450±0.5829	3.7098±0.4743
	3.0345±0.5608	2.9181±0.4050	3.4040±0.6726	3.3757±0.2515	4.2381±0.6610	4.5604±0.7587
	2.7653±0.7930	2.6110±0.7322	3.8499±0.3755	3.7591±0.4290	4.2118±0.7539	4.9550±0.6282
NC-CMTL	2.1895±0.5107	1.9796±0.9002	2.3260±0.6398	2.7319±0.2494	2.6456±0.4240	3.7135±0.2451
	3.1523±0.6546	2.4896±0.8178	2.3108±0.6115	2.9214±0.2780	3.8373±0.6570	4.8834±0.5967
	3.4705±0.9121	3.0252±0.5887	3.0929±0.4736	3.2041±0.3453	4.0527±0.8115	5.8707±0.4044
FTS-MTFL	1.9178±0.4717	2.0398±1.0532	2.0742±0.7679	1.8550±0.4091	2.6720±0.2375	3.6095±0.8001
	2.7331±0.5047	3.0912±0.6570	2.9464±0.6349	2.4752±0.2606	3.7807±0.3022	4.3441±0.9561
	2.5687±0.5317	2.6284±0.2075	3.2762±0.6910	2.7514±0.4172	3.4830±0.4707	4.7875±0.9369
GAMTL	1.7183±0.7254	2.3639±0.4500	2.4301±0.4398	2.3459±0.4785	2.4059±0.1056	3.5854±0.9655
	2.3226±0.4731	2.3373±0.2426	2.6215±0.3299	3.1765±0.3801	3.8801±0.2533	4.6775±0.7379
	2.2895±0.4654	2.9514±0.3508	2.8965±0.2807	3.3024±0.3110	3.7154±0.2031	4.6918±0.5034
dMTLc	1.7940±0.3447	2.4372±0.6470	2.1386±0.8401	2.2573±0.4613	2.4334±0.4058	3.3250±0.5132
	2.3115±0.2783	3.1723±0.4607	3.3346±0.4201	3.3424±0.7040	2.9423±0.3250	3.3418±0.4535
	2.2363±0.2171	2.6860±0.3264	3.0190±0.7856	2.7889±0.6381	3.6434±0.2945	3.4655±0.4804
4DTMTEL	1.4964±0.2937	1.4092±0.1424	1.5367±0.1827	1.4207±0.1808	1.6149±0.2787	2.1355±0.2624
	1.4738±0.3297	1.4176±0.1759	1.7165±0.1641	1.4282±0.1859	2.1556±0.2650	2.1457±0.3161
	1.5009±0.3171	**1.3997±0.1055**	1.6038±0.1701	1.4269±0.2358	**1.5216±0.1664**	2.2379±0.2473
4DTMTCL	1.4964±0.2937	1.4092±0.1424	1.5367±0.1827	1.4207±0.1808	1.6149±0.2787	2.1355±0.2624
	1.4652±0.2771	1.4418±0.1356	**1.5189±0.1612**	1.4731±0.1815	1.6242±0.1497	**2.1146±0.1842**
	1.5264±0.2860	1.4328±0.1039	1.6153±0.1523	**1.4031±0.1683**	1.5542±0.1548	2.1282±0.1487

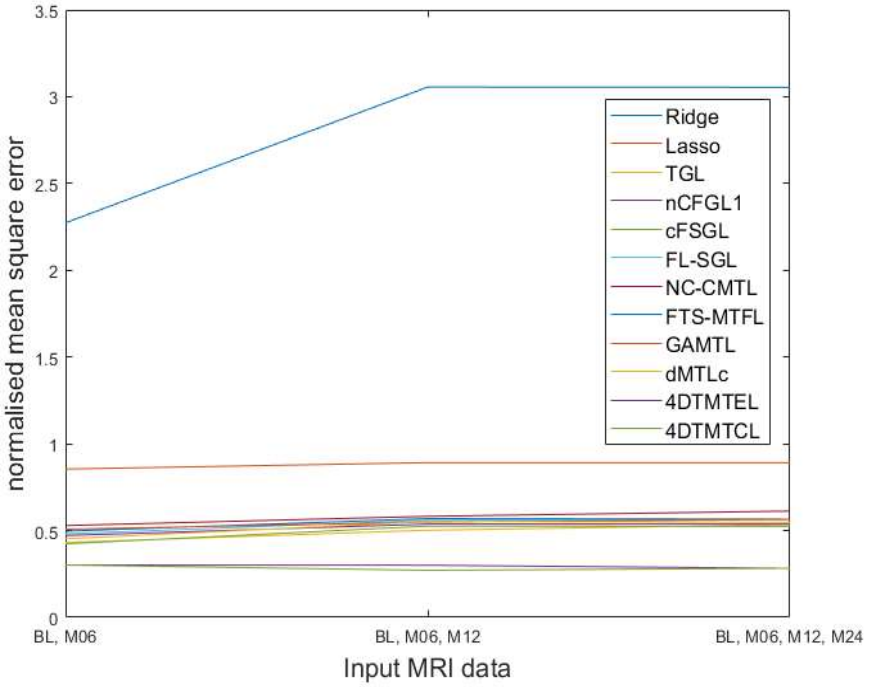


Figure 7.10: nMSE comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for FLU ANIM prediction.

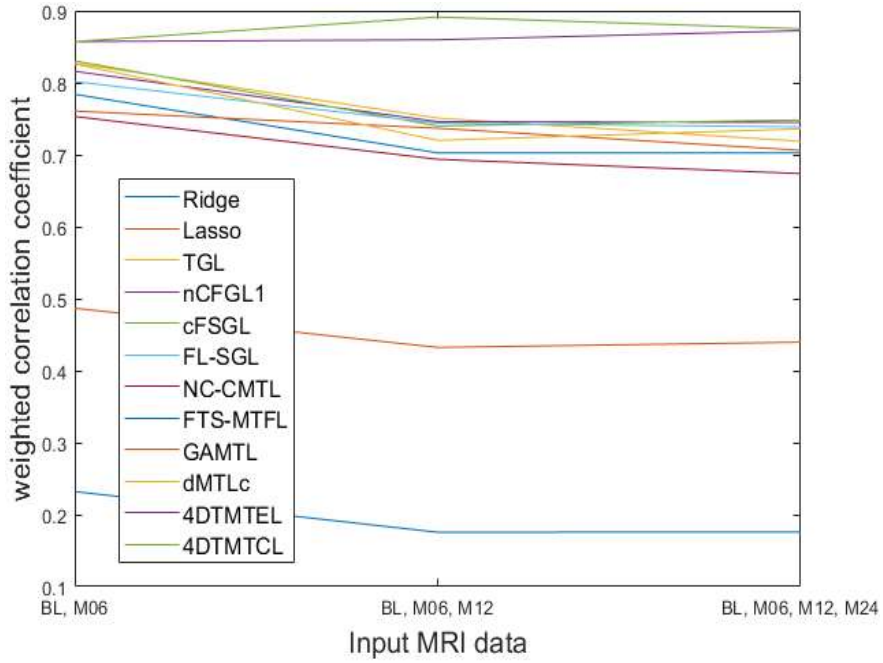


Figure 7.11: wR comparison of the results from our proposed approach with benchmarks and state-of-the-art methods for FLU ANIM prediction.

7.5 Summary

This chapter discussed, designed and constructed different multi-dimensional tensor multi-task learning regression based time-continuous algorithms to predict disease progression at different time points in neurological disease prediction scenarios in order to simultaneously overcome the problems of monotonic data forms, small datasets and time-continuous data scarcity. The problem of time-continuous data scarcity for AD dynamic prediction means that the number of available datasets decreases further as the disease progresses, and proposed approaches utilise a number of different methods to integrate time-continuous MRI recordings of patients in order to continuously improve the prediction accuracy of AD progression. Experimental results show that proposed approaches have the ability to diagnose and predict the AD progression, it can identify structural brain variants in individuals with AD, MCI and CN, and it only requires MRI data to achieve superior prediction performance, and by increasing the number of MRI data the performance of the models improves.

Chapter 8

Conclusions and Future Works

By now, it has reached the end of our investigation into multi-dimensional tensor multi-task learning in modelling disease progression. In this chapter, it would give a brief discussion about the strengths and weaknesses of the approach developed in this thesis and present a number of ideas for the future direction of the multi-dimensional tensor multi-task learning approach.

8.1 Conclusions

As reviewed in Chapter 2, the utilisation of machine learning algorithms to predict disease progression and identify biomarker information that can track disease progression is a hot topic in the field of computer science. However, existing prediction algorithms and frameworks cannot meet the ideal requirements for disease progression prediction applications due to various limitations. Which contain prediction accuracy and stability problems in medical small dataset scenarios, monotonic data formats (loss of multi-dimensional knowledge of the data and loss of correlation knowledge between biomarkers) and biomarker interpretability limitations. For instance, existing multi-task regression algorithms have problems with accuracy and stability in disease prediction applications, while deep learning-based algorithms have problems with low interpretability and high data requirements. The motivation for this research work is to investigate the possibility of quantifying and constructing AD biomarker data as a high-dimensional tensor, incorporating multi-task learning approaches to construct algorithms with high accuracy, stability and interpretability as a disease progression

prediction solution.

In order to achieve this goal, it is first necessary to investigate the types of tensor dimensions that are most applicable to AD progression prediction models and contain spatio-temporal information and knowledge of brain biomarkers, and approaches to quantify and construct multi-dimensional tensors. A quantification approach called Amalgamated magnitude-direction quantification in Chapter 4 offers a coherent solution with the ability to mine and incorporate multi-dimensional spatio-temporal information and biomarker correlation knowledge. It consists of two components, multi-dimensional knowledge vector construction and amalgamated magnitude-direction quantification of brain structural variation, it is a novel quantitative approach based on similarity calculations that simultaneously assesses and quantify the magnitude and direction information of brain structural variation. The quantification approach enables similarity of morphological trends between different biomarkers to be described as a third-order tensor with dimensions corresponding to the first biomarker, second biomarker and patient sample to address the problem of monotonic data forms. It contains comprehensive information about brain structural variation and can effectively differentiate between CN, MCI and AD patients.

Meanwhile, we designed and constructed multi-task learning regression algorithms and models incorporating multi-dimensional tensor data and mining MRI data for spatio-temporal structural variation information and knowledge to improve the accuracy, stability and interpretability of AD progression prediction in medical small dataset scenarios. In Chapter 5, we presented the components and workflow of the multi-dimensional tensor multi-task learning regression algorithm in detail. The algorithm consists of three components: supervised symmetric tensor decomposition for extracting biomarker latent factors, tensor multi-task learning regression and algorithmic regularisation terms. The proposed algorithm is designed to extract a set of first-order latent factors from the raw data, each represented by its first biomarker,

second biomarker and patient sample dimensions, to explicate the potential elements affecting the data variability in an interpretable manner and can be utilised as predictors for training prediction models, and these latent factors are shared by the predictions of each patient sample. The verification of multi-dimensional tensor multi-task learning regression in Chapter 5 illustrates that comparisons with single-task learning, benchmarks and state-of-the-art multi-task learning methods demonstrate that our approach outperforms competitors in terms of accuracy and stability of disease progression prediction. The main contributions are:

1. A novel similarity-based quantification approach (Amalgamated magnitude-direction quantification) is proposed to simultaneously capture the magnitude and directional correlation of structural variations among brain biomarkers, it contains comprehensive information on brain structural variations and can effectively differentiate CN, MCI and AD patients, and encodes MRI biomarker data into a third-order tensor, addressing the problem of monotonic data forms.
2. The proposed multi-dimensional tensor multi-task learning regression algorithm seamlessly integrates and shares spatio-temporal information and knowledge based on brain structural variation correlation and its biomarker latent factors. It utilises symmetric tensor decomposition techniques to learn task correlations from raw data and allows all samples to share latent knowledge of biomarkers based on brain structural variation, and significantly enhances the accuracy and stability of AD progression prediction in medical small dataset scenarios.
3. In Chapter 6 we identified and analysed important relative structural variation correlations between brain biomarkers that can be utilised to predict AD progression and can be applied as potential indicators for AD early identification.
4. For real-world applications, the 4D tensor multi-task learning approach for AD progression dynamic prediction presented in Chapter 7 integrates time-continuous MRI recordings of patients to continuously improve prediction accuracy of AD progression. It can additionally address the scarcity of time-continuous data for disease progression prediction while addressing the problem of monotonic data

forms and small medical datasets. Experimental results demonstrate that the prediction accuracy continues to improve as the number of MRI examinations increases.

8.2 Future directions

The main contributions of this research work are the significant attention given to the design and construction of multi-dimensional tensor multi-task learning for disease progression prediction with improved accuracy, stability and interpretability. There are still a lot of work that needs to be done in the next stage.

Firstly, the proposed quantification approach (Amalgamated magnitude-direction quantification) can utilise different concepts of multi-dimensional knowledge vector construction (in addition to brain structural variation quantification) from multiple perspectives to attempt to enhance the accuracy, stability and interpretability of disease progression prediction algorithms and results.

Secondly, the research work focuses on investigating the spatio-temporal information and knowledge of brain structural variation to predict disease progression, therefore only MRI biomarkers are utilised, with its ability to clearly demonstrate the structural conditions of the brain, but the proposed algorithms and models can be seamlessly implemented to other modalities of biomarkers (e.g., PET, CSF, genetic data) and it is also easy to apply multi-modal concepts on the proposed algorithms and models.

Thirdly, this research work focuses on the design and construction of tensor multi-task learning algorithms in three and four dimensions. Future work could further increase the dimensionality of the tensor to accommodate more dimensional information, thus improving the accuracy, stability and interpretability of the algorithms and results.

Fourthly, future plans are to engage and discuss this research with clinicians and to translate the research into a clinical management and care tool that can be utilised for AD patients. We will need to introduce clinicians to the principles, strengths, limitations and application scenarios of the model. The proposed model requires the ability to provide intuitive and clear outputs, explain the logic and rationale of its predictions, along with highlighting high-risk areas and key features in the patient's brain to gain the acceptance and trust of both clinicians and patients, avoiding the risk of causing unnecessary misinformation and anxiety. The proposed model will require testing and validation in various datasets and populations to ensure that it has a high level of specificity and sensitivity to distinguish between different stages and types of AD. A clinical utilisation system incorporating the model will be required to assist clinicians and patients in understanding the usage and considerations of the model along with enhancing the operability and convenience of the model. A model evaluation and feedback mechanism will be established to collect and analyse the data and results of the model usage by clinicians and patients, to evaluate the satisfaction and effectiveness of the model, to identify deficiencies and problems of the model, and to provide timely update and improvement of the model.

Finally, we will attempt to translate the approaches presented in this research to other domains, the current research has already been translated to the smart and sustainable agriculture domain for precision fertilisation prediction [337][338][339].

Reference

[1] Khachaturian, Z.S., 1985. Diagnosis of Alzheimer's disease. *Archives of neurology*, *42*(11), pp.1097-1105.

[2] Cummings, J.L., Vinters, H.V., Cole, G.M. and Khachaturian, Z.S., 1998. Alzheimer's disease: etiologies, pathophysiology, cognitive reserve, and treatment opportunities. *Neurology*, *51*(1 Suppl 1), pp.S2-S17.

[3] Jiang, T., Yu, J.T., Tian, Y. and Tan, L., 2013. Epidemiology and etiology of Alzheimer's disease: from genetic to non-genetic factors. *Current Alzheimer Research*, *10*(8), pp.852-867.

[4] Weller, J. and Budson, A., 2018. Current understanding of Alzheimer's disease diagnosis and treatment. *F1000Research*, *7*.

[5] DeTure, M.A. and Dickson, D.W., 2019. The neuropathological diagnosis of Alzheimer's disease. *Molecular neurodegeneration*, *14*(1), pp.1-18.

[6] Cao, P., Liu, X., Yang, J., Zhao, D., Huang, M. and Zaiane, O., 2018. ℓ2, 1− ℓ1 regularized nonlinear multi-task representation learning based cognitive performance prediction of Alzheimer's disease. *Pattern Recognition*, *79*, pp.195-215.

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

[8] Tabarestani, S., Aghili, M., Shojaie, M., Freytes, C. and Adjouadi, M., 2018, December. Profile-specific regression model for progression prediction of Alzheimer's disease using longitudinal data. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1353-1357). IEEE.

[9] Wan, J., Zhang, Z., Rao, B.D., Fang, S., Yan, J., Saykin, A.J. and Shen, L., 2014. Identifying the neuroanatomical basis of cognitive impairment in Alzheimer's disease by correlation-and nonlinearity-aware sparse Bayesian learning. *IEEE transactions on medical imaging*, *33*(7), pp.1475-1487.

[10] Wang, X., Zhen, X., Li, Q., Shen, D. and Huang, H., 2018. Cognitive assessment prediction in Alzheimer's disease by multi-layer multi-target regression. *Neuroinformatics*, *16*, pp.285-294.

[11] Zhou, J., Liu, J., Narayan, V.A., Ye, J. and Alzheimer's Disease Neuroimaging Initiative, 2013. Modeling disease progression via multi-task learning. *NeuroImage*, *78*, pp.233-248.

[12] Zhu, X., Suk, H.I., Wang, L., Lee, S.W., Shen, D. and Alzheimer's Disease Neuroimaging Initiative, 2017. A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. *Medical image analysis*, *38*, pp.205-214.

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

[14] Fenn, P. and Gray, A., 1999. Estimating long term cost savings from treatment of Alzheimer's disease: a modelling approach. *Pharmacoeconomics*, *16*, pp.165-174.

[15] Green, C., Shearer, J., Ritchie, C.W. and Zajicek, J.P., 2011. Model-based economic evaluation in Alzheimer's disease: a review of the methods available to model Alzheimer's disease progression. *Value in health*, *14*(5), pp.621-630.

[16] Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R. and Feng, D., 2014, April. Early diagnosis of Alzheimer's disease with deep learning. In *2014 IEEE 11th international symposium on biomedical imaging (ISBI)* (pp. 1015-1018). IEEE.

[17] Nguyen, M., He, T., An, L., Alexander, D.C., Feng, J., Yeo, B.T. and Alzheimer's Disease Neuroimaging Initiative, 2020. Predicting Alzheimer's disease progression using deep recurrent neural networks. *NeuroImage*, *222*, p.117203.

[18] van Oostveen, W.M. and de Lange, E.C., 2021. Imaging techniques in Alzheimer's disease: a review of applications in early diagnosis and longitudinal monitoring. *International journal of molecular sciences*, *22*(4), p.2110.

[19] Wang, T., Qiu, R.G. and Yu, M., 2018. Predictive modeling of the progression of Alzheimer's disease with recurrent neural networks. *Scientific reports*, *8*(1), p.9161.

[20] Jack, C.R. and Holtzman, D.M., 2013. Biomarker modeling of Alzheimer's disease. *Neuron*, *80*(6), pp.1347-1358.

[21] Counts, S.E., Ikonomovic, M.D., Mercado, N., Vega, I.E. and Mufson, E.J., 2017. Biomarkers for the early detection and progression of Alzheimer's disease. *Neurotherapeutics*, *14*, pp.35-53.

[22] Veitch, D.P., Weiner, M.W., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack Jr, C.R., Jagust, W., Morris, J.C. and Petersen, R.C., 2019. Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's & Dementia*, *15*(1), pp.106-152.

[23] Bertsch, M., Franchi, B., Marcello, N., Tesi, M.C. and Tosin, A., 2017. Alzheimer's disease: a mathematical model for onset and progression. *Mathematical medicine and biology: a journal of the IMA*, *34*(2), pp.193-214.

[24] Sanz-Arigita, E.J., Schoonheim, M.M., Damoiseaux, J.S., Rombouts, S.A., Maris, E., Barkhof, F., Scheltens, P. and Stam, C.J., 2010. Loss of 'small-world'networks in Alzheimer's disease: graph analysis of FMRI resting-state functional connectivity. *PloS one*, *5*(11), p.e13788.

[25] Liu, Y., Yu, C., Zhang, X., Liu, J., Duan, Y., Alexander-Bloch, A.F., Liu, B., Jiang, T. and Bullmore, E., 2014. Impaired long distance functional connectivity and weighted network architecture in Alzheimer's disease. *Cerebral Cortex*, *24*(6), pp.1422-1435.

[26] Hampel, H., O'Bryant, S.E., Durrleman, S., Younesi, E., Rojkova, K., Escott-Price, V., Corvol, J.C., Broich, K., Dubois, B., Lista, S. and Alzheimer Precision Medicine Initiative, 2017. A precision medicine initiative for Alzheimer's disease: the road ahead to biomarker-guided integrative disease modeling. *Climacteric*, *20*(2), pp.107-118.

[27] Fisher, C.K., Smith, A.M. and Walsh, J.R., 2019. Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Scientific reports*, *9*(1), p.13622.

[28] Samtani, M.N., Farnum, M., Lobanov, V., Yang, E., Raghavan, N., DiBernardo, A., Narayan, V. and Alzheimer's Disease Neuroimaging Initiative, 2012. An improved model for disease progression in patients from the Alzheimer's disease neuroimaging initiative. *The Journal of Clinical Pharmacology*, *52*(5), pp.629-644.

[29] Sun, B.L., Li, W.W., Zhu, C., Jin, W.S., Zeng, F., Liu, Y.H., Bu, X.L., Zhu, J., Yao, X.Q. and Wang, Y.J., 2018. Clinical research on Alzheimer's disease: progress and perspectives. *Neuroscience bulletin*, *34*, pp.1111-1118.

[30] Licher, S., Leening, M.J., Yilmaz, P., Wolters, F.J., Heeringa, J., Bindels, P.J., Alzheimer's Disease Neuroimaging Initiative, Vernooij, M.W., Stephan, B.C., Steyerberg, E.W. and Ikram, M.K., 2019. Development and validation of a dementia risk prediction model in the general population: an analysis of three longitudinal studies. *American Journal of Psychiatry*, *176*(7), pp.543-551.

[31] Ito, K., Corrigan, B., Zhao, Q., French, J., Miller, R., Soares, H., Katz, E., Nicholas, T., Billing, B., Anziano, R. and Fullerton, T., 2011. Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database. *Alzheimer's & Dementia*, *7*(2), pp.151-160.

[32] Ito, K., Ahadieh, S., Corrigan, B., French, J., Fullerton, T., Tensfeldt, T. and Alzheimer's Disease Working Group, 2010. Disease progression meta-analysis model in Alzheimer's disease. *Alzheimer's & Dementia*, *6*(1), pp.39-53.

[33] Wattmo, C., Wallin, Å.K. and Minthon, L., 2013. Progression of mild Alzheimer's disease: knowledge and prediction models required for future treatment strategies. *Alzheimer's research & therapy*, *5*(5), pp.1-15.

[34] Hu, C., Ju, R., Shen, Y., Zhou, P. and Li, Q., 2016, May. Clinical decision support for Alzheimer's disease based on deep learning and brain network. In *2016 IEEE international conference on communications (ICC)* (pp. 1-6). IEEE.

[35] Bucholc, M., Ding, X., Wang, H., Glass, D.H., Wang, H., Prasad, G., Maguire, L.P., Bjourson, A.J., McClean, P.L., Todd, S. and Finn, D.P., 2019. A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. *Expert systems with applications*, *130*, pp.157-171.

[36] Ballard, C., Gauthier, S., Corbett, A., Brayne, C., Aarsland, D. and Jones, E., 2011. Alzheimer's disease. *the Lancet*, *377*(9770), pp.1019-1031.

[37] Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C.E., Cummings, J. and van der Flier, W.M., 2021. Alzheimer's disease. *The Lancet*, *397*(10284), pp.1577-1590.

[38] Grant, W.B., Campbell, A., Itzhaki, R.F. and Savory, J., 2002. The significance of environmental factors in the etiology of Alzheimer's disease. *Journal of Alzheimer's disease*, *4*(3), pp.179-189.

[39] Rahman, M.A., Rahman, M.S., Uddin, M.J., Mamum-Or-Rashid, A.N.M., Pang, M.G. and Rhim, H., 2020. Emerging risk of environmental factors: insight mechanisms of Alzheimer's diseases. *Environmental Science and Pollution Research*, *27*, pp.44659-44672.

[40] Yegambaram, M., Manivannan, B., G Beach, T. and U Halden, R., 2015. Role of environmental contaminants in the etiology of Alzheimer's disease: a review. *Current Alzheimer Research*, *12*(2), pp.116-146.

[41] Francis, P.T., 2005. The interplay of neurotransmitters in Alzheimer's disease. *CNS spectrums*, *10*(S18), pp.6-9.

[42] Mohandas, E., Rajmohan, V. and Raghunath, B., 2009. Neurobiology of Alzheimer's disease. *Indian journal of psychiatry*, *51*(1), p.55.

[43] Chauhan, V. and Chauhan, A., 2006. Oxidative stress in Alzheimer's disease. *Pathophysiology*, *13*(3), pp.195-208.

[44] Huang, W.J., Zhang, X.I.A. and Chen, W.W., 2016. Role of oxidative stress in Alzheimer's disease. *Biomedical reports*, *4*(5), pp.519-522.

[45] Desai, A.K. and Grossberg, G.T., 2005. Diagnosis and treatment of Alzheimer's disease. *Neurology*, *64*(12 suppl 3), pp.S34-S39.

[46] Chu, L.W., 2012. Alzheimer's disease: early diagnosis and treatment. *Hong Kong Medical Journal*, *18*(3), p.228.

[47] De Roeck, E.E., De Deyn, P.P., Dierckx, E. and Engelborghs, S., 2019. Brief cognitive screening instruments for early detection of Alzheimer's disease: a systematic review. *Alzheimer's research & therapy*, *11*(1), pp.1-14.

[48] Porsteinsson, A.P., Isaacson, R.S., Knox, S., Sabbagh, M.N. and Rubino, I., 2021. Diagnosis of early Alzheimer's disease: clinical practice in 2021. *The journal of prevention of Alzheimer's disease*, *8*, pp.371-386.

[49] Johnson, K.A., Fox, N.C., Sperling, R.A. and Klunk, W.E., 2012. Brain imaging in Alzheimer disease. *Cold Spring Harbor perspectives in medicine*, *2*(4), p.a006213.

[50] Jagust, W., Gitcho, A., Sun, F., Kuczynski, B., Mungas, D. and Haan, M., 2006. Brain imaging evidence of preclinical Alzheimer's disease in normal aging. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *59*(4), pp.673-681.

[51] Alawode, D.O., Heslegrave, A.J., Ashton, N.J., Karikari, T.K., Simrén, J., Montoliu-Gaya, L., Pannee, J.A.O.C., O´ Connor, A., Weston, P.S., Lantero-Rodriguez, J. and Keshavan, A., 2021. Transitioning from cerebrospinal fluid to blood tests to facilitate diagnosis and disease monitoring in Alzheimer's disease. *Journal of internal medicine*, *290*(3), pp.583-601.

[52] O'Bryant, S.E., Edwards, M., Johnson, L., Hall, J., Villarreal, A.E., Britton, G.B., Quiceno, M., Cullum, C.M. and Graff-Radford, N.R., 2016. A blood screening test for Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *3*, pp.83-90.

[53] Vaz, M. and Silvestre, S., 2020. Alzheimer's disease: Recent treatment strategies. *European journal of pharmacology*, *887*, p.173554.

[54] Srivastava, S., Ahmad, R. and Khare, S.K., 2021. Alzheimer's disease and its treatment by different approaches: A review. *European Journal of Medicinal Chemistry*, *216*, p.113320.

[55] Hoerl, A.E. and Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), pp.55-67.

[56] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), pp.267-288

[57] Syed, A.H., Khan, T., Hassan, A., Alromema, N.A., Binsawad, M. and Alsayed, A.O., 2020. An ensemble-learning based application to predict the earlier stages of Alzheimer's disease (AD). *IEEE Access*, *8*, pp.222126-222143.

[58] Khoei, T.T., Labuhn, M.C., Caleb, T.D., Hu, W.C. and Kaabouch, N., 2021, May. A stacking-based ensemble learning model with genetic algorithm for detecting early stages of Alzheimer's disease. In *2021 IEEE International Conference on Electro Information Technology (EIT)* (pp. 215-222). IEEE.

[59] Lei, B., Yang, P., Wang, T., Chen, S. and Ni, D., 2017. Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis. *IEEE transactions on cybernetics*, *47*(4), pp.1102-1113.

[60] Zhu, X., Suk, H.I. and Shen, D., 2014. A novel multi-relation regularization method for regression and classification in AD diagnosis. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part III 17* (pp. 401-408). Springer International Publishing.

[61] Zhang, D., Shen, D. and Alzheimer's Disease Neuroimaging Initiative, 2012. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage*, *59*(2), pp.895-907.

[62] Brand, L., Nichols, K., Wang, H., Shen, L. and Huang, H., 2019. Joint multi-modal longitudinal regression and classification for alzheimer's disease prediction. *IEEE transactions on medical imaging*, *39*(6), pp.1845-1855.

[63] Pang, Z., Wang, X., Wang, X., Qi, J., Zhao, Z., Gao, Y., Yang, Y. and Yang, P., 2021. A multi-modal data platform for diagnosis and prediction of Alzheimer's disease using machine learning methods. *Mobile Networks and Applications*, *26*(6), pp.2341-2352.

[64] Mohs, R.C., Schmeidler, J. and Aryan, M., 2000. Longitudinal studies of cognitive, functional and behavioural change in patients with Alzheimer's disease. *Statistics in Medicine*, *19*(11-12), pp.1401-1409.

[65] Chen, X.R., Shao, Y., Sadowski, M.J. and Alzheimer's Disease Neuroimaging Initiative, 2021. Segmented Linear Mixed Model Analysis Reveals Association of the APOE ε4 Allele with Faster Rate of Alzheimer's Disease Dementia Progression. *Journal of Alzheimer's Disease*, *82*(3), pp.921-937.

[66] Guerrero, R., Schmidt-Richberg, A., Ledig, C., Tong, T., Wolz, R., Rueckert, D. and Alzheimer's Disease Neuroimaging Initiative (ADNI, 2016. Instantiated mixed effects modeling of Alzheimer's disease markers. *NeuroImage*, *142*, pp.113-125.

[67] Wang, Y.L., Chen, J., Du, Z.L., Weng, H., Zhang, Y., Li, R., Jia, Z., Sun, M., Jiang, J., Wang, F.Z. and Xu, J., 2021. Plasma p-tau181 level predicts neurodegeneration and progression to Alzheimer's dementia: a longitudinal study. *Frontiers in Neurology*, *12*, p.695696.

[68] Lei, B., Hou, W., Zou, W., Li, X., Zhang, C. and Wang, T., 2019. Longitudinal score prediction for Alzheimer's disease based on ensemble correntropy and spatial–temporal constraint. *Brain imaging and behavior*, *13*, pp.126-137.

[69] Hou, W., Lei, B., Zou, W., Li, X. and Zhang, C., 2017, April. Ensemble prediction of longitudinal scores of Alzheimer's disease based on ℓ 2, 1-norm regularized correntropy with spatial-temporal constraint. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (pp. 891-894). IEEE.

[70] Guan, H., Wang, L. and Liu, M., 2021, April. Multi-source domain adaptation via optimal transport for brain dementia identification. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)* (pp. 1514-1517). IEEE.

[71] Afzal, S., Maqsood, M., Khan, U., Mehmood, I., Nawaz, H., Aadil, F., Song, O.Y. and Yunyoung, N., 2021. Alzheimer disease detection techniques and methods: a review.

[72] Zhang, D. and Shen, D., 2011, September. Multi-modal multi-task learning for joint prediction of clinical scores in Alzheimer's disease. In *International Workshop on Multimodal Brain Image Analysis* (pp. 60-67). Berlin, Heidelberg: Springer Berlin Heidelberg.

[73] Prakash, M., Abdelaziz, M., Zhang, L., Strange, B.A., Tohka, J. and Alzheimer's Disease Neuroimaging Initiative, 2021. Quantitative longitudinal predictions of Alzheimer's disease by multi-modal predictive learning. *Journal of Alzheimer's Disease*, *79*(4), pp.1533-1546.

[74] Abd El Hamid, M.M., Omar, Y.M. and Mabrouk, M.S., 2016, December. Identifying genetic biomarkers associated to Alzheimer's disease using Support Vector Machine. In *2016 8th Cairo International Biomedical Engineering Conference (CIBEC)* (pp. 5-9). IEEE.

[75] Orru, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G. and Mechelli, A., 2012. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews*, *36*(4), pp.1140-1152.

[76] Lama, R.K. and Kwon, G.R., 2021. Diagnosis of Alzheimer's disease using brain network. *Frontiers in Neuroscience*, *15*, p.605115.

[77] Cui, X., Xiang, J., Guo, H., Yin, G., Zhang, H., Lan, F. and Chen, J., 2018. Classification of Alzheimer's disease, mild cognitive impairment, and normal controls with subnetwork selection

and graph Kernel principal component analysis based on minimum spanning tree brain functional network. *Frontiers in computational neuroscience*, *12*, p.31.

[78] Youssofzadeh, V., McGuinness, B., Maguire, L.P. and Wong-Lin, K., 2017. Multi-kernel learning with dartel improves combined MRI-PET classification of Alzheimer's disease in AIBL data: group and individual analyses. *Frontiers in human neuroscience*, *11*, p.380.

[79] Cheng, G. and Tong, X., 2018, July. Fuzzy clustering multiple kernel support vector machine. In *2018 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)* (pp. 7-12). IEEE.

[80] Rigatti, S.J., 2017. Random forest. *Journal of Insurance Medicine*, *47*(1), pp.31-39.

[81] Huang, L., Gao, Y., Jin, Y., Thung, K.H. and Shen, D., 2015. Soft-split sparse regression based random forest for predicting future clinical scores of Alzheimer's disease. In *Machine Learning in Medical Imaging: 6th International Workshop, MLMI 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Proceedings 6* (pp. 246-254). Springer International Publishing.

[82] Dimitriadis, S.I., Liparas, D. and Alzheimer's Disease Neuroimaging Initiative, 2018. How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI) database. *Neural regeneration research*, *13*(6), p.962.

[83] Huang, L., Jin, Y., Gao, Y., Thung, K.H., Shen, D. and Alzheimer's Disease Neuroimaging Initiative, 2016. Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. *Neurobiology of aging*, *46*, pp.180-191.

[84] Kim, J., Lee, M., Lee, M.K., Wang, S.M., Kim, N.Y., Kang, D.W., Um, Y.H., Na, H.R., Woo, Y.S., Lee, C.U. and Bahk, W.M., 2021. Development of random forest algorithm based prediction model of alzheimer's disease using neurodegeneration pattern. *Psychiatry Investigation*, *18*(1), p.69.

[85] Venugopalan, J., Tong, L., Hassanzadeh, H.R. and Wang, M.D., 2021. Multimodal deep learning models for early detection of Alzheimer's disease stage. *Scientific reports*, *11*(1), p.3254.

[86] Koga, S., Ikeda, A. and Dickson, D.W., 2022. Deep learning-based model for diagnosing Alzheimer's disease and tauopathies. *Neuropathology and Applied Neurobiology*, *48*(1), p.e12759.

[87] Zhang, F., Li, Z., Zhang, B., Du, H., Wang, B. and Zhang, X., 2019. Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease. *Neurocomputing*, *361*, pp.185-195.

[88] Spasov, S.E., Passamonti, L., Duggento, A., Lio, P. and Toschi, N., 2018, July. A multi-modal convolutional neural network framework for the prediction of Alzheimer's disease. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1271-1274). IEEE.

[89] Lee, G., Nho, K., Kang, B., Sohn, K.A. and Kim, D., 2019. Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Scientific reports*, *9*(1), p.1952.

[90] Zhang, T. and Shi, M., 2020. Multi-modal neuroimaging feature fusion for diagnosis of Alzheimer's disease. *Journal of Neuroscience Methods*, *341*, p.108795.

[91] Huang, Y., Xu, J., Zhou, Y., Tong, T., Zhuang, X. and Alzheimer's Disease Neuroimaging Initiative (ADNI), 2019. Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. *Frontiers in neuroscience*, *13*, p.509.

[92] Karasawa, H., Liu, C.L. and Ohwada, H., 2018. Deep 3d convolutional neural network architectures for alzheimer's disease diagnosis. In *Intelligent Information and Database Systems: 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, March 19-21, 2018, Proceedings, Part I 10* (pp. 287-296). Springer International Publishing.

[93] Soliman, S.A., El-Dahshan, E.S.A. and Salem, A.B.M., 2022, March. Deep learning 3D convolutional neural networks for predicting Alzheimer's disease (ALD). In *New Approaches for Multidimensional Signal Processing: Proceedings of International Workshop, NAMSP 2021* (pp. 151-162). Singapore: Springer Singapore.

[94] Song, X., Mao, M. and Qian, X., 2021. Auto-metric graph neural network based on a meta-learning strategy for the diagnosis of Alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, *25*(8), pp.3141-3152.

[95] Kim, M., Kim, J., Qu, J., Huang, H., Long, Q., Sohn, K.A., Kim, D. and Shen, L., 2021, December. Interpretable temporal graph neural network for prognostic prediction of Alzheimer's disease using longitudinal neuroimaging data. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1381-1384). IEEE.

[96] Zhang, Y., He, X., Chan, Y.H., Teng, Q. and Rajapakse, J.C., 2023. Multi-modal graph neural network for early diagnosis of Alzheimer's disease from sMRI and PET scans. *Computers in Biology and Medicine*, p.107328.

[97] Tang, Z., Chuang, K.V., DeCarli, C., Jin, L.W., Beckett, L., Keiser, M.J. and Dugger, B.N., 2019. Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline. *Nature communications*, *10*(1), p.2173.

[98] Qiu, S., Joshi, P.S., Miller, M.I., Xue, C., Zhou, X., Karjadi, C., Chang, G.H., Joshi, A.S., Dwyer, B., Zhu, S. and Kaku, M., 2020. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain*, *143*(6), pp.1920-1933.

[99] Pohl, T., Jakab, M. and Benesova, W., 2022. Interpretability of deep neural networks used for the diagnosis of Alzheimer's disease. *International Journal of Imaging Systems and Technology*, *32*(2), pp.673-686.

[100] Zhang, J., Zheng, B., Gao, A., Feng, X., Liang, D. and Long, X., 2021. A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification. *Magnetic Resonance Imaging*, *78*, pp.119-126.

[101] Qin, Z., Liu, Z., Guo, Q. and Zhu, P., 2022. 3D convolutional neural networks with hybrid attention mechanism for early diagnosis of Alzheimer's disease. *Biomedical Signal Processing and Control*, *77*, p.103828.

[102] Lam, P., Zhu, A.H., Gari, I.B., Jahanshad, N. and Thompson, P.M., 2020. 3D Grid-Attention Networks for Interpretable Age and Alzheimer's Disease Prediction from Structural MRI. *arXiv preprint arXiv:2011.09115*.

[103] Mahyoub, M., Randles, M., Baker, T. and Yang, P., 2018, June. Effective use of data science toward early prediction of Alzheimer's disease. In *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)* (pp. 1455-1461). IEEE.

[104] Zhang, J. and Wang, Y., 2019, October. Continually modeling Alzheimer's disease progression via deep multi-order preserving weight consolidation. In *International Conference on*

Medical Image Computing and Computer-Assisted Intervention (pp. 850-859). Cham: Springer International Publishing.

[105] Dhanusha, C. and Kumar, A.S., 2021, February. Deep recurrent Q reinforcement learning model to predict the Alzheimer disease using smart home sensor data. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1074, No. 1, p. 012014). IOP Publishing.

[106] Saboo, K., Choudhary, A., Cao, Y., Worrell, G., Jones, D. and Iyer, R., 2021. Reinforcement learning based disease progression model for Alzheimer's disease. *Advances in Neural Information Processing Systems*, *34*, pp.20903-20915.

[107] Zhang, Q., Du, Q. and Liu, G., 2021. A whole-process interpretable and multi-modal deep reinforcement learning for diagnosis and analysis of Alzheimer's disease∗. *Journal of Neural Engineering*, *18*(6), p.066032.

[108] Zhu, J., Tan, Y., Lin, R., Miao, J., Fan, X., Zhu, Y., Liang, P., Gong, J. and He, H., 2022. Efficient self-attention mechanism and structural distilling model for Alzheimer's disease diagnosis. *Computers in Biology and Medicine*, *147*, p.105737.

[109] Li, Y., Luo, J. and Zhang, J., 2022. Classification of Alzheimer's disease in MRI images using knowledge distillation framework: an investigation. *International Journal of Computer Assisted Radiology and Surgery*, *17*(7), pp.1235-1243.

[110] Kotia, J., Kotwal, A., Bharti, R. and Mangrulkar, R., 2021. Few shot learning for medical imaging. *Machine learning algorithms for industrial applications*, pp.107-132.

[111] Candemir, S., Nguyen, X.V., Folio, L.R. and Prevedello, L.M., 2021. Training strategies for radiology deep learning models in data-limited scenarios. *Radiology: Artificial Intelligence*, *3*(6), p.e210014.

[112] Ren, F., Yang, C. and Nanehkaran, Y.A., 2023. MRI-based model for MCI conversion using deep zero-shot transfer learning. *The Journal of Supercomputing*, *79*(2), pp.1182-1200.

[113] Lei, B., Yang, M., Yang, P., Zhou, F., Hou, W., Zou, W., Li, X., Wang, T., Xiao, X. and Wang, S., 2020. Deep and joint learning of longitudinal data for Alzheimer's disease prediction. *Pattern Recognition*, *102*, p.107247.

[114] Tang, S., Cao, P., Huang, M., Liu, X. and Zaiane, O., 2022. Dual feature correlation guided multi-task learning for Alzheimer's disease prediction. *Computers in Biology and Medicine*, *140*, p.105090.

[115] Caruana, R., 1997. Multitask learning. *Machine learning*, *28*, pp.41-75.

[116] Dong, D., Wu, H., He, W., Yu, D. and Wang, H., 2015, July. Multi-task learning for multiple language translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 1723-1732).

[117] Greenlaw, K., Szefer, E., Graham, J., Lesperance, M., Nathoo, F.S. and Alzheimer's Disease Neuroimaging Initiative, 2017. A Bayesian group sparse multi-task regression model for imaging genetics. *Bioinformatics*, *33*(16), pp.2513-2522.

[118] Bi, J., Xiong, T., Yu, S., Dundar, M. and Rao, R.B., 2008. An improved multi-task learning approach with applications in medical diagnosis. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part I 19* (pp. 117-132). Springer Berlin Heidelberg.

[119] Xue, Y., Liao, X., Carin, L. and Krishnapuram, B., 2007. Multi-Task Learning for Classification with Dirichlet Process Priors. *Journal of Machine Learning Research*, *8*(1).

[120] Evgeniou, T., Micchelli, C.A., Pontil, M. and Shawe-Taylor, J., 2005. Learning multiple tasks with kernel methods. *Journal of machine learning research*, *6*(4).

[121] Zhou, Y., Jin, R. and Hoi, S.C.H., 2010, March. Exclusive lasso for multi-task feature selection. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 988-995). JMLR Workshop and Conference Proceedings.

[122] Oliveira, S.H., Goncalves, A.R. and Von Zuben, F.J., 2022. Asymmetric multi-task learning with local transference. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *16*(5), pp.1-30.

[123] Nie, F., Hu, Z. and Li, X., 2018, July. Calibrated multi-task learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2012-2021).

[124] Widmer, C. and Rätsch, G., 2012, June. Multitask learning in computational biology.

In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning* (pp. 207-216). JMLR Workshop and Conference Proceedings.

[125] Zhang, Y. and Yeung, D.Y., 2012. A convex formulation for learning task relationships in multi-task learning. *arXiv preprint arXiv:1203.3536*.

[126] Weinberger, K., Dasgupta, A., Langford, J., Smola, A. and Attenberg, J., 2009, June. Feature hashing for large scale multitask learning. In *Proceedings of the 26th annual international conference on machine learning* (pp. 1113-1120).

[127] Zhou, J., Liu, J., Narayan, V.A. and Ye, J., 2012, August. Modeling disease progression via fused sparse group lasso. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1095-1103).

[128] Zhang, Y. and Yang, Q., 2021. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, *34*(12), pp.5586-5609.

[129] Hao, X., Bao, Y., Guo, Y., Yu, M., Zhang, D., Risacher, S.L., Saykin, A.J., Yao, X., Shen, L. and Alzheimer's Disease Neuroimaging Initiative, 2020. Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease. *Medical image analysis*, *60*, p.101625.

[130] Zhu, X., Suk, H.I. and Shen, D., 2014. Multi-modality canonical feature selection for Alzheimer's disease diagnosis. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part II 17* (pp. 162-169). Springer International Publishing.

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

[132] Suk, H.I., Lee, S.W., Shen, D. and Alzheimer's Disease Neuroimaging Initiative, 2016. Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. *Brain Structure and Function*, *221*, pp.2569-2587.

[133] Liu, X., Cao, P., Gonçalves, A.R., Zhao, D. and Banerjee, A., 2018. Modeling alzheimer's disease progression with fused laplacian sparse group LASSO. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *12*(6), pp.1-35.

[134] Cao, P., Liang, W., Zhang, K., Tang, S. and Yang, J., 2021, December. Joint feature and task aware multi-task feature learning for Alzheimer's disease diagnosis. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2643-2650). IEEE.

[135] Liu, X., Cao, P., Wang, J., Kong, J. and Zhao, D., 2019. Fused group lasso regularized multi-task feature learning and its application to the cognitive performance prediction of Alzheimer's disease. *Neuroinformatics*, *17*, pp.271-294.

[136] Cao, P., Tang, S., Huang, M., Yang, J., Zhao, D., Trabelsi, A. and Zaiane, O., 2019, November. Feature-aware Multi-task feature learning for Predicting Cognitive Outcomes in Alzheimer's disease. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1-5). IEEE.

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

[138] Gong, P., Zhou, J., Fan, W. and Ye, J., 2014, August. Efficient multi-task feature learning with calibration. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 761-770).

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

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

[141] Lei, B., Cheng, N., Frangi, A.F., Tan, E.L., Cao, J., Yang, P., Elazab, A., Du, J., Xu, Y. and Wang, T., 2020. Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease. *Medical image analysis*, *61*, p.101652.

[142] Jiang, P., Wang, X., Li, Q., Jin, L. and Li, S., 2018. Correlation-aware sparse and low-rank constrained multi-task learning for longitudinal analysis of Alzheimer's disease. *IEEE journal of biomedical and health informatics*, *23*(4), pp.1450-1456.

[143] Hamdi, S.M. and Angryk, R., 2019, November. Interpretable feature learning of graphs using tensor decomposition. In *2019 IEEE International Conference on Data Mining (ICDM)* (pp. 270-279). IEEE.

[144] Latchoumane, C.F.V., Vialatte, F.B., Solé-Casals, J., Maurice, M., Wimalaratna, S.R., Hudson, N., Jeong, J. and Cichocki, A., 2012. Multiway array decomposition analysis of EEGs in Alzheimer's disease. *Journal of neuroscience methods*, *207*(1), pp.41-50.

[145] Fang, J., 2019. Tightly integrated genomic and epigenomic data mining using tensor decomposition. *Bioinformatics*, *35*(1), pp.112-118.

[146] Jeong, J.Y. and Jun, C.H., 2019, November. Sparse tensor decomposition for multi-task interaction selection. In *2019 IEEE International Conference on Big Knowledge (ICBK)* (pp. 105-114). IEEE.

[147] Wang, M., Fischer, J. and Song, Y.S., 2019. Three-way clustering of multi-tissue multi-individual gene expression data using semi-nonnegative tensor decomposition. *The annals of applied statistics*, *13*(2), p.1103.

[148] Bahadori, M.T., Yu, Q.R. and Liu, Y., 2014. Fast multivariate spatio-temporal analysis via low rank tensor learning. *Advances in neural information processing systems*, *27*.

[149] Atluri, G., Karpatne, A. and Kumar, V., 2018. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, *51*(4), pp.1-41.

[150] Ding, Z., Newton, A.T., Xu, R., Anderson, A.W., Morgan, V.L. and Gore, J.C., 2013. Spatio-temporal correlation tensors reveal functional structure in human brain. *PloS one*, *8*(12), p.e82107.

[151] Yuan, H. and Li, G., 2021. A survey of traffic prediction: from spatio-temporal data to intelligent transportation. *Data Science and Engineering*, *6*, pp.63-85.

[152] Han, Y. and Moutarde, F., 2016. Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization. *International Journal of Intelligent Transportation Systems Research*, *14*, pp.36-49.

[153] Davidson, I., Gilpin, S., Carmichael, O. and Walker, P., 2013, August. Network discovery via constrained tensor analysis of fmri data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 194-202).

[154] McLeod, K., Sermesant, M., Beerbaum, P. and Pennec, X., 2015. Spatio-temporal tensor decomposition of a polyaffine motion model for a better analysis of pathological left ventricular dynamics. *IEEE transactions on medical imaging*, *34*(7), pp.1562-1575.

[155] Hung, H. and Wang, C.C., 2013. Matrix variate logistic regression model with application to EEG data. *Biostatistics*, *14*(1), pp.189-202.

[156] Li, X., Xu, D., Zhou, H. and Li, L., 2018. Tucker tensor regression and neuroimaging analysis. *Statistics in Biosciences*, *10*, pp.520-545.

[157] Zhou, H., Li, L. and Zhu, H., 2013. Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, *108*(502), pp.540-552.

[158] Latif, J., Xiao, C., Imran, A. and Tu, S., 2019, January. Medical imaging using machine learning and deep learning algorithms: a review. In *2019 2nd International conference on computing, mathematics and engineering technologies (iCoMET)* (pp. 1-5). IEEE.

[159] Qayyum, A., Qadir, J., Bilal, M. and Al-Fuqaha, A., 2020. Secure and robust machine learning for healthcare: A survey. *IEEE Reviews in Biomedical Engineering*, *14*, pp.156-180.

[160] Gaikwad, S.K., Gawali, B.W. and Yannawar, P., 2010. A review on speech recognition technique. *International Journal of Computer Applications*, *10*(3), pp.16-24.

[161] Muti, D. and Bourennane, S., 2007. Survey on tensor signal algebraic filtering. *Signal Processing*, *87*(2), pp.237-249.

[162] De Lathauwer, L., 2009, May. A survey of tensor methods. In *2009 IEEE international symposium on circuits and systems* (pp. 2773-2776). IEEE.

[163] Cong, F., Lin, Q.H., Kuang, L.D., Gong, X.F., Astikainen, P. and Ristaniemi, T., 2015. Tensor decomposition of EEG signals: a brief review. *Journal of neuroscience methods*, *248*, pp.59-69.

[164] Krishnan, S. and Athavale, Y., 2018. Trends in biomedical signal feature extraction. *Biomedical Signal Processing and Control*, *43*, pp.41-63.

[165] Guo, W., Kotsia, I. and Patras, I., 2011. Tensor learning for regression. *IEEE Transactions on Image Processing*, *21*(2), pp.816-827.

[166] Gahrooei, M.R., Yan, H., Paynabar, K. and Shi, J., 2021. Multiple tensor-on-tensor regression: An approach for modeling processes with heterogeneous sources of data. *Technometrics*, *63*(2), pp.147-159.

[167] Lock, E.F., 2018. Tensor-on-tensor regression. *Journal of Computational and Graphical Statistics*, *27*(3), pp.638-647.

[168] Hoff, P.D., 2015. Multilinear tensor regression for longitudinal relational data. *The annals of applied statistics*, *9*(3), p.1169.

[169] Kossaifi, J., Lipton, Z.C., Kolbeinsson, A., Khanna, A., Furlanello, T. and Anandkumar, A., 2020. Tensor regression networks. *The Journal of Machine Learning Research*, *21*(1), pp.4862-4882.

[170] Cao, X. and Rabusseau, G., 2017. Tensor regression networks with various low-rank tensor approximations. *arXiv preprint arXiv:1712.09520*.

[171] Kolbeinsson, A., Kossaifi, J., Panagakis, Y., Bulat, A., Anandkumar, A., Tzoulaki, I. and Matthews, P., 2019. Robust deep networks with randomized tensor regression layers. *arXiv*.

[172] Gao, C. and Wu, X.J., 2012. Kernel support tensor regression. *Procedia Engineering*, *29*, pp.3986-3990.

[173] Liu, H. and Wei, Z., 2010, May. An edge-adaptive structure tensor kernel regression for image interpolation. In *2010 2nd International Conference on Future Computer and Communication* (Vol. 2, pp. V2-681). IEEE.

[174] Zhao, Q., Zhou, G., Adali, T., Zhang, L. and Cichocki, A., 2013. Kernelization of tensor-based models for multiway data analysis: Processing of multidimensional structured data. *IEEE Signal Processing Magazine*, *30*(4), pp.137-148.

[175] Friedlander, M.P. and Hatz, K., 2008. Computing non-negative tensor factorizations. *Optimisation Methods and Software*, *23*(4), pp.631-647.

[176] Morup, M., Madsen, K.H. and Hansen, L.K., 2008, May. Approximate l 0 constrained non-negative matrix and tensor factorization. In *2008 IEEE international symposium on circuits and systems* (pp. 1328-1331). IEEE.

[177] Qiu, Y., Zhou, G., Wang, Y., Zhang, Y. and Xie, S., 2020. A generalized graph regularized non-negative tucker decomposition framework for tensor data representation. *IEEE transactions on cybernetics*, *52*(1), pp.594-607.

[178] Hazan, T., Polak, S. and Shashua, A., 2005, October. Sparse image coding using a 3D non-negative tensor factorization. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (Vol. 1, pp. 50-57). IEEE.

[179] Karahan, E., Rojas-Lopez, P.A., Bringas-Vega, M.L., Valdés-Hernández, P.A. and Valdes-Sosa, P.A., 2015. Tensor analysis and fusion of multimodal brain images. *Proceedings of the IEEE*, *103*(9), pp.1531-1559.

[180] He, L., Chen, K., Xu, W., Zhou, J. and Wang, F., 2018. Boosted sparse and low-rank tensor regression. *Advances in Neural Information Processing Systems*, *31*.

[181] Park, J.Y., Carr, K., Zheng, S., Yue, Y. and Yu, R., 2020, November. Multiresolution tensor learning for efficient and interpretable spatial analysis. In *International Conference on Machine Learning* (pp. 7499-7509). PMLR.

[182] Kolda, T.G. and Bader, B.W., 2009. Tensor decompositions and applications. *SIAM review*, *51*(3), pp.455-500.

[183] Wang, Y.X. and Zhang, Y.J., 2012. Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on knowledge and data engineering*, *25*(6), pp.1336-1353.

[184] Ji, Y., Wang, Q., Li, X. and Liu, J., 2019. A survey on tensor techniques and applications in machine learning. *IEEE Access*, *7*, pp.162950-162990.

[185] Anandkumar, A., Ge, R., Hsu, D., Kakade, S.M. and Telgarsky, M., 2014. Tensor decompositions for learning latent variable models. *Journal of machine learning research*, *15*, pp.2773-2832.

[186] Grasedyck, L., Kressner, D. and Tobler, C., 2013. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, *36*(1), pp.53-78.

[187] Tang, Y., Salakhutdinov, R. and Hinton, G., 2013, May. Tensor analyzers. In *International conference on machine learning* (pp. 163-171). PMLR.

[188] Bi, X., Tang, X., Yuan, Y., Zhang, Y. and Qu, A., 2021. Tensors in statistics. *Annual review of statistics and its application*, *8*, pp.345-368.

[189] Anandkumar, A., Ge, R. and Janzamin, M., 2015, June. Learning overcomplete latent variable models through tensor methods. In *Conference on Learning Theory* (pp. 36-112). PMLR.

[190] Ermiş, B., Acar, E. and Cemgil, A.T., 2015. Link prediction in heterogeneous data via generalized coupled tensor factorization. *Data Mining and Knowledge Discovery*, *29*, pp.203-236.

[191] Rabanser, S., Shchur, O. and Günnemann, S., 2017. Introduction to tensor decompositions and their applications in machine learning. *arXiv preprint arXiv:1711.10781*.

[192] Nickel, M. and Tresp, V., 2013. Tensor factorization for multi-relational learning. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part III 13* (pp. 617-621). Springer Berlin Heidelberg.

[193] Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C. and Phan, H.A., 2015. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE signal processing magazine*, *32*(2), pp.145-163.

[194] Papalexakis, E.E., Faloutsos, C. and Sidiropoulos, N.D., 2016. Tensors for data mining and data fusion: Models, applications, and scalable algorithms. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *8*(2), pp.1-44.

[195] Xu, J., Zhou, J., Tan, P.N., Liu, X. and Luo, L., 2019. Spatio-temporal multi-task learning via tensor decomposition. *IEEE Transactions on Knowledge and Data Engineering*, *33*(6), pp.2764-2775.

[196] Sivera, R., Delingette, H., Lorenzi, M., Pennec, X., Ayache, N. and Alzheimer's Disease Neuroimaging Initiative, 2019. A model of brain morphological changes related to aging and Alzheimer's disease from cross-sectional assessments. *NeuroImage*, *198*, pp.255-270.

[197] Reitz, C. and Mayeux, R., 2009. Endophenotypes in normal brain morphology and Alzheimer's disease: a review. *Neuroscience*, *164*(1), pp.174-190.

[198] Ohnishi, T., Matsuda, H., Tabira, T., Asada, T. and Uno, M., 2001. Changes in brain morphology in Alzheimer disease and normal aging: is Alzheimer disease an exaggerated aging process?. *American Journal of Neuroradiology*, *22*(9), pp.1680-1685.

[199] Roshchupkin, G.V., Adams, H.H., van der Lee, S.J., Vernooij, M.W., van Duijn, C.M., Uitterlinden, A.G., van der Lugt, A., Hofman, A., Niessen, W.J. and Ikram, M.A., 2016. Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology. *Neurobiology of aging*, *48*, pp.204-211.

[200] Vemuri, P., Wiste, H.J., Weigand, S.D., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Knopman, D.S., Petersen, R.C. and Jack, C.R., 2009. MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. *Neurology*, *73*(4), pp.294-301.

[201] Zhuang, L., Wen, W., Zhu, W., Trollor, J., Kochan, N., Crawford, J., Reppermund, S., Brodaty, H. and Sachdev, P., 2010. White matter integrity in mild cognitive impairment: a tract-based spatial statistics study. *Neuroimage*, *53*(1), pp.16-25.

[202] Chen, G., Ward, B.D., Xie, C., Li, W., Wu, Z., Jones, J.L., Franczak, M., Antuono, P. and Li, S.J., 2011. Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging. *Radiology*, *259*(1), pp.213-221.

[203] Si, T., Xing, G. and Han, Y., 2020. Subjective cognitive decline and related cognitive deficits. *Frontiers in Neurology*, *11*, p.247.

[204] Singh, V., Chertkow, H., Lerch, J.P., Evans, A.C., Dorr, A.E. and Kabani, N.J., 2006. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. *Brain*, *129*(11), pp.2885-2893.

[205] Pacheco, J., Goh, J.O., Kraut, M.A., Ferrucci, L. and Resnick, S.M., 2015. Greater cortical thinning in normal older adults predicts later cognitive impairment. *Neurobiology of aging*, *36*(2), pp.903-908.

[206] Cheng, C.P.W., Cheng, S.T., Tam, C.W.C., Chan, W.C., Chu, W.C.W. and Lam, L.C.W., 2018. Relationship between cortical thickness and neuropsychological performance in normal older adults and those with mild cognitive impairment. *Aging and disease*, *9*(6), p.1020.

[207] Putcha, D., Brickhouse, M., O'Keefe, K., Sullivan, C., Rentz, D., Marshall, G., Dickerson, B. and Sperling, R., 2011. Hippocampal hyperactivation associated with cortical thinning in Alzheimer's disease signature regions in non-demented elderly adults. *Journal of Neuroscience*, *31*(48), pp.17680-17688.

[208] Wee, C.Y., Yap, P.T., Shen, D. and Alzheimer's Disease Neuroimaging Initiative, 2013. Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Human brain mapping*, *34*(12), pp.3411-3425.

[209] Li, Y., Wang, Y., Wu, G., Shi, F., Zhou, L., Lin, W., Shen, D. and Alzheimer's Disease Neuroimaging Initiative, 2012. Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. *Neurobiology of aging*, *33*(2), pp.427-e15.

[210] Bakkour, A., Morris, J.C. and Dickerson, B.C., 2009. The cortical signature of prodromal AD: regional thinning predicts mild AD dementia. *Neurology*, *72*(12), pp.1048-1055.

[211] Dai, D., He, H., Vogelstein, J.T. and Hou, Z., 2013. Accurate prediction of AD patients using cortical thickness networks. *Machine vision and applications*, *24*, pp.1445-1457.

[212] He, Y., Chen, Z. and Evans, A., 2008. Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease. *Journal of Neuroscience*, *28*(18), pp.4756-4766.

[213] Hojjati, S.H., Ebrahimzadeh, A., Khazaee, A., Babajani-Feremi, A. and Alzheimer's Disease Neuroimaging Initiative, 2017. Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM. *Journal of neuroscience methods*, *282*, pp.69-80.

[214] He, Y., Chen, Z., Gong, G. and Evans, A., 2009. Neuronal networks in Alzheimer's disease. *The Neuroscientist*, *15*(4), pp.333-350.

[215] Tijms, B.M., Wink, A.M., de Haan, W., van der Flier, W.M., Stam, C.J., Scheltens, P. and Barkhof, F., 2013. Alzheimer's disease: connecting findings from graph theoretical studies of brain networks. *Neurobiology of aging*, *34*(8), pp.2023-2036.

[216] Asuero, A.G., Sayago, A. and González, A.G., 2006. The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, *36*(1), pp.41-59.

[217] Gogtay, N.J. and Thatte, U.M., 2017. Principles of correlation analysis. *Journal of the Association of Physicians of India*, *65*(3), pp.78-81.

[218] Huberty, C.J., 1975. Discriminant analysis. *Review of Educational Research*, *45*(4), pp.543-598.

[219] Cacoullos, T. ed., 2014. *Discriminant analysis and applications*. Academic Press.

[220] Johnson, R.A. and Wichern, D.W., 2002. Applied multivariate statistical analysis.

[221] Izenman, A.J., 2008. *Modern multivariate statistical techniques* (Vol. 1). New York: Springer.

[222] Vihinen, M., 2001. Bioinformatics in proteomics. *Biomolecular Engineering*, *18*(5), pp.241-248.

[223] Horner, D.S., Pavesi, G., Castrignano, T., De Meo, P.D.O., Liuni, S., Sammeth, M., Picardi, E. and Pesole, G., 2010. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Briefings in bioinformatics*, *11*(2), pp.181-197.

[224] Liu, J., Li, M., Pan, Y., Lan, W., Zheng, R., Wu, F.X. and Wang, J., 2017. Complex brain network analysis and its applications to brain disorders: a survey. *Complexity*, *2017*.

[225] Adourian, A., Jennings, E., Balasubramanian, R., Hines, W.M., Damian, D., Plasterer, T.N., Clish, C.B., Stroobant, P., McBurney, R., Verheij, E.R. and Bobeldijk, I., 2008. Correlation network analysis for data integration and biomarker selection. *Molecular BioSystems*, *4*(3), pp.249-259.

[226] Wagner, K.H., Cameron-Smith, D., Wessner, B. and Franzke, B., 2016. Biomarkers of aging: from function to molecular biology. *Nutrients*, *8*(6), p.338.

[227] Gomez Ravetti, M., Rosso, O.A., Berretta, R. and Moscato, P., 2010. Uncovering molecular biomarkers that correlate cognitive decline with the changes of hippocampus' gene expression profiles in Alzheimer's disease. *PloS one*, *5*(4), p.e10153.

[228] Brown, C. and Liebovitch, L., 2010. *Fractal analysis* (Vol. 165). Sage.

[229] King, R.D., Brown, B., Hwang, M., Jeon, T., George, A.T. and Alzheimer's Disease Neuroimaging Initiative, 2010. Fractal dimension analysis of the cortical ribbon in mild Alzheimer's disease. *Neuroimage*, *53*(2), pp.471-479.

[230] Freund, R.J., Wilson, W.J. and Sa, P., 2006. *Regression analysis*. Elsevier.

[231] Chatterjee, S. and Hadi, A.S., 2006. *Regression analysis by example*. John Wiley & Sons.

[232] Liu, Yang, et al. "A time-incorporated SOFA score-based machine learning model for predicting mortality in critically ill patients: A multicenter, real-world study." *International Journal of Medical Informatics* 163 (2022): 104776.

[233] Wang, H., Nie, F., Huang, H., Yan, J., Kim, S., Risacher, S., Saykin, A. and Shen, L., 2012. High-order multi-task feature learning to identify longitudinal phenotypic markers for alzheimer's disease progression prediction. *Advances in neural information processing systems*, *25*.

[234] Nawaz, S. and Yuan, Y., 2016. Computational pathology: Exploring the spatial dimension of tumor ecology. *Cancer letters*, *380*(1), pp.296-303.

[235] Gao, J., Tian, L., Wang, J., Chen, Y., Song, B. and Hu, X., 2020. Similar disease prediction with heterogeneous disease information networks. *IEEE Transactions on NanoBioscience*, *19*(3), pp.571-578.

[236] Abdulkareem, K.H., Al-Mhiqani, M.N., Dinar, A.M., Mohammed, M.A., Al-Imari, M.J., Al-Waisy, A.S., Alghawli, A.S. and Al-Qaness, M.A., 2022. MEF: Multidimensional Examination Framework for Prioritization of COVID-19 Severe Patients and Promote Precision Medicine Based on Hybrid Multi-Criteria Decision-Making Approaches. *Bioengineering*, *9*(9), p.457.

[237] Kwak, G.H.J. and Hui, P., 2019. DeepHealth: Deep Learning for Health Informatics reviews, challenges, and opportunities on medical imaging, electronic health records, genomics, sensing, and online communication health. *arXiv preprint arXiv:1909.00384*.

[238] Kellogg, R.A., Dunn, J. and Snyder, M.P., 2018. Personal omics for precision health. *Circulation Research*, *122*(9), pp.1169-1171.

[239] Kasl, S.V., 1996. The influence of the work environment on cardiovascular health: a historical, conceptual, and methodological perspective. *Journal of occupational health psychology*, *1*(1), p.42.

[240] Navarro-Peternella, F.M. and Marcon, S.S., 2012. Quality of life of a person with Parkinson's disease and the relationship between the time of evolution and the severity of the disease. *Revista latino-americana de enfermagem*, *20*, pp.384-391.

[241] Cordella, M., Poiani, A., Cordella, M. and Poiani, A., 2014. The Behavioural Dimension of Cancer and Sickness. *Behavioural Oncology: Psychological, Communicative, and Social Dimensions*, pp.11-79.

[242] Unger-Saldaña, K., Peláez-Ballestas, I. and Infante-Castañeda, C., 2012. Development and validation of a questionnaire to assess delay in treatment for breast cancer. *BMC cancer*, *12*(1), pp.1-13.

[243] Weinstein, N.D., 2003. Exploring the links between risk perceptions and preventive health behavior. *Social psychological foundations of health and illness*, *22*, p.53.

[244] Wang, J., Ma, J.J., Liu, J., Zeng, D.D., Song, C. and Cao, Z., 2017. Prevalence and risk factors of comorbidities among hypertensive patients in China. *International journal of medical sciences*, *14*(3), p.201.

[245] Hammond, T.C., Xing, X., Wang, C., Ma, D., Nho, K., Crane, P.K., Elahi, F., Ziegler, D.A., Liang, G., Cheng, Q. and Yanckello, L.M., 2020. β-amyloid and tau drive early Alzheimer's disease decline while glucose hypometabolism drives late decline. *Communications biology*, *3*(1), p.352.

[246] Hwang, J., Na, S., Lee, H. and Lee, D., 2009. Correlation between preoperative serum levels of five biomarkers and relationships between these biomarkers and cancer stage in epithelial overian cancer. *Journal of Gynecologic Oncology*, *20*(3), pp.169-175.

[247] Köbel, M., Kalloger, S.E., Boyd, N., McKinney, S., Mehl, E., Palmer, C., Leung, S., Bowen, N.J., Ionescu, D.N., Rajput, A. and Prentice, L.M., 2008. Ovarian carcinoma subtypes are different diseases: implications for biomarker studies. *PLoS medicine*, *5*(12), p.e232.

[248] Xia, D., Casanova, R., Machiraju, D., McKee, T.D., Weder, W., Beck, A.H. and Soltermann, A., 2018. Computationally-guided development of a stromal inflammation histologic biomarker in lung squamous cell carcinoma. *Scientific reports*, *8*(1), pp.1-13.

[249] Malta-Santos, H., Fukutani, K.F., Sorgi, C.A., Queiroz, A.T., Nardini, V., Silva, J., Lago, A., Carvalho, L.P., Machado, P.L., Bozza, P.T. and Franca-Costa, J., 2020. Multi-omic analyses of plasma cytokines, lipidomics, and transcriptomics distinguish treatment outcomes in cutaneous leishmaniasis. *Iscience*, *23*(12), p.101840.

[250] Sambyal, N., Saini, P. and Syal, R., 2021. A review of statistical and machine learning techniques for microvascular complications in type 2 diabetes. *Current Diabetes Reviews*, *17*(2), pp.143-155.

[251] Santiago, J.A. and Potashkin, J.A., 2015. Network-based metaanalysis identifies HNF4A and PTBP1 as longitudinally dynamic biomarkers for Parkinson's disease. *Proceedings of the National Academy of Sciences*, *112*(7), pp.2257-2262.

[252] Silverman, E.K., Schmidt, H.H., Anastasiadou, E., Altucci, L., Angelini, M., Badimon, L., Balligand, J.L., Benincasa, G., Capasso, G., Conte, F. and Di Costanzo, A., 2020. Molecular networks in Network Medicine: Development and applications. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *12*(6), p.e1489.

[253] Wu, S., Zhang, H.R. and Ré, C., 2020. Understanding and improving information transfer in multi-task learning. *arXiv preprint arXiv:2005.00944*.

[254] Søgaard, A. and Goldberg, Y., 2016, August. Deep multi-task learning with low level tasks supervised at lower layers. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 231-235).

[255] Bonilla, E.V., Agakov, F.V. and Williams, C.K., 2007, March. Kernel multi-task learning using task-specific features. In *Artificial Intelligence and Statistics* (pp. 43-50). PMLR.

[256] Fei, H. and Huan, J., 2013. Structured feature selection and task relationship inference for multi-task learning. *Knowledge and information systems*, *35*, pp.345-364.

[257] Sener, O. and Koltun, V., 2018. Multi-task learning as multi-objective optimization. *Advances in neural information processing systems*, *31*.

[258] Lin, X., Zhen, H.L., Li, Z., Zhang, Q.F. and Kwong, S., 2019. Pareto multi-task learning. *Advances in neural information processing systems*, *32*.

[259] Zheng, Z., Wang, Y., Dai, Q., Zheng, H. and Wang, D., 2019, August. Metadata-driven Task Relation Discovery for Multi-task Learning. In *IJCAI* (pp. 4426-4432).

[260] Liu, S., Qu, M., Zhang, Z., Cai, H. and Tang, J., 2021. Multi-task learning with domain knowledge for molecular property prediction. In *NeurIPS 2021 AI for Science Workshop*.

[261] Han, L. and Zhang, Y., 2015, August. Learning tree structure in multi-task learning. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 397-406).

[262] Malakouti, S. and Hauskrecht, M., 2019, November. Hierarchical adaptive multi-task learning framework for patient diagnoses and diagnostic category classification. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 701-706). IEEE.

[263] Zhao, L., Sun, Q., Ye, J., Chen, F., Lu, C.T. and Ramakrishnan, N., 2015, August. Multi-task learning for spatio-temporal event forecasting. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1503-1512).

[264] Zhang, Y., Yang, Y., Zhou, W., Wang, H. and Ouyang, X., 2021. Multi-city traffic flow forecasting via multi-task learning. *Applied Intelligence*, pp.1-19.

[265] Tan, Z., De, G., Li, M., Lin, H., Yang, S., Huang, L. and Tan, Q., 2020. Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine. *Journal of cleaner production*, *248*, p.119252.

[266] Fan, H., Yang, Z., Zhang, X., Wu, S., Long, J. and Liu, L., 2022. A novel multi-satellite and multi-task scheduling method based on task network graph aggregation. *Expert Systems with Applications*, *205*, p.117565.

[267] Yadwadkar, N.J., Hariharan, B., Gonzalez, J.E. and Katz, R., 2016. Multi-task learning for straggler avoiding predictive job scheduling. *The Journal of Machine Learning Research*, *17*(1), pp.3692-3728.

[268] Sun, X., Kashima, H., Tomioka, R., Ueda, N. and Li, P., 2011, December. A new multi-task learning method for personalized activity recognition. In *2011 IEEE 11th International Conference on Data Mining* (pp. 1218-1223). IEEE.

[269] Ruder, S., 2017. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

[270] Guo, P., Lee, C.Y. and Ulbricht, D., 2020, November. Learning to branch for multi-task learning. In *International Conference on Machine Learning* (pp. 3854-3863). PMLR.

[271] Zhang, Y. and Yeung, D.Y., 2013. Learning high-order task relationships in multi-task learning. In *IJCAI International Joint Conference on Artificial Intelligence* (p. 1917).

[272] Murugesan, K., Liu, H., Carbonell, J. and Yang, Y., 2016. Adaptive smoothed online multi-task learning. *Advances in Neural Information Processing Systems*, *29*.

[273] Crawshaw, M., 2020. Multi-task learning with deep neural networks: A survey. *arXiv preprint arXiv:2009.09796*.

[274] Thung, K.H. and Wee, C.Y., 2018. A brief review on multi-task learning. *Multimedia Tools and Applications*, *77*, pp.29705-29725.

[275] Liang, W., Zhang, K., Cao, P., Liu, X., Yang, J. and Zaiane, O., 2021. Rethinking modeling Alzheimer's disease progression from a multi-task learning perspective with deep recurrent neural network. *Computers in Biology and Medicine*, *138*, p.104935.

[276] Maggio, V., Chierici, M., Jurman, G. and Furlanello, C., 2018. Distillation of the clinical algorithm improves prognosis by multi-task deep learning in high-risk neuroblastoma. *PloS one*, *13*(12), p.e0208924.

[277] Zhang, W., Yang, G., Zhang, N., Xu, L., Wang, X., Zhang, Y., Zhang, H., Del Ser, J. and de Albuquerque, V.H.C., 2021. Multi-task learning with multi-view weighted fusion attention for artery-specific calcification analysis. *Information Fusion*, *71*, pp.64-76.

[278] Liu, X., Cao, P., Zhao, D. and Banerjee, A., 2016, May. Multi-task spare group lasso for characterizing Alzheimers disease. In *5th Workshop on Data Mining for Medicine and Healthcare* (Vol. 49).

[279] Xie, Q., Wang, S., Zhu, J., Zhang, X. and Alzheimer′ s Disease Neuroimaging Initiative, 2016. Modeling and predicting AD progression by regression analysis of sequential clinical data. *Neurocomputing*, *195*, pp.50-55.

[280] Chen, W., Long, G., Yao, L. and Sheng, Q.Z., 2020. AMRNN: attended multi-task recurrent neural networks for dynamic illness severity prediction. *World Wide Web*, *23*, pp.2753-2770.

[281] Zhang, Y. and Yang, Q., 2018. An overview of multi-task learning. *National Science Review*, *5*(1), pp.30-43.

[282] Li, X., Kao, Y., Shen, W., Li, X. and Xie, G., 2017, March. Lung nodule malignancy prediction using multi-task convolutional neural network. In *Medical Imaging 2017: Computer-Aided Diagnosis* (Vol. 10134, pp. 551-557). SPIE.

[283] Emrani, S., McGuirk, A. and Xiao, W., 2017, August. Prognosis and diagnosis of Parkinson's disease using multi-task learning. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1457-1466).

[284] Li, Y., Yang, T., Zhou, J. and Ye, J., 2018, May. Multi-task learning based survival analysis for predicting alzheimer's disease progression with multi-source block-wise missing data. In *proceedings of the 2018 SIAM international conference on data mining* (pp. 288-296). Society for Industrial and Applied Mathematics.

[285] Tan, K., Huang, W., Liu, X., Hu, J. and Dong, S., 2022. A multi-modal fusion framework based on multi-task correlation learning for cancer prognosis prediction. *Artificial Intelligence in Medicine*, *126*, p.102260.

[286] El-Sappagh, S., Abuhmed, T., Islam, S.R. and Kwak, K.S., 2020. Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. *Neurocomputing*, *412*, pp.197-215.

[287] Chen, L., Saykin, A.J., Yao, B., Zhao, F. and Alzheimer's Disease Neuroimaging Initiative (ADNI, 2022. Multi-task deep autoencoder to predict Alzheimer's disease progression using temporal DNA methylation data in peripheral blood. *Computational and Structural Biotechnology Journal*, *20*, pp.5761-5774.

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

[289] Cao, P., Liu, X., Liu, H., Yang, J., Zhao, D., Huang, M. and Zaiane, O., 2018. Generalized fused group lasso regularized multi-task feature learning for predicting cognitive outcomes in Alzheimers disease. *Computer methods and programs in biomedicine*, *162*, pp.19-45.

[290] Brand, L., Wang, H., Huang, H., Risacher, S., Saykin, A., Shen, L. and ADNI, 2018. Joint high-order multi-task feature learning to predict the progression of alzheimer's disease. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I* (pp. 555-562). Springer International Publishing.

[291] Lu, L., Wang, H., Yao, X., Risacher, S., Saykin, A. and Shen, L., 2018, April. Predicting progressions of cognitive outcomes via high-order multi-modal multi-task feature learning. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 545-548). IEEE.

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

[293] Standley, T., Zamir, A., Chen, D., Guibas, L., Malik, J. and Savarese, S., 2020, November. Which tasks should be learned together in multi-task learning?. In *International Conference on Machine Learning* (pp. 9120-9132). PMLR.

[294] Argyriou, A., Evgeniou, T. and Pontil, M., 2008. Convex multi-task feature learning. *Machine learning*, *73*, pp.243-272.

[295] Stonnington, C.M., Chu, C., Klöppel, S., Jack Jr, C.R., Ashburner, J., Frackowiak, R.S. and Alzheimer Disease Neuroimaging Initiative, 2010. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *Neuroimage*, *51*(4), pp.1405-1413.

[296] Hillar, C.J. and Lim, L.H., 2013. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, *60*(6), pp.1-39.

[297] Shitov, Y., 2016. How hard is the tensor rank?. *arXiv preprint arXiv:1611.01559*.

[298] Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K. and Cilar, L., 2020. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(5), p.e1379.

[299] Vellido, A., 2020. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural computing and applications*, *32*(24), pp.18069-18083.

[300] Fan, F.L., Xiong, J., Li, M. and Wang, G., 2021. On interpretability of artificial neural networks: A survey. *IEEE Transactions on Radiation and Plasma Medical Sciences*, *5*(6), pp.741-760.

[301] ElShawi, R., Sherif, Y., Al-Mallah, M. and Sakr, S., 2021. Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Computational Intelligence*, *37*(4), pp.1633-1650.

[302] Han, L. and Zhang, Y., 2015, February. Learning multi-level task groups in multi-task learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 29, No. 1).

[303] Zhang, Y., Liu, T., Lanfranchi, V. and Yang, P., 2022. 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*, *11*, pp.1-12.

[304] Parikh, N. and Boyd, S., 2014. Proximal algorithms. *Foundations and trends® in Optimization*, *1*(3), pp.127-239.

[305] Xia, M., Wang, J. and He, Y., 2013. BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one*, *8*(7), p.e68910.

[306] Greene, S.J., Killiany, R.J. and Alzheimer's Disease Neuroimaging Initiative, 2010. Subregions of the inferior parietal lobule are affected in the progression to Alzheimer's disease. *Neurobiology of aging*, *31*(8), pp.1304-1311.

[307] Karas, G., Scheltens, P., Rombouts, S., Van Schijndel, R., Klein, M., Jones, B., Van Der Flier, W., Vrenken, H. and Barkhof, F., 2007. Precuneus atrophy in early-onset Alzheimer's disease: a morphometric structural MRI study. *Neuroradiology*, *49*, pp.967-976.

[308] Grignon, Y., Duyckaerts, C., Bennecib, M. and Hauw, J.J., 1998. Cytoarchitectonic alterations in the supramarginal gyrus of late onset Alzheimer's disease. *Acta neuropathologica*, *95*, pp.395-406.

[309] Alegret, M., Cuberas-Borrós, G., Espinosa, A., Valero, S., Hernández, I., Ruíz, A., Becker, J.T., Rosende-Roca, M., Mauleón, A., Sotolongo, O. and Castell-Conesa, J., 2014. Cognitive, genetic, and brain perfusion factors associated with four year incidence of Alzheimer's disease from mild cognitive impairment. *Journal of Alzheimer's Disease*, *41*(3), pp.739-748.

[310] Jacobs, H.I., Van Boxtel, M.P., Jolles, J., Verhey, F.R. and Uylings, H.B., 2012. Parietal cortex matters in Alzheimer's disease: an overview of structural, functional and metabolic findings. *Neuroscience & Biobehavioral Reviews*, *36*(1), pp.297-309.

[311] Parker, T.D., Slattery, C.F., Zhang, J., Nicholas, J.M., Paterson, R.W., Foulkes, A.J., Malone, I.B., Thomas, D.L., Modat, M., Cash, D.M. and Crutch, S.J., 2018. Cortical microstructure in young onset Alzheimer's disease using neurite orientation dispersion and density imaging. *Human brain mapping*, *39*(7), pp.3005-3017.

[312] Brun, A. and Englund, E., 1981. Regional pattern of degeneration in Alzheimer's disease: neuronal loss and histopathological grading. *Histopathology*, *5*(5), pp.549-564.

[313] Jacobs, H.I.L., Van Boxtel, M.P.J., Heinecke, A., Gronenschild, E.H.B.M., Backes, W.H., Ramakers, I.H.G.B., Jolles, J. and Verhey, F.R.J., 2012. Functional integration of parietal lobe activity in early Alzheimer disease. *Neurology*, *78*(5), pp.352-360.

[314] Liang, P., Wang, Z., Yang, Y. and Li, K., 2012. Three subsystems of the inferior parietal cortex are differently affected in mild cognitive impairment. *Journal of Alzheimer's Disease*, *30*(3), pp.475-487.

[315] Yang, H., Xu, H., Li, Q., Jin, Y., Jiang, W., Wang, J., Wu, Y., Li, W., Yang, C., Li, X. and Xiao, S., 2019. Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls. *General psychiatry*, *32*(2).

[316] Li, C., Wang, J., Gui, L., Zheng, J., Liu, C. and Du, H., 2011. Alterations of whole-brain cortical area and thickness in mild cognitive impairment and Alzheimer's disease. *Journal of Alzheimer's Disease*, *27*(2), pp.281-290.

[317] Kim, J.H., Lee, J.W., Kim, G.H., Roh, J.H., Kim, M.J., Seo, S.W., Kim, S.T., Jeon, S., Lee, J.M., Heilman, K.M. and Na, D.L., 2012. Cortical asymmetries in normal, mild cognitive impairment, and Alzheimer's disease. *Neurobiology of aging*, *33*(9), pp.1959-1966.

[318] Pereira, J.B., Mijalkov, M., Kakaei, E., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Spenger, C., Lovestone, S. and Simmons, A., 2016. Disrupted network topology in patients with stable and progressive mild cognitive impairment and Alzheimer's disease. *Cerebral Cortex*, *26*(8), pp.3476-3493.

[319] Zhou, B., Liu, Y., Zhang, Z., An, N., Yao, H., Wang, P., Wang, L., Zhang, X. and Jiang, T., 2013. Impaired functional connectivity of the thalamus in Alzheimer's disease and mild cognitive impairment: a resting-state fMRI study. *Current Alzheimer Research*, *10*(7), pp.754-766.

[320] Hiscox, L.V., Johnson, C.L., McGarry, M.D., Marshall, H., Ritchie, C.W., Van Beek, E.J., Roberts, N. and Starr, J.M., 2020. Mechanical property alterations across the cerebral cortex due to Alzheimer's disease. *Brain Communications*, *2*(1), p.fcz049.

[321] Lim, H.K., Juh, R., Pae, C.U., Lee, B.T., Yoo, S.S., Ryu, S.H., Kwak, K.R., Lee, C. and Lee, C.U., 2008. Altered verbal working memory process in patients with Alzheimer's disease. *Neuropsychobiology*, *57*(4), pp.181-187.

[322] Davies, C.A., Mann, D.M.A., Sumpter, P.Q. and Yates, P.O., 1987. A quantitative morphometric analysis of the neuronal and synaptic content of the frontal and temporal cortex in patients with Alzheimer's disease. *Journal of the neurological sciences*, *78*(2), pp.151-164.

[323] Scheff, S.W. and Price, D.A., 2003. Synaptic pathology in Alzheimer's disease: a review of ultrastructural studies. *Neurobiology of aging*, *24*(8), pp.1029-1046.

[324] Yetkin, F.Z., Rosenberg, R.N., Weiner, M.F., Purdy, P.D. and Cullum, C.M., 2006. FMRI of working memory in patients with mild cognitive impairment and probable Alzheimer's disease. *European radiology*, *16*, pp.193-206.

[325] Convit, A., De Asis, J., De Leon, M.J., Tarshish, C.Y., De Santi, S. and Rusinek, H., 2000. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease☆. *Neurobiology of aging*, *21*(1), pp.19-26.

[326] Peters, F., Collette, F., Degueldre, C., Sterpenich, V., Majerus, S. and Salmon, E., 2009. The neural correlates of verbal short-term memory in Alzheimer's disease: an fMRI study. *Brain*, *132*(7), pp.1833-1846.

[327] Shefer, V.F., 1973. Absolute number of neurons and thickness of the cerebral cortex during aging, senile and vascular dementia, and Pick's and Alzheimer's diseases. *Neuroscience and behavioral physiology*, *6*(4), pp.319-324.

[328] Seidenberg, M., Guidotti, L., Nielson, K.A., Woodard, J.L., Durgerian, S., Antuono, P., Zhang, Q. and Rao, S.M., 2009. Semantic memory activation in individuals at risk for developing Alzheimer disease. *Neurology*, *73*(8), pp.612-620.

[329] Weston, P.S., Nicholas, J.M., Lehmann, M., Ryan, N.S., Liang, Y., Macpherson, K., Modat, M., Rossor, M.N., Schott, J.M., Ourselin, S. and Fox, N.C., 2016. Presymptomatic cortical thinning in familial Alzheimer disease: A longitudinal MRI study. *Neurology*, *87*(19), pp.2050-2057.

[330] Lehmann, M., Rohrer, J.D., Clarkson, M.J., Ridgway, G.R., Scahill, R.I., Modat, M., Warren, J.D., Ourselin, S., Barnes, J., Rossor, M.N. and Fox, N.C., 2010. Reduced cortical thickness in the

posterior cingulate gyrus is characteristic of both typical and atypical Alzheimer's disease. *Journal of Alzheimer's Disease*, *20*(2), pp.587-598.

[331] Verfaillie, S.C., Tijms, B., Versteeg, A., Benedictus, M.R., Bouwman, F.H., Scheltens, P., Barkhof, F., Vrenken, H. and van der Flier, W.M., 2016. Thinner temporal and parietal cortex is related to incident clinical progression to dementia in patients with subjective cognitive decline. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *5*, pp.43-52.

[332] Damoiseaux, J.S., Prater, K.E., Miller, B.L. and Greicius, M.D., 2012. Functional connectivity tracks clinical deterioration in Alzheimer's disease. *Neurobiology of aging*, *33*(4), pp.828-e19.

[333] Golby, A., Silverberg, G., Race, E., Gabrieli, S., O'Shea, J., Knierim, K., Stebbins, G. and Gabrieli, J., 2005. Memory encoding in Alzheimer's disease: an fMRI study of explicit and implicit memory. *Brain*, *128*(4), pp.773-787.

[334] Prvulovic, D., Hubl, D., Sack, A.T., Melillo, L., Maurer, K., Frölich, L., Lanfermann, H., Zanella, F.E., Goebel, R., Linden, D.E. and Dierks, T., 2002. Functional imaging of visuospatial processing in Alzheimer's disease. *Neuroimage*, *17*(3), pp.1403-1414.

[335] Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp.1189-1232.

[336] Ogutu, J.O., Piepho, H.P. and Schulz-Streeck, T., 2011, December. A comparison of random forests, boosting and support vector machines for genomic selection. In *BMC proceedings* (Vol. 5, No. 3, pp. 1-5). BioMed Central.

[337] Zhang, Y., Wang, X., Liu, T., Wang, R., Li, Y., Xue, Q. and Yang, P., 2023. Sustainable fertilisation management via tensor multi-task learning using multi-dimensional agricultural data. *Journal of Industrial Information Integration*, *34*, p.100461.

[338] Zhang, Y., Liu, T., Li, Y., Wang, R., Huang, H. and Yang, P., 2022, October. Spatio-temporal Tensor Multi-Task Learning for Precision Fertilisation with Real-world Agricultural Data. In *IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society* (pp. 1-6). IEEE.

[339] Zhang, Y., Liu, T., Li, Y., Wang, R., Huang, H. and Yang, P., 2021, December. Spatio-temporal Tensor Multi-Task Learning for Precision fertilization. In *2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS)* (pp. 398-405). IEEE.

Appendix A: Mainstream similarity calculation approaches

A.1 Cosine similarity

Cosine similarity is a major similarity measurement technique that expresses the degree of similarity by calculating the cosine of the angle between two vectors, with values ranging from -1 to 1. In practice, cosine similarity is widely utilised in areas such as text classification, information retrieval and recommendation systems. It has the advantage of being simple and easy to utilise, and the result is an intuitive representation of the similarity of the two vectors. Cosine similarity utilises the cosine value of the angle between two vectors in a vector space as a measurement for the difference between two individuals. When the two vectors are more similar, the angle is closer to 0 degrees, and the cosine value is closer to 1. Cosine similarity between two vectors $\mathrm{A}(a_1, a_2)$ and $\mathrm{B}(b_1, b_2)$ is defined as $\cos\theta = \frac{a_1 b_1 + a_2 b_2}{\sqrt{a_1^2 + a_2^2} \times \sqrt{b_1^2 + b_2^2}}$.

Compared with Euclidean distance, cosine similarity focuses more on the difference in direction of two vectors (Figure A.1). If the position of A keeps unvaried and B away from the origin of the coordinate axis in the original direction, then the cosine similarity cos θ keeps unvaried, because the angle remains unchanged, and the distance between A and B changed, which is the difference between Cosine similarity and Euclidean distance.

To sum up, cosine similarity has the following advantages: 1) Simplicity: cosine similarity is a simple and easy to understand approach to similarity calculation, which only requires the calculation of the product of the dot product of two vectors and their modulus lengths. 2) Computational stability: the results of the cosine similarity

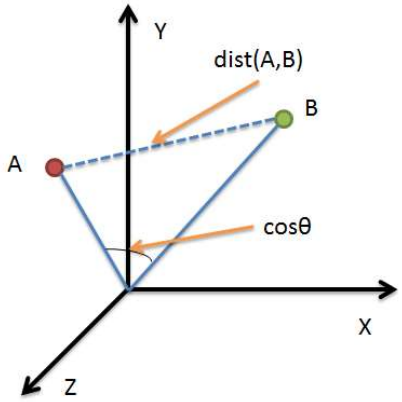


Figure A.1: The difference between Cosine similarity and Euclidean distance.

calculation is highly stable and are not affected by small errors. 3) Vector space model applicability: cosine similarity can be directly applied to vector space models, thus enabling similarity analysis of data such as text and images. However, cosine similarity has the following disadvantages: 1) Cosine similarity only considers the direction of a vector, ignoring its size, therefore in certain cases where the size of two similar vectors differs significantly, the cosine similarity result is smaller. 2) For sparse data, the result of cosine similarity cannot guarantee the accuracy because numerous eigenvalues in sparse data are zero, thus the calculated cosine similarity often cannot accurately reflect the similarity of two vectors. 3) For high-dimensional dense data, it can produce highly similar results. This is because in dense data, numerous dimensions have high values and their inner products are large, thus it is easy to produce high similarity.

A.2 Euclidean distance

Euclidean distance is the easiest to understand distance calculation method. It is derived from distance formula between two points in Euclidean space, which measures the absolute distance between points in a multi-dimensional space. It is often utilised in data analysis and mining, such as clustering or calculating similarity. Euclidean distance between two points a $(x_1,\ y_1)$ and b $(x_2,\ y_2)$ for a two-dimensional plane: $\sqrt{(x_1 - x_2)^2\ + (y_1 - y_2)^2}$.

Euclidean distance is utilised in a wide range of applications, for example in machine learning, k-nearest neighbour algorithms, cluster analysis etc. In data mining, it is frequently utilised as an evaluation metric for clustering algorithms. In addition, it is widely used in image processing, speech recognition, bioinformatics and other fields.

To sum up, the Euclidean distance has the following advantages: 1) Simplicity and ease of utilise: the Euclidean distance is sufficiently simple to be calculated directly using standard mathematical formulas. 2) Clear geometric meaning: the geometric meaning of Euclidean distance is the true distance between two points; thus, it has wide applicability in numerous spatial geometry problems. 3) Strong mathematical support: the Euclidean distance has a solid mathematical backing and can therefore be used in a wide range of mathematical methods. 4) Good stability: the Euclidean distance has a high degree of stability and cannot be significantly affected by small changes in the data. 5) Utilisation for multi-dimensional data: Euclidean distances can be used for multi-dimensional data and distances can be calculated directly on multi-dimensional data.

However, the Euclidean distance has the following disadvantages: 1) Does not consider the direction of the vectors: the Euclidean distance only considers the length of the vectors, not their direction. Therefore, in certain cases, it does not accurately reflect the similarity between data. 2) Cannot handle non-linear relationships: Euclidean distance is based on linear relationships, which can lead to inaccurate results if there are non-linear relationships between the data. 3) Numerical discontinuity: Euclidean distance is a numerically discontinuous distance, which means that even small variations in distance can lead to large variations in the output value. 4) High-dimensional catastrophe: Euclidean distances become complex as dimensionality increases, thus when dimensionality is high, Euclidean distances can suffer from a "high-dimensional catastrophe", meaning that the distances between points can no longer be valid in a

high-dimensional space. 5) Does not deal with densities: Euclidean distances do not consider the density of data points, which implies that if two points are close but the density of surrounding data points is low, Euclidean distance will still recognise them as similar, but it may not be realistic. 6) Sensitive to outliers: Euclidean distances are sensitive to outliers, and if there are outliers in the data, the results of the Euclidean distance can be significantly affected.

A.3 Mahalanobis distance

Mahalanobis distance is a distance measurement with weights that is utilised to determine the relative distance of a sample from the mean of multi-dimensional data. It is more complex to calculate than Euclidean distance, but it can consider the correlation between multi-dimensional data. Mahalanobis distance is widely utilised in various fields, such as image recognition, data classification, data clustering, etc. In the field of image recognition, the Mahalanobis distance can be utilised to calculate the similarity of images and to identify objects by comparing the similarity of images. In the field of data classification, the Mahalanobis distance can be used to determine whether a sample is anomalous or not. In the field of data clustering, the Mahalanobis distance can be used to assess the degree of similarity between samples and to group data. It can be used to perform statistical analysis, such as dimensionality reduction of data, removal of redundant data, and to assess the degree of outliers of a sample. In the field of data mining, the Mahalanobis distance can be used to identify correlations in a data set. In the field of biology, the Mahalanobis distance can be used to identify differences between different biological samples and can be used to assess the similarity of biological samples and it can alternatively be used to assess the similarity of biological data, such as genomic data and protein structure data. Mahalanobis distance between the vectors X_i and X_j is defined as: $\mathrm{D}(\mathrm{X}_i, \mathrm{X}_j) = \sqrt{(\mathrm{X}_i - \mathrm{X}_j)^{\mathrm{T}} \mathrm{S}^{-1} (\mathrm{X}_i - \mathrm{X}_j)}$, where S is the covariance matrix. Divided by the covariance matrix, which removes the variance

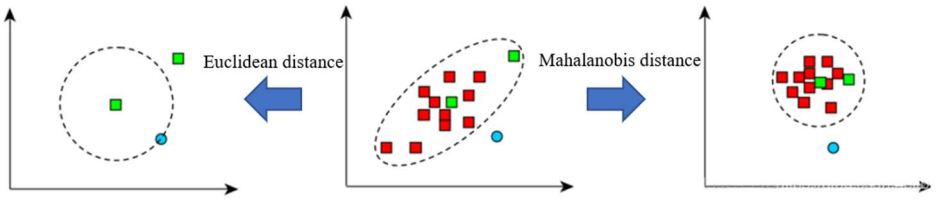


Figure A.2: The difference between Mahalanobis distance and Euclidean distance.

between each component, eliminating the dimensionality.

The Mahalanobis distance is a valid approach to calculate the similarity between two sample sets. The difference from Euclidean distance is that it considers the relationship between different features and it is scale-invariant, that is, independent of the measurement scale (Figure A.2). The advantage of the Mahalanobis distance is that it is not affected by the scale, which means Mahalanobis distance between two points is irrelevant with the unit of measurement of the original data. The disadvantage is that the effect of the small change can be exaggerated.

To sum up, the Mahalanobis distance has the following advantages: 1) Intuition: the Mahalanobis distance compares the distance between two points in a multi-dimensional space and is easy to understand and intuitive. 2) Stability: Mahalanobis distance is a stable metric and can therefore be used to assess similarity in multi-dimensional data analysis. 3) Not affected by the scale: the Mahalanobis distance between two vectors is irrelevant with the unit of measurement of the original data. 4) Compatibility: the Mahalanobis distance can be applied to data of any number of dimensions, making it extremely suitable for multi-dimensional data analysis. 5) Ability to handle outliers: Mahalanobis distance can handle outliers in the data, which is an essential advantage, as numerous data analysis approaches cannot handle outliers effectively. 6) Scalability: this means that it can be applied to data sets of different sizes and distances can be calculated for multi-dimensional features. 7) Correlation of multi-dimensional features between samples can be considered and the results are not affected by significant deviations in one dimensional feature. 8) Effective data analysis can be performed on

high-dimensional data and is not susceptible to dimensional catastrophes. However, there are a number of disadvantages: 1) Does not consider the direction of the vectors: the Mahalanobis distance only considers the length of the vectors, not their direction. Therefore, in certain cases, it does not accurately reflect the similarity between data. 2) The computation of the distance requires the utilisation of a covariance matrix, thus the complexity of the computation can be extremely high when the dimensionality of the data is high, which affects efficiency of the algorithm. 3) Mahalanobis distance cannot handle a mixture of dense and sparse data effectively. 4) Mahalanobis distance cannot handle missing data effectively. If there are missing values in the data, the calculation of the Marcian distance can be inaccurate. 5) Mahalanobis distances cannot handle the different importance of different dimensions of data effectively. 6) The Mahalanobis distance is a method for calculating the distance between two samples, but cannot handle distances between multiple samples.

Appendix B: MRI biomarker features utilised in this research

Vol(C). refers to the volumes of cortical parcellations

Surf. Area refers to the total surface area of the cortex

Vol(WM). refers to the volumes of specific white matter parcellations

CTA. refers to the average cortical thickness

CTStd. refers to the standard deviation in cortical thickness

L for left, R for right

Labels	MRI biomarker features
1	Vol(WM). of R.Pallidum
2	Vol(C). of R.Paracentral
3	Surf. Area of R.Paracentral
4	CTA. of R.Paracentral
5	CTStd. of R.Paracentral
6	Vol(C). of R.Parahippocampal
7	Surf. Area of R.Parahippocampal
8	CTA. of R.Parahippocampal
9	CTStd. of R.Parahippocampal
10	Vol(C). of R.ParsOpercularis
11	Surf. Area of R.ParsOpercularis
12	CTA. of R.ParsOpercularis
13	CTStd. of R.ParsOpercularis
14	Vol(C). of R.ParsOrbitalis
15	Surf. Area of R.ParsOrbitalis
16	CTA. of R.ParsOrbitalis
17	CTStd. of R.ParsOrbitalis
18	Vol(C). of R.ParsTriangularis
19	Surf. Area of R.ParsTriangularis
20	CTA. of R.ParsTriangularis
21	CTStd. of R.ParsTriangularis
22	Vol(C). of R.Pericalcarine
23	Surf. Area of R.Pericalcarine
24	CTA. of R.Pericalcarine
25	CTStd. of R.Pericalcarine
26	Vol(C). of R.Postcentral
27	Surf. Area of R.Postcentral
28	CTA. of R.Postcentral
29	CTStd. of R.Postcentral
30	Vol(C). of R.PosteriorCingulate
31	Surf. Area of R.PosteriorCingulate

32	CTA. of R.PosteriorCingulate
33	CTStd. of R.PosteriorCingulate
34	Vol(C). of Icv
35	Vol(C). of R.Precentral
36	Surf. Area of R.Precentral
37	CTA. of R.Precentral
38	CTStd. of R.Precentral
39	Vol(C). of R.Precuneus
40	Surf. Area of R.Precuneus
41	CTA. of R.Precuneus
42	CTStd. of R.Precuneus
43	Vol(WM). of R.Putamen
44	Vol(C). of R.RostralAnteriorCingulate
45	Surf. Area of R.RostralAnteriorCingulate
46	CTA. of R.RostralAnteriorCingulate
47	CTStd. of R.RostralAnteriorCingulate
48	Vol(C). of R.RostralMiddleFrontal
49	Surf. Area of R.RostralMiddleFrontal
50	CTA. of R.RostralMiddleFrontal
51	CTStd. of R.RostralMiddleFrontal
52	Vol(C). of R.SuperiorFrontal
53	Surf. Area of R.SuperiorFrontal
54	CTA. of R.SuperiorFrontal
55	CTStd. of R.SuperiorFrontal
56	Vol(C). of R.SuperiorParietal
57	Surf. Area of R.SuperiorParietal
58	CTA. of R.SuperiorParietal
59	CTStd. of R.SuperiorParietal
60	Vol(C). of R.SuperiorTemporal
61	Surf. Area of R.SuperiorTemporal
62	CTA. of R.SuperiorTemporal
63	CTStd. of R.SuperiorTemporal
64	Vol(C). of R.Supramarginal
65	Surf. Area of R.Supramarginal
66	CTA. of R.Supramarginal
67	CTStd. of R.Supramarginal
68	Vol(C). of R.TemporalPole
69	Surf. Area of R.TemporalPole
70	CTA. of R.TemporalPole
71	CTStd. of R.TemporalPole
72	Vol(WM). of L.AccumbensArea
73	Vol(WM). of R.Thalamus
74	Vol(C). of R.TransverseTemporal

75	Surf. Area of R.TransverseTemporal
76	CTA. of R.TransverseTemporal
77	CTStd. of R.TransverseTemporal
78	Vol(WM). of R.VentralDC
79	Vol(WM). of R.Vessel
80	Vol(WM). of ThirdVentricle
81	Vol(WM). of WMHypoIntensities
82	Vol(C). of L.Insula
83	Surf. Area of L.Insula
84	CTA. of L.Insula
85	CTStd. of L.Insula
86	Vol(WM). of L.Amygdala
87	Vol(C). of R.Insula
88	Surf. Area of R.Insula
89	CTA. of R.Insula
90	CTStd. of R.Insula
91	Vol(C). of L.Bankssts
92	Surf. Area of L.Bankssts
93	CTA. of L.Bankssts
94	CTStd. of L.Bankssts
95	Vol(C). of L.CaudalAnteriorCingulate
96	Surf. Area of L.CaudalAnteriorCingulate
97	CTA. of L.CaudalAnteriorCingulate
98	CTStd. of L.CaudalAnteriorCingulate
99	Vol(C). of L.CaudalMiddleFrontal
100	Surf. Area of L.CaudalMiddleFrontal
101	CTA. of L.CaudalMiddleFrontal
102	CTStd. of L.CaudalMiddleFrontal
103	Vol(WM). of L.Caudate
104	Vol(WM). of L.CerebellumCortex
105	Vol(WM). of L.CerebellumWM
106	Vol(WM). of Brainstem
107	Vol(WM). of L.ChoroidPlexus
108	Vol(C). of L.Cuneus
109	Surf. Area of L.Cuneus
110	CTA. of L.Cuneus
111	CTStd. of L.Cuneus
112	Vol(C). of L.Entorhinal
113	Surf. Area of L.Entorhinal
114	CTA. of L.Entorhinal
115	CTStd. of L.Entorhinal
116	Vol(C). of L.FrontalPole
117	Surf. Area of L.FrontalPole

118	CTA. of L.FrontalPole
119	CTStd. of L.FrontalPole
120	Vol(C). of L.Fusiform
121	Surf. Area of L.Fusiform
122	CTA. of L.Fusiform
123	CTStd. of L.Fusiform
124	Vol(WM). of L.Hippocampus
125	Vol(WM). of CorpusCallosumAnterior
126	Vol(WM). of L.InferiorLateralVentricle
127	Vol(C). of L.InferiorParietal
128	Surf. Area of L.InferiorParietal
129	CTA. of L.InferiorParietal
130	CTStd. of L.InferiorParietal
131	Vol(C). of L.InferiorTemporal
132	Surf. Area of L.InferiorTemporal
133	CTA. of L.InferiorTemporal
134	CTStd. of L.InferiorTemporal
135	Vol(C). of L.IsthmusCingulate
136	Surf. Area of L.IsthmusCingulate
137	CTA. of L.IsthmusCingulate
138	CTStd. of L.IsthmusCingulate
139	Vol(C). of L.LateralOccipital
140	Surf. Area of L.LateralOccipital
141	CTA. of L.LateralOccipital
142	CTStd. of L.LateralOccipital
143	Vol(C). of L.LateralOrbitofrontal
144	Surf. Area of L.LateralOrbitofrontal
145	CTA. of L.LateralOrbitofrontal
146	CTStd. of L.LateralOrbitofrontal
147	Vol(WM). of L.LateralVentricle
148	Vol(C). of L.Lingual
149	Surf. Area of L.Lingual
150	CTA. of L.Lingual
151	CTStd. of L.Lingual
152	Vol(C). of L.MedialOrbitofrontal
153	Surf. Area of L.MedialOrbitofrontal
154	CTA. of L.MedialOrbitofrontal
155	CTStd. of L.MedialOrbitofrontal
156	Vol(WM). of CorpusCallosumCentral
157	Vol(C). of L.MiddleTemporal
158	Surf. Area of L.MiddleTemporal
159	CTA. of L.MiddleTemporal
160	CTStd. of L.MiddleTemporal

161	Vol(WM). of L.Pallidum
162	Vol(C). of L.Paracentral
163	Surf. Area of L.Paracentral
164	CTA. of L.Paracentral
165	CTStd. of L.Paracentral
166	Vol(C). of L.Parahippocampal
167	Surf. Area of L.Parahippocampal
168	CTA. of L.Parahippocampal
169	CTStd. of L.Parahippocampal
170	Vol(C). of L.ParsOpercularis
171	Surf. Area of L.ParsOpercularis
172	CTA. of L.ParsOpercularis
173	CTStd. of L.ParsOpercularis
174	Vol(C). of L.ParsOrbitalis
175	Surf. Area of L.ParsOrbitalis
176	CTA. of L.ParsOrbitalis
177	CTStd. of L.ParsOrbitalis
178	Vol(C). of L.ParsTriangularis
179	Surf. Area of L.ParsTriangularis
180	CTA. of L.ParsTriangularis
181	CTStd. of L.ParsTriangularis
182	Vol(C). of L.Pericalcarine
183	Surf. Area of L.Pericalcarine
184	CTA. of L.Pericalcarine
185	CTStd. of L.Pericalcarine
186	Vol(C). of L.Postcentral
187	Surf. Area of L.Postcentral
188	CTA. of L.Postcentral
189	CTStd. of L.Postcentral
190	Vol(WM). of CorpusCallosumMidAnterior
191	Vol(C). of L.PosteriorCingulate
192	Surf. Area of L.PosteriorCingulate
193	CTA. of L.PosteriorCingulate
194	CTStd. of L.PosteriorCingulate
195	Vol(C). of L.Precentral
196	Surf. Area of L.Precentral
197	CTA. of L.Precentral
198	CTStd. of L.Precentral
199	Vol(C). of L.Precuneus
200	Surf. Area of L.Precuneus
201	CTA. of L.Precuneus
202	CTStd. of L.Precuneus
203	Vol(WM). of L.Putamen

204	Vol(C). of L.RostralAnteriorCingulate
205	Surf. Area of L.RostralAnteriorCingulate
206	CTA. of L.RostralAnteriorCingulate
207	CTStd. of L.RostralAnteriorCingulate
208	Vol(C). of L.RostralMiddleFrontal
209	Surf. Area of L.RostralMiddleFrontal
210	CTA. of L.RostralMiddleFrontal
211	CTStd. of L.RostralMiddleFrontal
212	Vol(C). of L.SuperiorFrontal
213	Surf. Area of L.SuperiorFrontal
214	CTA. of L.SuperiorFrontal
215	CTStd. of L.SuperiorFrontal
216	Vol(C). of L.SuperiorParietal
217	Surf. Area of L.SuperiorParietal
218	CTA. of L.SuperiorParietal
219	CTStd. of L.SuperiorParietal
220	Vol(C). of L.SuperiorTemporal
221	Surf. Area of L.SuperiorTemporal
222	CTA. of L.SuperiorTemporal
223	CTStd. of L.SuperiorTemporal
224	Vol(C). of L.Supramarginal
225	Surf. Area of L.Supramarginal
226	CTA. of L.Supramarginal
227	CTStd. of L.Supramarginal
228	Vol(WM). of CorpusCallosumMidPosterior
229	Vol(C). of L.TemporalPole
230	Surf. Area of L.TemporalPole
231	CTA. of L.TemporalPole
232	CTStd. of L.TemporalPole
233	Vol(WM). of L.Thalamus
234	Vol(C). of L.TransverseTemporal
235	Surf. Area of L.TransverseTemporal
236	CTA. of L.TransverseTemporal
237	CTStd. of L.TransverseTemporal
238	Vol(WM). of L.VentralDC
239	Vol(WM). of L.Vessel
240	Vol(WM). of NonWMHypoIntensities
241	Vol(WM). of OpticChiasm
242	Vol(WM). of CorpusCallosumPosterior
243	Vol(WM). of R.AccumbensArea
244	Vol(WM). of R.Amygdala
245	Vol(C). of R.Bankssts
246	Surf. Area of R.Bankssts

247	CTA. of R.Bankssts
248	CTStd. of R.Bankssts
249	Vol(C). of R.CaudalAnteriorCingulate
250	Surf. Area of R.CaudalAnteriorCingulate
251	CTA. of R.CaudalAnteriorCingulate
252	CTStd. of R.CaudalAnteriorCingulate
253	Vol(C). of R.CaudalMiddleFrontal
254	Surf. Area of R.CaudalMiddleFrontal
255	CTA. of R.CaudalMiddleFrontal
256	CTStd. of R.CaudalMiddleFrontal
257	Vol(WM). of R.Caudate
258	Vol(WM). of R.CerebellumCortex
259	Vol(WM). of R.CerebellumWM
260	Vol(WM). of Csf
261	Vol(WM). of R.ChoroidPlexus
262	Vol(C). of R.Cuneus
263	Surf. Area of R.Cuneus
264	CTA. of R.Cuneus
265	CTStd. of R.Cuneus
266	Vol(C). of R.Entorhinal
267	Surf. Area of R.Entorhinal
268	CTA. of R.Entorhinal
269	CTStd. of R.Entorhinal
270	Vol(C). of R.FrontalPole
271	Surf. Area of R.FrontalPole
272	CTA. of R.FrontalPole
273	CTStd. of R.FrontalPole
274	Vol(C). of R.Fusiform
275	Surf. Area of R.Fusiform
276	CTA. of R.Fusiform
277	CTStd. of R.Fusiform
278	Vol(WM). of R.Hippocampus
279	Vol(WM). of R.InferiorLateralVentricle
280	Vol(C). of R.InferiorParietal
281	Surf. Area of R.InferiorParietal
282	CTA. of R.InferiorParietal
283	CTStd. of R.InferiorParietal
284	Vol(C). of R.InferiorTemporal
285	Surf. Area of R.InferiorTemporal
286	CTA. of R.InferiorTemporal
287	CTStd. of R.InferiorTemporal
288	Vol(C). of R.IsthmusCingulate
289	Surf. Area of R.IsthmusCingulate

290	CTA. of R.IsthmusCingulate
291	CTStd. of R.IsthmusCingulate
292	Vol(C). of R.LateralOccipital
293	Surf. Area of R.LateralOccipital
294	CTA. of R.LateralOccipital
295	CTStd. of R.LateralOccipital
296	Vol(C). of R.LateralOrbitofrontal
297	Surf. Area of R.LateralOrbitofrontal
298	CTA. of R.LateralOrbitofrontal
299	CTStd. of R.LateralOrbitofrontal
300	Vol(WM). of R.LateralVentricle
301	Vol(C). of R.Lingual
302	Surf. Area of R.Lingual
303	CTA. of R.Lingual
304	CTStd. of R.Lingual
305	Vol(C). of R.MedialOrbitofrontal
306	Surf. Area of R.MedialOrbitofrontal
307	CTA. of R.MedialOrbitofrontal
308	CTStd. of R.MedialOrbitofrontal
309	Vol(C). of R.MiddleTemporal
310	Surf. Area of R.MiddleTemporal
311	CTA. of R.MiddleTemporal
312	CTStd. of R.MiddleTemporal
313	Vol(WM). of FourthVentricle

Appendix C: Abbreviations for regions in BrainNet viewer visualisation

Abbreviations for regions in BrainNet viewer visualisation (L for left, R for right).

Labels	Regions	abbr.
1	Precentral gyrus	PreCG
2	Superior parietal gyrus	SPG
3	Superior frontal gyrus	SFGdor
4	Postcentral gyrus	PoCG
5	Inferior parietal lobule	IPL
6	Supramarginal gyrus	SMG
7	Caudal middle frontal gyrus	CMF
8	Precuneus	PCUN
9	Pars opercularis	POPE
10	Rostral middle frontal gyrus	RMF
11	Superior temporal gyrus	STG
12	Middle temporal gyrus	MTG
13	Lateral occipital sulcus	LOCC
14	Cerebellum cortex	CERC
15	Inferior temporal gyrus	ITG
16	Paracentral lobule	PCL
17	Fusiform gyrus	FFG
18	Posterior cingulate gyrus	PCG
19	Bankssts	BAN
20	Pericalcarine cortex	CAL
21	Hippocampus	HIP
22	Pars triangularis	PTRI
23	Corpus callosum mid anterior	CCMA
24	Lingual gyrus	LING
25	Medial orbitofrontal cortex	MORB
26	Cerebellum white matter	CERW
27	Caudal anterior cingulate cortex	CAC
28	Isthmus of cingulate gyrus	ISC
29	Putamen	PUT
30	Parahippocampal gyrus	PHG
31	Pars orbitalis	PORB
32	Amygdala	AMYG
33	Insula	INS
34	Lateral orbitofrontal cortex	LORB

35	Caudate nucleus	CAU
36	Transverse temporal gyrus	TTEM
37	Cuneus	CUN
38	Frontal pole	FPO
39	Thalamus	THA
40	Rostral anterior cingulate cortex	RAC
41	Temporal pole	TPOsup
42	Corpus callosum mid posterior	CCMP
43	Entorhinal cortex	ENT
44	Corpus callosum central	CCCe

Appendix D: The top 10 brain biomarker structural variation correlations of the proposed AMDQ-TMTL approach

The top 10 brain biomarker structural variation correlations of the proposed AMDQ-TMTL approach are shown in descending order of the weighted parameter values by MMSE prediction at various time points in Tables D.1-D.10 and visualised in Figure D.1, by ADAS-Cog prediction at various time points in Tables D.11-D.20 and visualised in Figure D.2, by RAVLT TOTAL prediction at various time points in Tables D.21-D.30 and visualised in Figure D.3, by FLU ANIM prediction at various time points in Tables D.31-D.40 and visualised in Figure D.4.

D.1 For MMSE prediction targets

Table D.1: The top-10 rank brain biomarker correlations in time point M12 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.5301
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.5244
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.5145
CTA. of R.Precuneus - Vol(C). of R.Supramarginal	0.4916
Vol(C). of R.Precentral - CTA. of R.Supramarginal	0.4889
CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal	0.4679
CTA. of R.SuperiorTemporal - CTA. of L.SuperiorFrontal	0.4582
CTA. of L.InferiorParietal - CTA. of L.Precentral	0.4479
Vol(C). of L.SuperiorFrontal - Vol(C). of L.SuperiorParietal	0.4289
Vol(C). of R.Supramarginal - Vol(C). of L.Precuneus	0.4261

Table D.2: The top-10 rank brain biomarker correlations in time point M24 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.6296
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.6097
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.5821
CTA. of R.Precuneus - Vol(C). of R.Supramarginal	0.5656
Vol(C). of R.Precentral - CTA. of R.Supramarginal	0.5526
CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal	0.5384
CTA. of L.InferiorParietal - CTA. of L.Precentral	0.5352
CTA. of R.SuperiorTemporal - CTA. of L.SuperiorFrontal	0.5201
CTA. of L.Precuneus - CTA. of R.InferiorParietal	0.4933
Vol(C). of L.Precentral - CTA. of L.Supramarginal	0.4815

Table D.3: The top-10 rank brain biomarker correlations in time point M36 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.6898
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.6364
CTA. of R.Precuneus - Vol(C). of R.Supramarginal	0.6172
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.5977
CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal	0.5869
CTA. of L.InferiorParietal - CTA. of L.Precentral	0.5759
Vol(C). of R.Precentral - CTA. of R.Supramarginal	0.5558
CTA. of L.Precuneus - CTA. of R.InferiorParietal	0.5304
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.5239
CTA. of R.ParsOpercularis - CTA. of R.Supramarginal	0.5232

Table D.4: The top-10 rank brain biomarker correlations in time point M48 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.6978
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.6616
CTA. of R.Precuneus - Vol(C). of R.Supramarginal	0.6566
CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal	0.6259
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.6177
CTA. of L.InferiorParietal - CTA. of L.Precentral	0.6067
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.5979
Vol(C). of R.Precentral - CTA. of R.Supramarginal	0.5742
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.5534
Vol(C). of L.SuperiorFrontal - Vol(C). of L.SuperiorParietal	0.5532

Table D.5: The top-10 rank brain biomarker correlations in time point M60 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.9068
CTA. of R.Precuneus - Vol(C). of R.Supramarginal	0.7687
CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal	0.7439
CTA. of L.InferiorParietal - CTA. of L.Precentral	0.7194
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.6874
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.6762
Vol(C). of L.SuperiorFrontal - Vol(C). of L.SuperiorParietal	0.6706
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.6698
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.6673
CTA. of L.Precuneus - CTA. of R.InferiorParietal	0.6534

Table D.6: The top-10 rank brain biomarker correlations in time point M72 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.7843
CTA. of L.InferiorParietal - CTA. of L.Precentral	0.6337
CTA. of L.Precuneus - CTA. of R.InferiorParietal	0.6102
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.6096
CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal	0.6015
CTA. of R.ParsTriangularis - CTA. of L.CaudalMiddleFrontal	0.5928
CTA. of R.ParsOpercularis - CTA. of R.Supramarginal	0.5664
Vol(C). of L.SuperiorFrontal - Vol(C). of L.SuperiorParietal	0.5607
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.5581
CTA. of R.SuperiorParietal - CTA. of R.InferiorParietal	0.5518

Table D.7: The top-10 rank brain biomarker correlations in time point M84 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.7096
Vol(C). of L.SuperiorFrontal - Vol(C). of L.SuperiorParietal	0.6812
CTA. of L.InferiorParietal - CTA. of L.Precentral	0.6645
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.6549
CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal	0.6445
CTA. of R.Precuneus - Vol(C). of R.Supramarginal	0.6176
Vol(C). of L.PosteriorCingulate - Vol(C). of L.Precuneus	0.5986
Vol(C). of R.SuperiorParietal - Vol(C). of L.SuperiorParietal	0.5978
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.5938
Vol(C). of L.Postcentral - Vol(C). of L.SuperiorParietal	0.5909

Table D.8: The top-10 rank brain biomarker correlations in time point M96 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.5756
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.5453
Vol(C). of L.SuperiorFrontal - Vol(C). of L.SuperiorParietal	0.5378
CTA. of R.Postcentral - Vol(C). of R.Fusiform	0.5258
CTA. of R.ParsTriangularis - CTA. of L.CaudalMiddleFrontal	0.5219
CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal	0.5138
CTA. of R.Precuneus - Vol(C). of R.Supramarginal	0.5118
CTA. of L.InferiorParietal - CTA. of L.Precentral	0.5015
Vol(C). of R.SuperiorParietal - Vol(C). of L.SuperiorParietal	0.4972
CTA. of R.ParsOpercularis - Vol(C). of R.RostralMiddleFrontal	0.4956

Table D.9: The top-10 rank brain biomarker correlations in time point M108 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.6538
Vol(C). of L.PosteriorCingulate - Vol(C). of L.Precuneus	0.6268
CTA. of L.InferiorParietal - CTA. of L.Precentral	0.6190
CTA. of R.Precuneus - Vol(C). of R.Supramarginal	0.6093
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.5947
CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal	0.5853
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.5839
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.5609
Vol(C). of L.SuperiorFrontal - Vol(C). of L.SuperiorParietal	0.5520
CTA. of R.ParsTriangularis - CTA. of L.CaudalMiddleFrontal	0.5478

Table D.10: The top-10 rank brain biomarker correlations in time point M120 for the AMDQ-TMTL approach on MMSE prediction.

Brain biomarker correlation	Weight
CTA. of R.Precuneus - Vol(C). of R.Supramarginal	0.8201
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.8014
Vol(C). of L.PosteriorCingulate - Vol(C). of L.Precuneus	0.7853
CTA. of R.Postcentral - Vol(C). of R.Fusiform	0.7421
CTA. of R.ParsTriangularis - CTA. of L.CaudalMiddleFrontal	0.7420
CTA. of R.ParsOpercularis - CTA. of R.TransverseTemporal	0.7366
CTA. of R.ParsOpercularis - CTA. of R.Supramarginal	0.7052
Vol(C). of R.ParsOpercularis - Vol(C). of L.SuperiorTemporal	0.6423
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.6374
CTA. of L.RostralMiddleFrontal - CTA. of R.IsthmusCingulate	0.6189

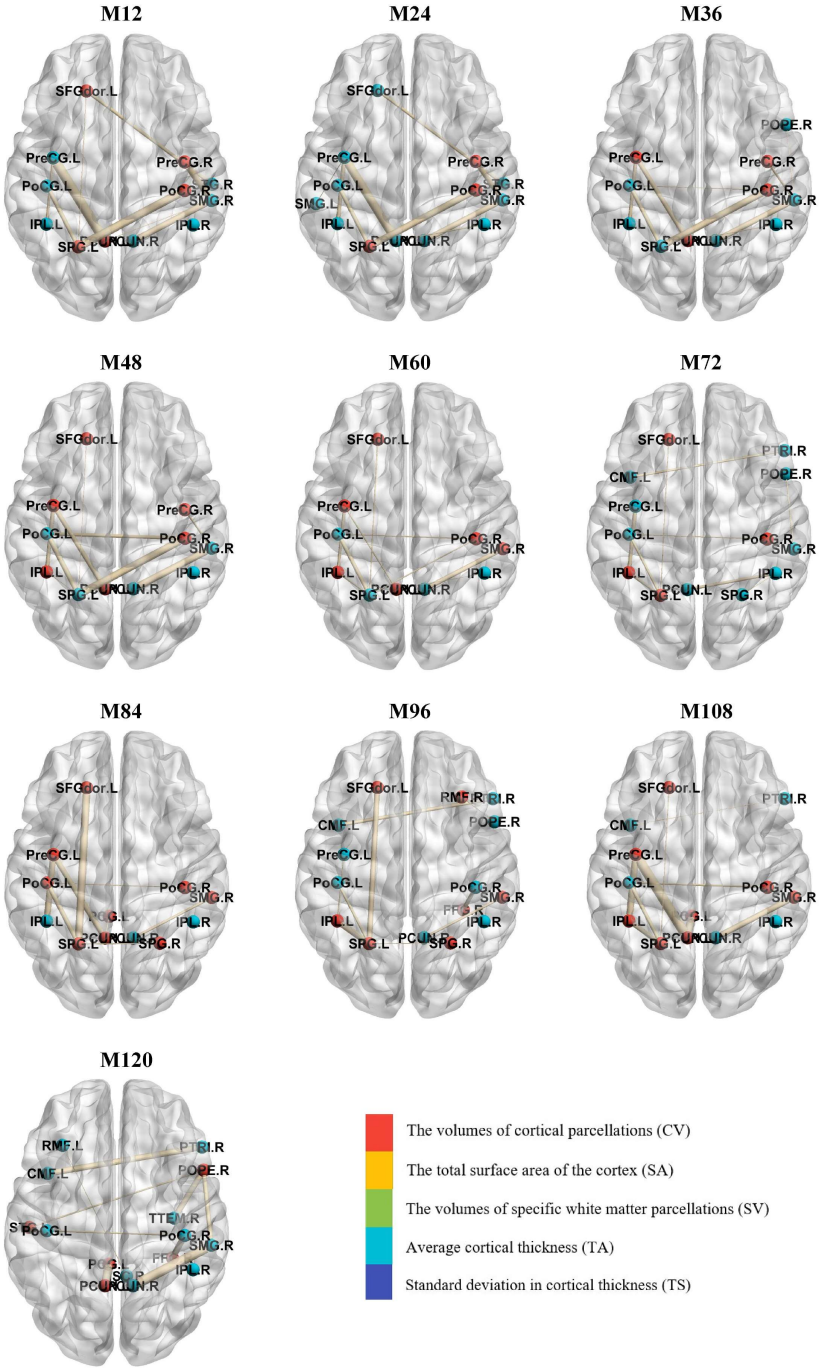


Figure D.1: Visualisation for the top-10 ranked brain biomarker correlations at various time points for the proposed AMDQ-TMTL approach for MMSE prediction. BrainNet Viewer's toolkit [305] was utilised for visualisation. The colors of the nodes indicate the various biomarker categories, while the thickness of edges represents the implications of biomarker correlations, with thicker edges signifying more important correlations between biomarkers. The abbreviations for the brain regions are in Appendix C.

D.2 For ADAS-Cog prediction targets

Table D.11: The top-10 rank brain biomarker correlations in time point M12 for the AMDQ-TMTL approach on ADAS-Cog prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.7289
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.6490
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.5172
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.4847
CTA. of R.ParsOpercularis - Vol(C). of L.SuperiorFrontal	0.4768
Vol(C). of R.Supramarginal - Vol(C). of L.Precuneus	0.4615
CTA. of R.ParsOpercularis - CTA. of R.Supramarginal	0.4555
Vol(C). of L.Precuneus - Vol(C). of L.Supramarginal	0.4450
CTA. of L.Precentral - CTA. of L.SuperiorFrontal	0.4384
CTA. of L.Postcentral - Vol(C). of L.SuperiorFrontal	0.4347

Table D.12: The top-10 rank brain biomarker correlations in time point M24 for the AMDQ-TMTL approach on ADAS-Cog prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.7953
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.6769
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.5577
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.5453
CTA. of R.ParsOpercularis - Vol(C). of L.SuperiorFrontal	0.5072
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.5037
CTA. of R.Precuneus - CTA. of R.Bankssts	0.4934
CTA. of R.ParsOpercularis - CTA. of R.Supramarginal	0.4870
Vol(C). of R.Supramarginal - Vol(C). of L.Precuneus	0.4857
Vol(C). of R.Postcentral - CTA. of R.Precentral	0.4814

Table D.13: The top-10 rank brain biomarker correlations in time point M36 for the AMDQ-TMTL approach on ADAS-Cog prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.8192
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.6541
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.5823
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.5703
Vol(C). of R.Paracentral - CTA. of R.Postcentral	0.5289
CTA. of R.ParsOpercularis - Vol(C). of L.SuperiorFrontal	0.5227
CTA. of L.SuperiorFrontal - CTA. of R.InferiorParietal	0.5172
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.5145
CTA. of R.Precuneus - CTA. of R.Bankssts	0.5133
Vol(C). of R.Postcentral - CTA. of R.Precentral	0.4993

Table D.14: The top-10 rank brain biomarker correlations in time point M48 for the AMDQ-TMTL approach on ADAS-Cog prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.9132
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.7031
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.6334
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.5822
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.5760
Vol(C). of R.Postcentral - CTA. of R.Precentral	0.5720
Vol(C). of R.Paracentral - CTA. of R.Postcentral	0.5608
Vol(C). of R.SuperiorParietal - CTA. of R.SuperiorParietal	0.5513
CTA. of R.Postcentral - Vol(C). of L.Supramarginal	0.5444
CTA. of R.ParsOpercularis - Vol(C). of L.SuperiorFrontal	0.5397

Table D.15: The top-10 rank brain biomarker correlations in time point M60 for the AMDQ-TMTL approach on ADAS-Cog prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.9073
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.7903
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.7598
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.7464
Vol(C). of R.Paracentral - CTA. of R.Postcentral	0.6387
CTA. of L.SuperiorFrontal - CTA. of R.InferiorParietal	0.6362
CTA. of R.RostralMiddleFrontal - CTA. of L.RostralMiddleFrontal	0.6359
Vol(C). of R.Supramarginal - Vol(C). of R.InferiorParietal	0.6346
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.6231
Vol(C). of L.Postcentral - Vol(C). of L.SuperiorParietal	0.6229

Table D.16: The top-10 rank brain biomarker correlations in time point M72 for the AMDQ-TMTL approach on ADAS-Cog prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.7080
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.6656
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.6249
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.5758
CTA. of R.RostralMiddleFrontal - CTA. of L.RostralMiddleFrontal	0.5757
Vol(C). of R.SuperiorParietal - CTA. of R.SuperiorParietal	0.5489
CTA. of L.CaudalMiddleFrontal - Vol(C). of R.InferiorParietal	0.5327
Vol(C). of L.LateralOrbitofrontal - CTA. of L.LateralOrbitofrontal	0.5288
CTA. of R.ParsOpercularis - CTA. of R.Supramarginal	0.5275
CTA. of L.SuperiorFrontal - CTA. of R.InferiorParietal	0.5217

Table D.17: The top-10 rank brain biomarker correlations in time point M84 for the AMDQ-TMTL approach on ADAS-Cog prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.7057
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.6537
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.6530
Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal	0.6083
CTA. of R.ParsOpercularis - Vol(C). of L.SuperiorFrontal	0.5662
Vol(C). of R.SuperiorParietal - CTA. of R.SuperiorParietal	0.5567
Vol(C). of R.SuperiorParietal - Vol(C). of L.MiddleTemporal	0.5442
Vol(C). of R.Paracentral - CTA. of R.Postcentral	0.5441
CTA. of L.CaudalMiddleFrontal - Vol(C). of L.Fusiform	0.5433
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.5400

Table D.18: The top-10 rank brain biomarker correlations in time point M96 for the AMDQ-TMTL approach on ADAS-Cog prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - CTA. of L.SuperiorParietal	0.6454
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.5994
CTA. of R.RostralMiddleFrontal - CTA. of L.RostralMiddleFrontal	0.5306
CTA. of R.RostralMiddleFrontal - CTA. of L.ParsOpercularis	0.5286
CTA. of R.ParsOpercularis - Vol(C). of L.SuperiorFrontal	0.5269
Vol(C). of L.LateralOrbitofrontal - CTA. of L.LateralOrbitofrontal	0.5226
Vol(C). of L.LateralOccipital - Vol(C). of R.LateralOccipital	0.5124
Vol(C). of R.Postcentral - Vol(C). of R.RostralMiddleFrontal	0.4964
CTA. of R.ParsOpercularis - CTA. of R.Supramarginal	0.4907
CTA. of R.Supramarginal - CTA. of R.Cuneus	0.4906

Table D.19: The top-10 rank brain biomarker correlations in time point M108 for the AMDQ-TMTL approach on ADAS-Cog prediction.

Brain biomarker correlation	Weight
Vol(C). of L.LateralOrbitofrontal - CTA. of L.LateralOrbitofrontal	0.6108
CTA. of R.TransverseTemporal - CTA. of L.Precentral	0.5695
Vol(C). of L.PosteriorCingulate - Vol(C). of L.Precuneus	0.5521
CTA. of L.LateralOrbitofrontal - Vol(C). of L.SuperiorParietal	0.5486
CTA. of R.Supramarginal - CTA. of R.Cuneus	0.5341
Vol(C). of R.SuperiorParietal - CTA. of R.SuperiorParietal	0.5236
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.5193
Vol(C). of L.Postcentral - Vol(C). of L.SuperiorParietal	0.5117
Vol(C). of R.SuperiorParietal - Vol(C). of L.MiddleTemporal	0.4988
Vol(C). of L.Postcentral - CTA. of L.Postcentral	0.4979

Table D.20: The top-10 rank brain biomarker correlations in time point M120 for the AMDQ-TMTL approach on ADAS-Cog prediction.

Brain biomarker correlation	Weight
Vol(C). of L.LateralOrbitofrontal - CTA. of L.LateralOrbitofrontal	0.6601
Vol(C). of R.ParsOpercularis - Vol(C). of L.SuperiorTemporal	0.6141
Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal	0.5697
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.5539
CTA. of L.CaudalMiddleFrontal - Vol(C). of L.Fusiform	0.5325
CTA. of R.ParsOpercularis - CTA. of R.Supramarginal	0.5255
Vol(C). of R.SuperiorParietal - Vol(C). of L.MiddleTemporal	0.5205
CTA. of R.TransverseTemporal - CTA. of L.Precentral	0.5135
Vol(C). of L.Precuneus - Vol(C). of L.Supramarginal	0.5117
Vol(C). of R.Supramarginal - Vol(C). of R.InferiorParietal	0.5114

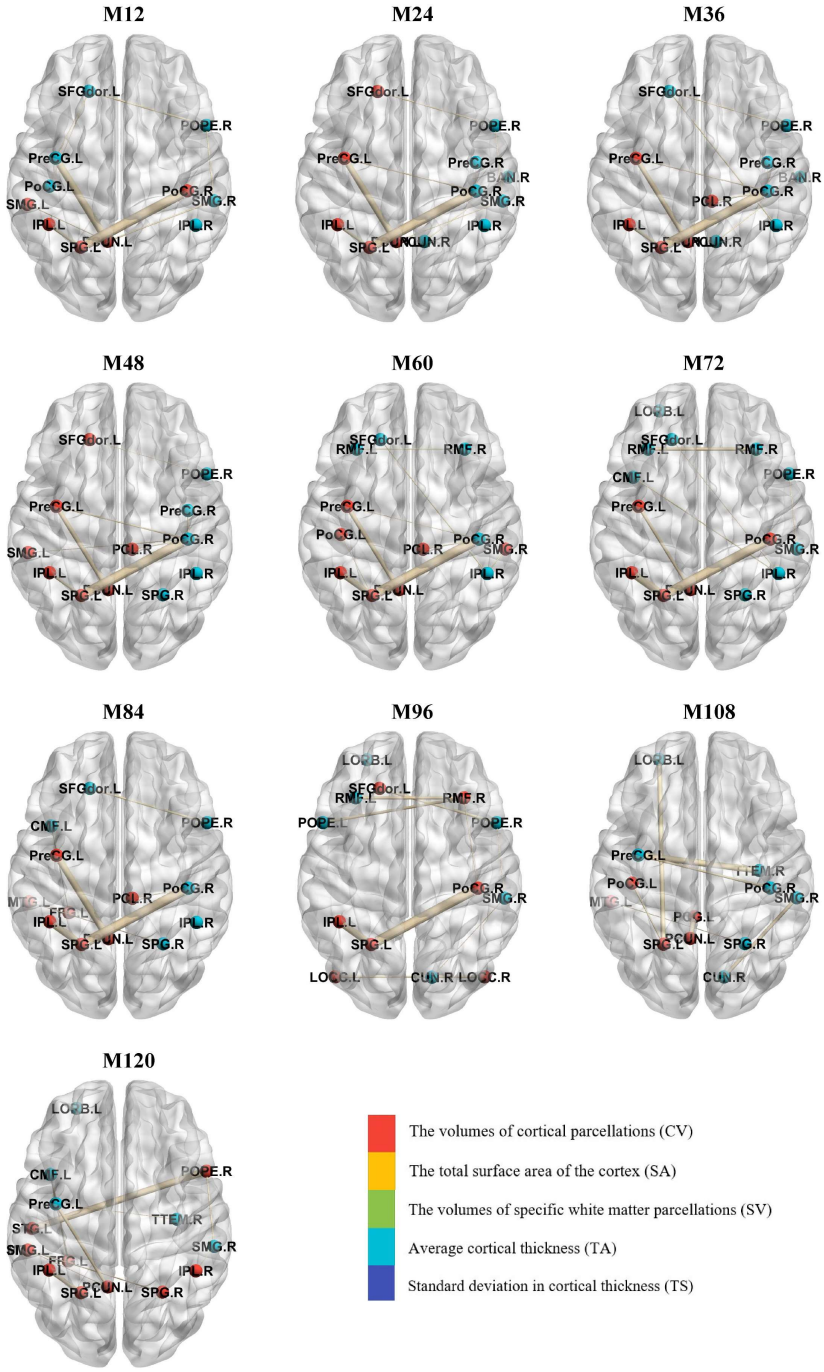


Figure D.2: Visualisation for the top-10 ranked brain biomarker correlations at various time points for the proposed AMDQ-TMTL approach for ADAS-Cog prediction. BrainNet Viewer's toolkit [305] was utilised for visualisation. The colors of the nodes indicate the various biomarker categories, while the thickness of edges represents the implications of biomarker correlations, with thicker edges signifying more important correlations between biomarkers. The abbreviations for the brain regions are in Appendix C.

D.3 For RAVLT TOTAL prediction targets

Table D.21:The top-10 rank brain biomarker correlations in time point M12 for the AMDQ-TMTL approach on RAVLT TOTAL prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal	0.5040
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.5015
CTA. of R.SuperiorParietal - Vol(C). of L.Precuneus	0.4905
Vol(C). of R.Postcentral - CTA. of L.SuperiorFrontal	0.4678
CTA. of R.Paracentral - Vol(C). of R.Supramarginal	0.4482
CTA. of R.ParsTriangularis - CTA. of L.Postcentral	0.4293
CTA. of L.LateralOccipital - CTA. of L.Precuneus	0.4209
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.4181
CTA. of R.Precuneus - CTA. of L.InferiorParietal	0.4077
Vol(C). of R.SuperiorFrontal - Vol(C). of R.Supramarginal	0.3938

Table D.22: The top-10 rank brain biomarker correlations in time point M24 for the AMDQ-TMTL approach on RAVLT TOTAL prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal	0.5266
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.5187
CTA. of R.SuperiorParietal - Vol(C). of L.Precuneus	0.4959
Vol(C). of R.Postcentral - CTA. of L.SuperiorFrontal	0.4920
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.4651
CTA. of L.LateralOccipital - CTA. of L.Precuneus	0.4419
CTA. of R.Paracentral - Vol(C). of R.Supramarginal	0.4411
CTA. of R.ParsTriangularis - CTA. of L.Postcentral	0.4307
CTA. of R.Precuneus - CTA. of L.InferiorParietal	0.4297
CTA. of L.Postcentral - CTA. of L.SuperiorParietal	0.4264

Table D.23: The top-10 rank brain biomarker correlations in time point M36 for the AMDQ-TMTL approach on RAVLT TOTAL prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal	0.5371
Vol(C). of R.Postcentral - CTA. of L.SuperiorFrontal	0.5364
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.5323
CTA. of L.Postcentral - CTA. of L.SuperiorParietal	0.5053
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.4939
CTA. of R.SuperiorParietal - Vol(C). of L.Precuneus	0.4737
CTA. of R.ParsTriangularis - CTA. of L.Postcentral	0.4684
CTA. of R.Paracentral - Vol(C). of R.Supramarginal	0.4593
CTA. of R.Postcentral - CTA. of L.SuperiorFrontal	0.4353
CTA. of L.SuperiorFrontal - CTA. of L.Supramarginal	0.4280

Table D.24: The top-10 rank brain biomarker correlations in time point M48 for the AMDQ-TMTL approach on RAVLT TOTAL prediction.

Brain biomarker correlation	Weight
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.6089
Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal	0.5930
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.5521
CTA. of L.Postcentral - CTA. of L.SuperiorParietal	0.5510
Vol(C). of R.Postcentral - CTA. of L.SuperiorFrontal	0.5311
CTA. of R.SuperiorParietal - Vol(C). of L.Precuneus	0.5220
CTA. of R.ParsTriangularis - CTA. of L.Postcentral	0.5009
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.4653
CTA. of R.Postcentral - CTA. of L.SuperiorFrontal	0.4489
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.4439

Table D.25: The top-10 rank brain biomarker correlations in time point M60 for the AMDQ-TMTL approach on RAVLT TOTAL prediction.

Brain biomarker correlation	Weight
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.5343
Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal	0.5120
CTA. of R.SuperiorParietal - Vol(C). of L.Precuneus	0.5014
CTA. of L.Postcentral - CTA. of L.SuperiorParietal	0.4699
Vol(C). of R.Postcentral - CTA. of L.SuperiorFrontal	0.4693
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.4598
CTA. of R.SuperiorParietal - CTA. of L.RostralMiddleFrontal	0.4531
CTA. of R.Paracentral - Vol(C). of R.Supramarginal	0.4493
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.4353
CTA. of L.SuperiorFrontal - CTA. of L.Supramarginal	0.4320

Table D.26: The top-10 rank brain biomarker correlations in time point M72 for the AMDQ-TMTL approach on RAVLT TOTAL prediction.

Brain biomarker correlation	Weight
CTA. of L.Postcentral - CTA. of L.SuperiorParietal	0.5490
CTA. of R.ParsTriangularis - CTA. of L.Postcentral	0.5462
Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal	0.5091
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.4983
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.4888
CTA. of L.SuperiorFrontal - CTA. of L.Supramarginal	0.4852
CTA. of L.CaudalMiddleFrontal - CTA. of L.InferiorParietal	0.4795
CTA. of R.ParsOpercularis - CTA. of R.SuperiorFrontal	0.4766
CTA. of R.SuperiorParietal - CTA. of L.RostralMiddleFrontal	0.4723
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.4703

Table D.27: The top-10 rank brain biomarker correlations in time point M84 for the AMDQ-TMTL approach on RAVLT TOTAL prediction.

Brain biomarker correlation	Weight
Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal	0.5090
Vol(C). of L.Precentral - Vol(C). of L.Precuneus	0.4820
CTA. of R.ParsTriangularis - CTA. of L.Postcentral	0.4790
CTA. of L.Postcentral - CTA. of L.SuperiorParietal	0.4671
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.4646
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.4475
Vol(C). of R.Postcentral - CTA. of L.SuperiorFrontal	0.4468
CTA. of R.RostralMiddleFrontal - Vol(C). of L.SuperiorFrontal	0.4317
CTA. of L.SuperiorFrontal - CTA. of L.Supramarginal	0.4230
CTA. of R.SuperiorParietal - Vol(C). of L.Precuneus	0.4185

Table D.28: The top-10 rank brain biomarker correlations in time point M96 for the AMDQ-TMTL approach on RAVLT TOTAL prediction.

Brain biomarker correlation	Weight
CTA. of R.ParsTriangularis - CTA. of L.Postcentral	0.5427
Vol(C). of R.Precuneus - Vol(C). of L.SuperiorTemporal	0.5002
Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal	0.4884
CTA. of R.ParsTriangularis - Vol(C). of R.Precuneus	0.4737
CTA. of L.Postcentral - CTA. of L.SuperiorParietal	0.4605
CTA. of L.SuperiorFrontal - CTA. of L.Supramarginal	0.4577
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.4547
Vol(C). of R.Postcentral - CTA. of L.SuperiorFrontal	0.4503
CTA. of R.RostralMiddleFrontal - Vol(C). of L.SuperiorFrontal	0.4493
CTA. of R.ParsOpercularis - CTA. of R.SuperiorFrontal	0.4452

Table D.29: The top-10 rank brain biomarker correlations in time point M108 for the AMDQ-TMTL approach on RAVLT TOTAL prediction.

Brain biomarker correlation	Weight
CTA. of R.ParsTriangularis - CTA. of L.Postcentral	0.4888
CTA. of R.TransverseTemporal - CTA. of L.Precentral	0.4479
Vol(C). of L.Postcentral - Vol(C). of R.CaudalMiddleFrontal	0.4418
CTA. of R.Postcentral - Vol(C). of L.Precentral	0.4291
Vol(C). of R.Precuneus - Vol(C). of L.SuperiorTemporal	0.4266
CTA. of R.SuperiorParietal - Vol(C). of L.Precuneus	0.4043
CTA. of L.ParsOpercularis - Vol(C). of L.SuperiorParietal	0.3974
CTA. of L.Paracentral - CTA. of R.LateralOrbitofrontal	0.3939
Vol(C). of R.SuperiorFrontal - Vol(C). of R.Supramarginal	0.3849
CTA. of L.Cuneus - CTA. of L.MiddleTemporal	0.3838

Table D.30: The top-10 rank brain biomarker correlations in time point M120 for the AMDQ-TMTL approach on RAVLT TOTAL prediction.

Brain biomarker correlation	Weight
CTA. of R.ParsTriangularis - CTA. of L.Postcentral	0.5027
Vol(C). of R.RostralMiddleFrontal - Vol(C). of L.InferiorParietal	0.4293
CTA. of L.Postcentral - CTA. of L.SuperiorParietal	0.4256
Vol(C). of R.Paracentral - CTA. of L.Cuneus	0.4180
CTA. of R.ParsTriangularis - CTA. of L.InferiorParietal	0.4044
Vol(C). of R.ParsOpercularis - CTA. of R.TransverseTemporal	0.3983
Vol(C). of R.SuperiorFrontal - CTA. of R.CaudalMiddleFrontal	0.3924
CTA. of L.SuperiorFrontal - CTA. of L.Supramarginal	0.3745
CTA. of R.SuperiorParietal - Vol(C). of L.Precuneus	0.3509
CTA. of R.SuperiorFrontal - CTA. of L.Paracentral	0.3506

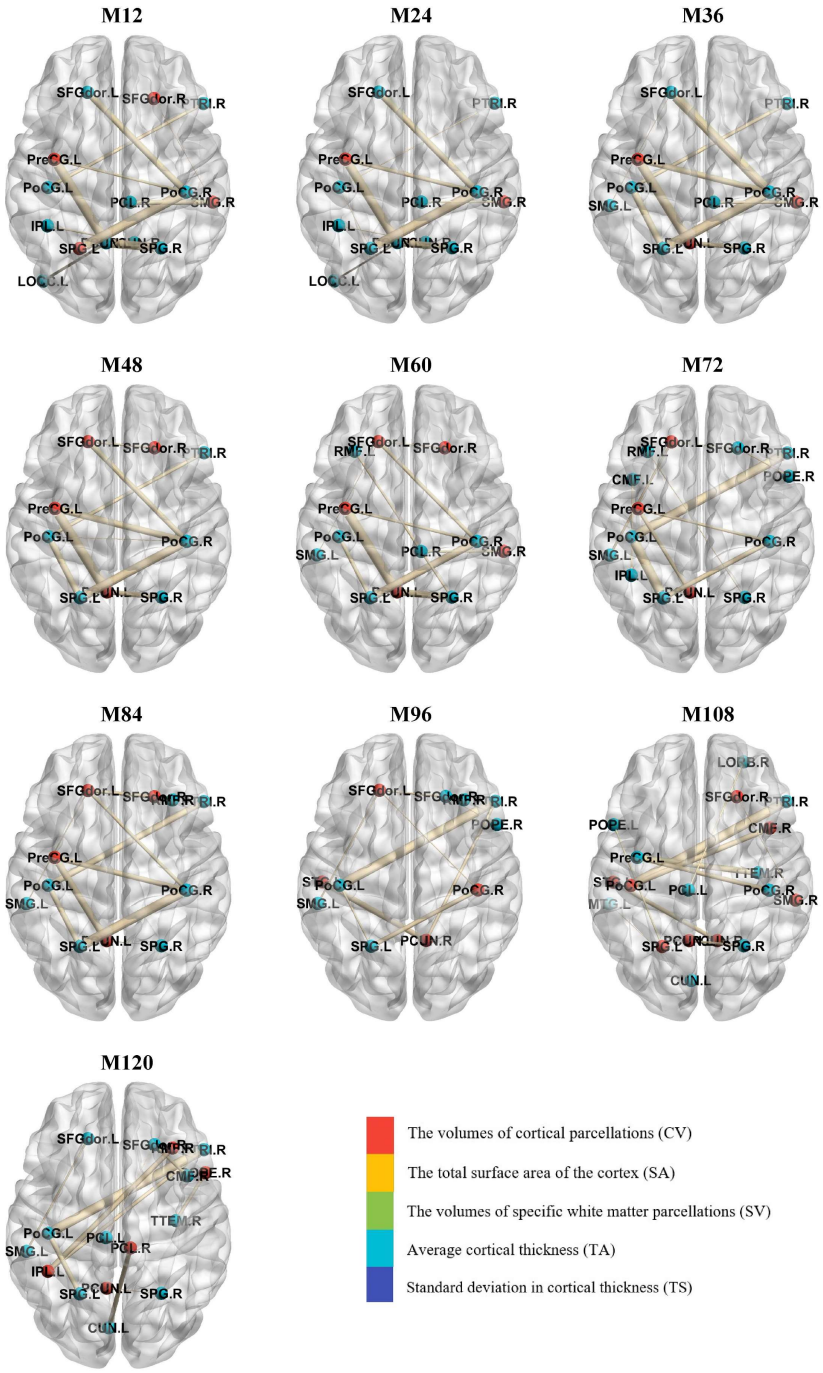


Figure D.3: Visualisation for the top-10 ranked brain biomarker correlations at various time points for the proposed AMDQ-TMTL approach for RAVLT TOTAL prediction. BrainNet Viewer's toolkit [305] was utilised for visualisation. The colors of the nodes indicate the various biomarker categories, while the thickness of edges represents the implications of biomarker correlations, with thicker edges signifying more important correlations between biomarkers. The abbreviations for the brain regions are in Appendix C.

D.4 For FLU ANIM prediction targets

Table D.31: The top-10 rank brain biomarker correlations in time point M12 for the AMDQ-TMTL approach on FLU ANIM prediction.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.7702
Vol(C). of L.Precentral - CTA. of L.Supramarginal	0.5841
Vol(C). of L.Precuneus - CTA. of L.SuperiorParietal	0.5585
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.5528
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.5498
Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal	0.5155
Vol(C). of R.SuperiorFrontal - Vol(C). of L.Precentral	0.5103
Vol(C). of R.Precentral - CTA. of R.CaudalMiddleFrontal	0.4937
CTA. of R.Postcentral - Vol(C). of L.Supramarginal	0.4882
Vol(C). of R.RostralMiddleFrontal - CTA. of L.SuperiorParietal	0.4790

Table D.32: The top-10 rank brain biomarker correlations in time point M24 for the AMDQ-TMTL approach on FLU ANIM prediction.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.8434
Vol(C). of L.Precentral - CTA. of L.Supramarginal	0.6257
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.6192
Vol(C). of L.Precuneus - CTA. of L.SuperiorParietal	0.5727
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.5697
Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal	0.5486
CTA. of R.Postcentral - Vol(C). of L.Supramarginal	0.5274
Vol(C). of R.Precentral - CTA. of R.CaudalMiddleFrontal	0.5178
Vol(C). of R.SuperiorFrontal - Vol(C). of L.Precentral	0.5130
Vol(C). of L.Precentral - CTA. of R.CaudalMiddleFrontal	0.4889

Table D.33: The top-10 rank brain biomarker correlations in time point M36 for the AMDQ-TMTL approach on FLU ANIM prediction.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.8043
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.6619
Vol(C). of L.Precentral - CTA. of L.Supramarginal	0.6173
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.6101
Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal	0.5565
Vol(C). of L.Precuneus - CTA. of L.SuperiorParietal	0.5346
Vol(C). of R.Precentral - CTA. of R.CaudalMiddleFrontal	0.5140
CTA. of R.Postcentral - Vol(C). of L.Supramarginal	0.5006
CTA. of R.Paracentral - CTA. of R.InferiorParietal	0.4872
Vol(C). of R.SuperiorFrontal - Vol(C). of L.Precentral	0.4802

Table D.34: The top-10 rank brain biomarker correlations in time point M48 for the AMDQ-TMTL approach on FLU ANIM prediction.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.8478
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.6719
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.6642
Vol(C). of L.Precentral - CTA. of L.Supramarginal	0.6328
Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal	0.5968
CTA. of R.Postcentral - Vol(C). of L.Supramarginal	0.5732
Vol(C). of L.Precuneus - CTA. of L.SuperiorParietal	0.5709
Vol(C). of L.InferiorParietal - Vol(C). of L.Precuneus	0.5158
Vol(C). of R.Postcentral - Vol(C). of L.SuperiorParietal	0.5132
Vol(C). of L.Supramarginal - Vol(C). of R.LateralOccipital	0.4995

Table D.35: The top-10 rank brain biomarker correlations in time point M60 for the AMDQ-TMTL approach on FLU ANIM prediction.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.8928
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.7570
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.6999
Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal	0.6586
Vol(C). of L.Precentral - CTA. of L.Supramarginal	0.6411
Vol(C). of R.Precuneus - CTA. of L.RostralMiddleFrontal	0.6095
Vol(C). of L.Precuneus - CTA. of L.SuperiorParietal	0.5741
Vol(C). of L.InferiorParietal - Vol(C). of L.Precuneus	0.5740
Vol(C). of L.Postcentral - CTA. of L.Postcentral	0.5577
CTA. of R.Postcentral - Vol(C). of L.Supramarginal	0.5512

Table D.36: The top-10 rank brain biomarker correlations in time point M72 for the AMDQ-TMTL approach on FLU ANIM prediction.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.7293
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.6453
Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal	0.6270
Vol(C). of L.Precentral - CTA. of L.Supramarginal	0.6077
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.5662
Vol(C). of R.Precuneus - CTA. of L.RostralMiddleFrontal	0.5509
CTA. of R.RostralMiddleFrontal - CTA. of L.RostralMiddleFrontal	0.5302
CTA. of R.Paracentral - CTA. of R.InferiorParietal	0.4919
CTA. of R.ParsTriangularis - CTA. of L.InferiorParietal	0.4898
CTA. of L.CaudalMiddleFrontal - CTA. of L.RostralMiddleFrontal	0.4749

Table D.37: The top-10 rank brain biomarker correlations in time point M84 for the AMDQ-TMTL approach on FLU ANIM prediction.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.8459
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.7549
Vol(C). of L.Precentral - CTA. of L.Supramarginal	0.6278
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.5966
Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal	0.5369
Vol(C). of R.Precuneus - CTA. of L.RostralMiddleFrontal	0.5172
Vol(C). of L.Postcentral - CTA. of L.Postcentral	0.5161
CTA. of L.Postcentral - CTA. of R.MiddleTemporal	0.5017
CTA. of R.RostralMiddleFrontal - CTA. of L.RostralMiddleFrontal	0.4996
CTA. of R.Paracentral - CTA. of R.InferiorParietal	0.4894

Table D.38: The top-10 rank brain biomarker correlations in time point M96 for the AMDQ-TMTL approach on FLU ANIM prediction.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.6547
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.5685
Vol(C). of L.Precentral - CTA. of L.Supramarginal	0.4725
Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal	0.4515
CTA. of R.ParsTriangularis - CTA. of L.InferiorParietal	0.4479
CTA. of R.Precentral - CTA. of R.SuperiorTemporal	0.4409
Vol(C). of L.Postcentral - CTA. of L.Postcentral	0.4377
CTA. of L.Postcentral - CTA. of R.MiddleTemporal	0.4113
CTA. of R.Postcentral - Vol(C). of L.Supramarginal	0.4102
CTA. of R.RostralMiddleFrontal - CTA. of L.RostralMiddleFrontal	0.4084

Table D.39: The top-10 rank brain biomarker correlations in time point M108 for the AMDQ-TMTL approach on FLU ANIM prediction.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.8272
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.6381
Vol(C). of R.Postcentral - CTA. of L.Postcentral	0.6088
Vol(C). of R.Postcentral - CTA. of L.Pericalcarine	0.5796
Vol(C). of L.Postcentral - CTA. of L.Postcentral	0.5566
Vol(C). of R.Precuneus - CTA. of L.RostralMiddleFrontal	0.5398
Vol(C). of L.InferiorParietal - Vol(C). of L.Precuneus	0.5392
CTA. of R.ParsTriangularis - CTA. of L.InferiorParietal	0.5239
Vol(C). of R.SuperiorFrontal - CTA. of L.InferiorParietal	0.4998
CTA. of L.LateralOrbitofrontal - Vol(C). of L.SuperiorParietal	0.4955

Table D.40: The top-10 rank brain biomarker correlations in time point M120 for the AMDQ-TMTL approach on FLU ANIM prediction.

Brain biomarker correlation	Weight
Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal	0.8582
Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal	0.7873
Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal	0.7067
Vol(C). of R.Precuneus - CTA. of L.RostralMiddleFrontal	0.6769
CTA. of R.RostralMiddleFrontal - CTA. of L.RostralMiddleFrontal	0.6116
Vol(C). of R.Postcentral - CTA. of L.Pericalcarine	0.5824
CTA. of L.CaudalMiddleFrontal - CTA. of L.RostralMiddleFrontal	0.5710
Vol(C). of L.Postcentral - CTA. of L.Postcentral	0.5656
Vol(C). of R.Precuneus - Vol(C). of R.Supramarginal	0.5431
CTA. of L.Postcentral - CTA. of R.MiddleTemporal	0.5412

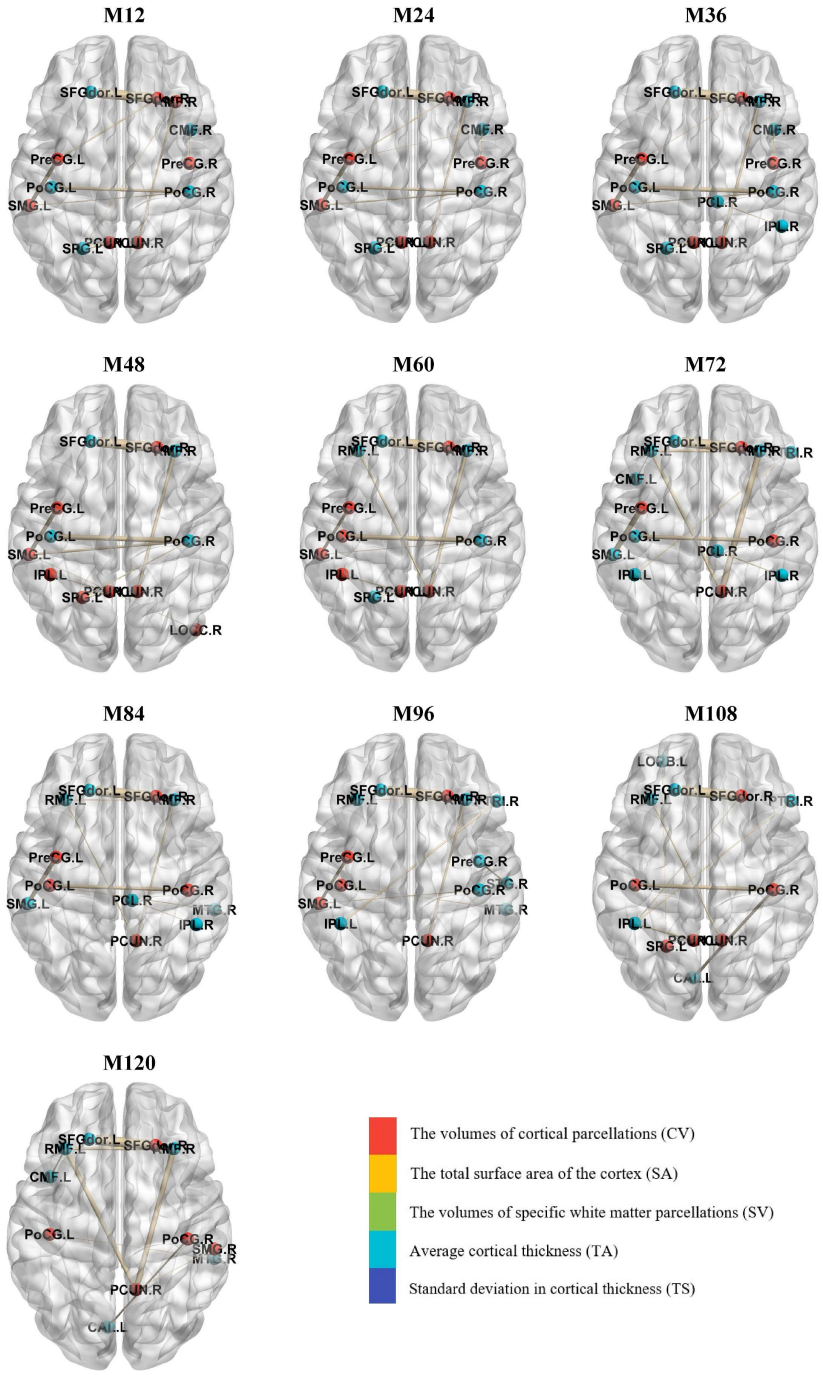


Figure D.4: Visualisation for the top-10 ranked brain biomarker correlations at various time points for the proposed AMDQ-TMTL approach for FLU ANIM prediction. BrainNet Viewer's toolkit [305] was utilised for visualisation. The colors of the nodes indicate the various biomarker categories, while the thickness of edges represents the implications of biomarker correlations, with thicker edges signifying more important correlations between biomarkers. The abbreviations for the brain regions are in Appendix C.

Appendix E: Information on potential indicators for AD early detection

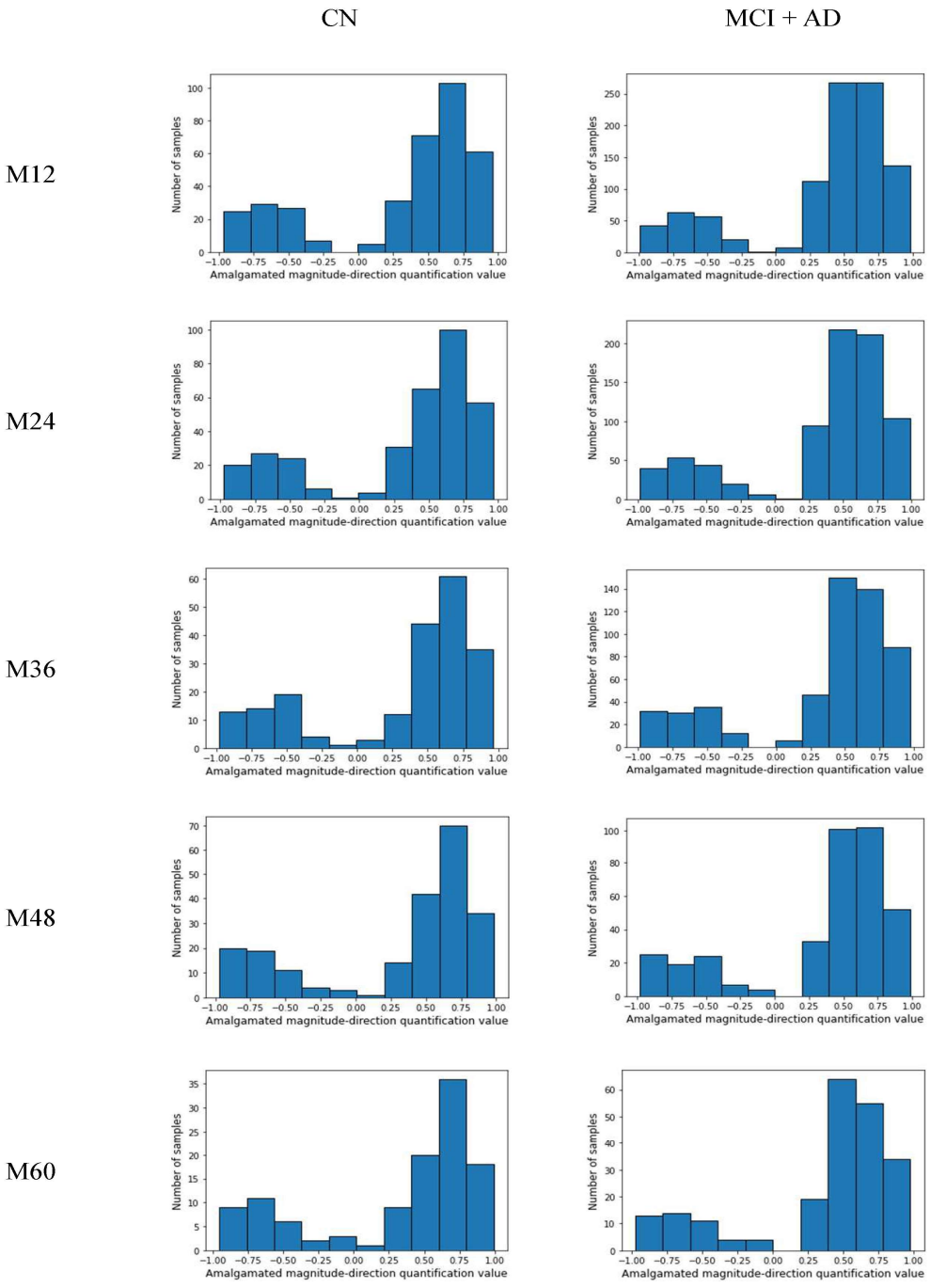


Figure E.1: Differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

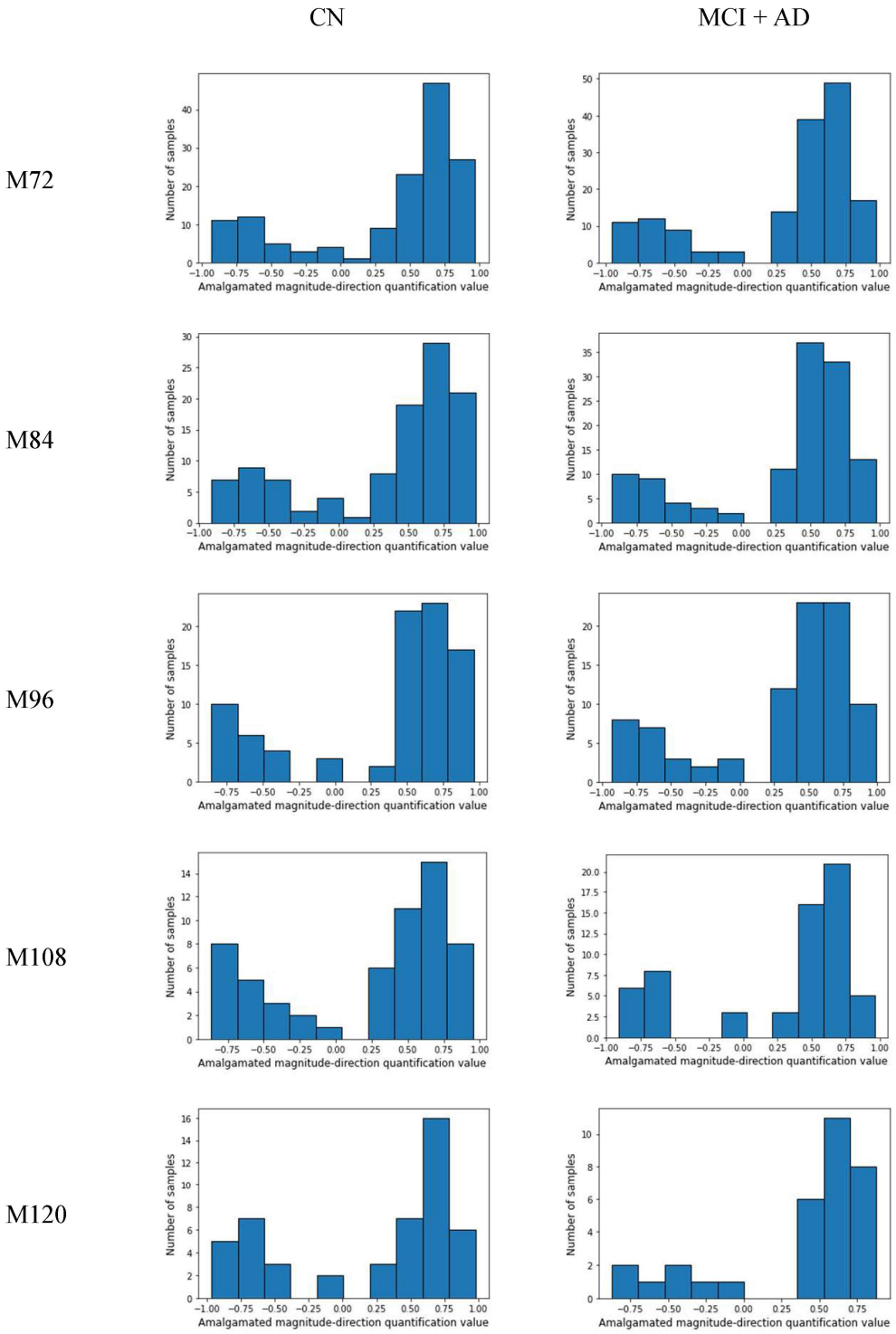


Figure E.2: Continuation of Figure E.1, differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

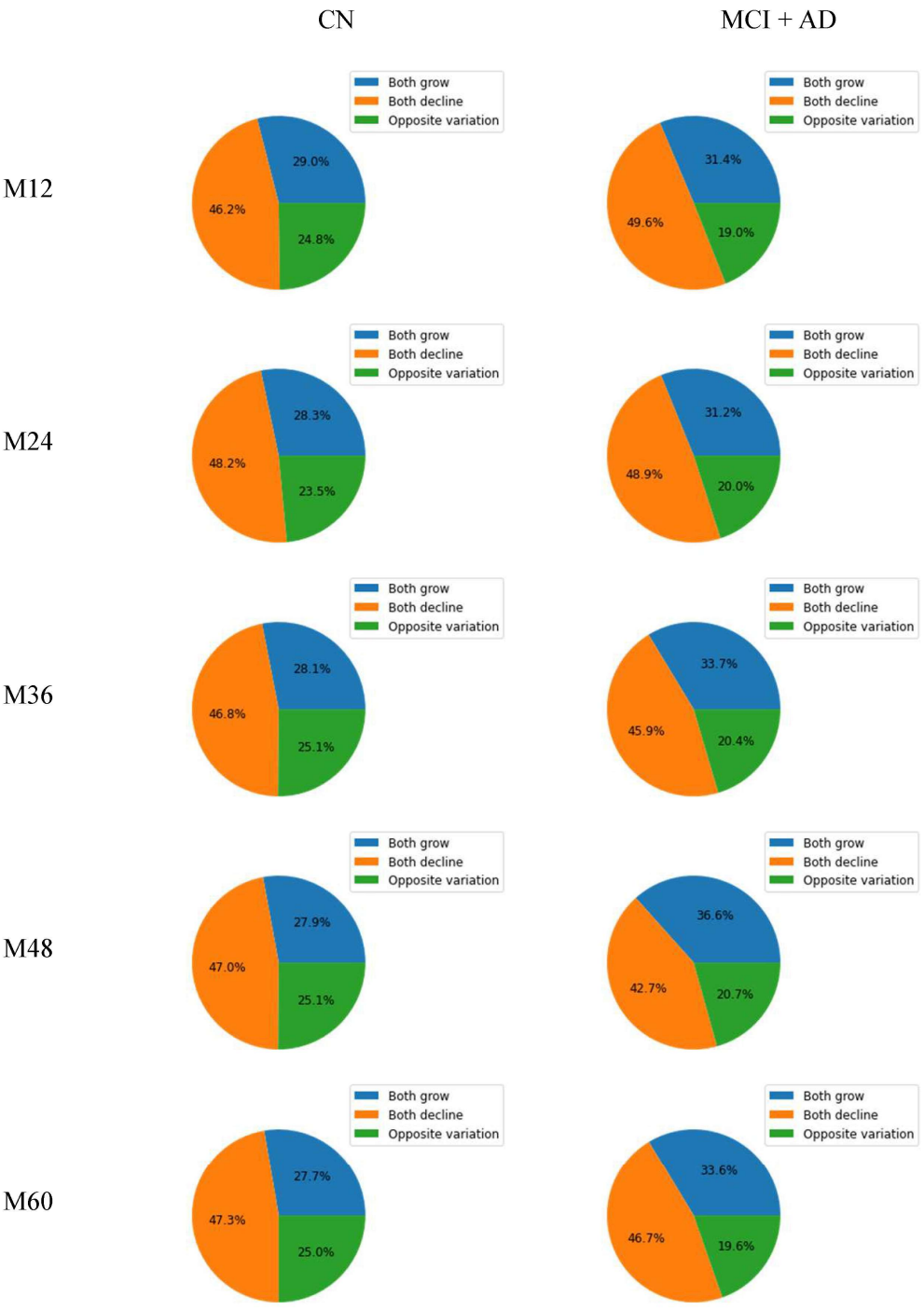


Figure E.3: Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

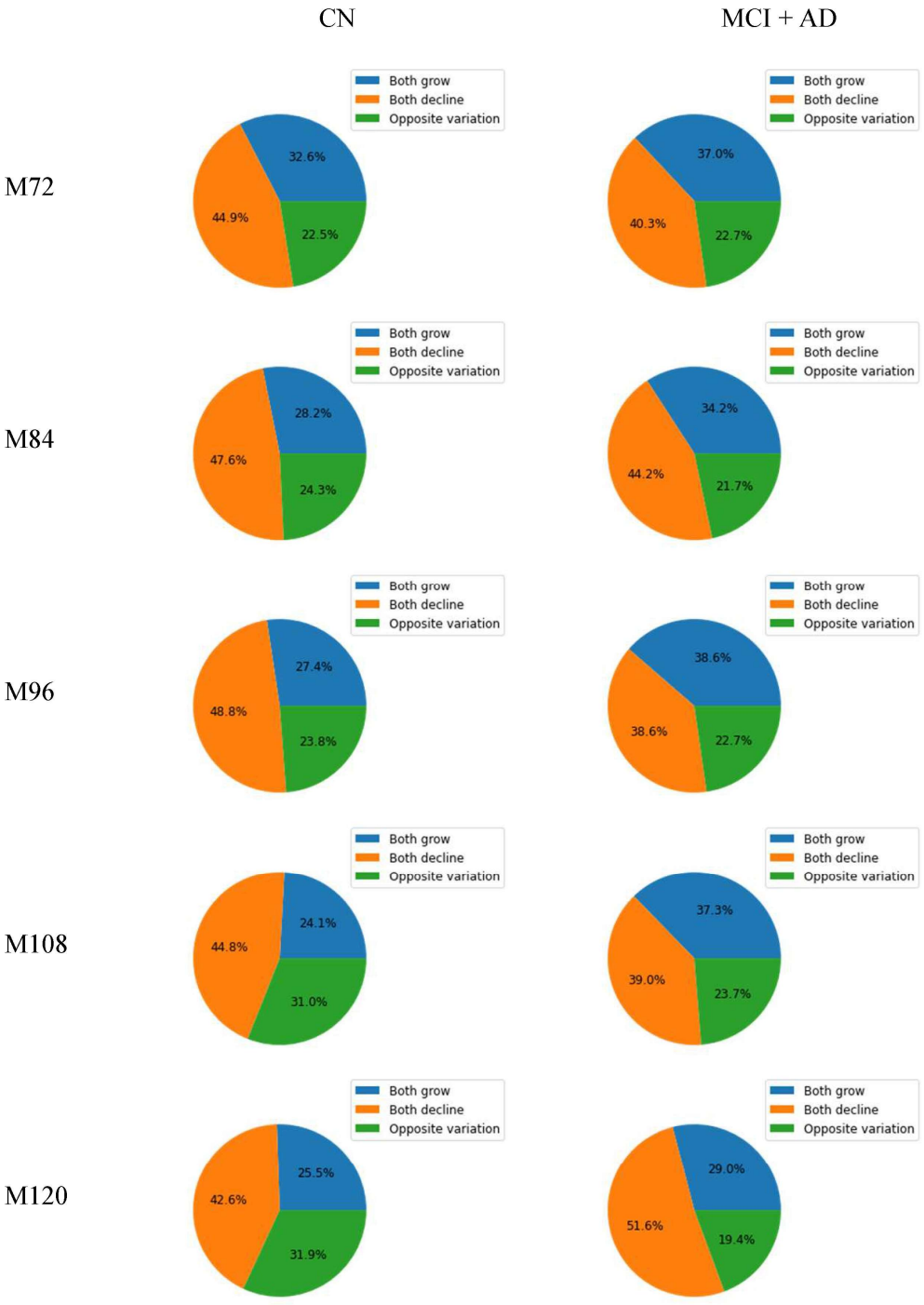


Figure E.4: Continuation of Figure E.3, Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of R.InferiorParietal - CTA. of R.InferiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

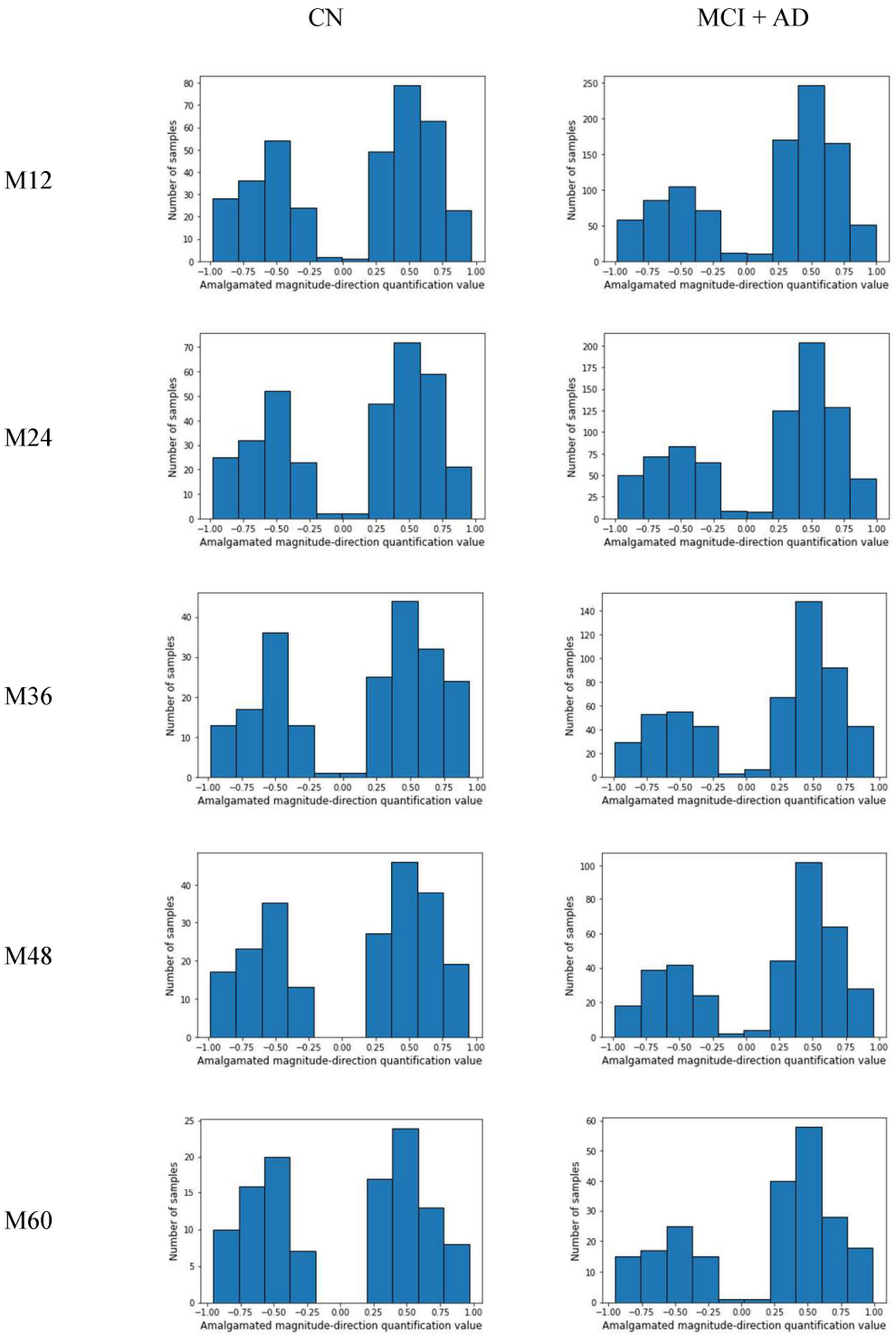


Figure E.5: Differences in the distribution of early stage (BL-M06) AMDQ quantitative values for CTA. of R.Precuneus - Vol(C). of R.Supramarginal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

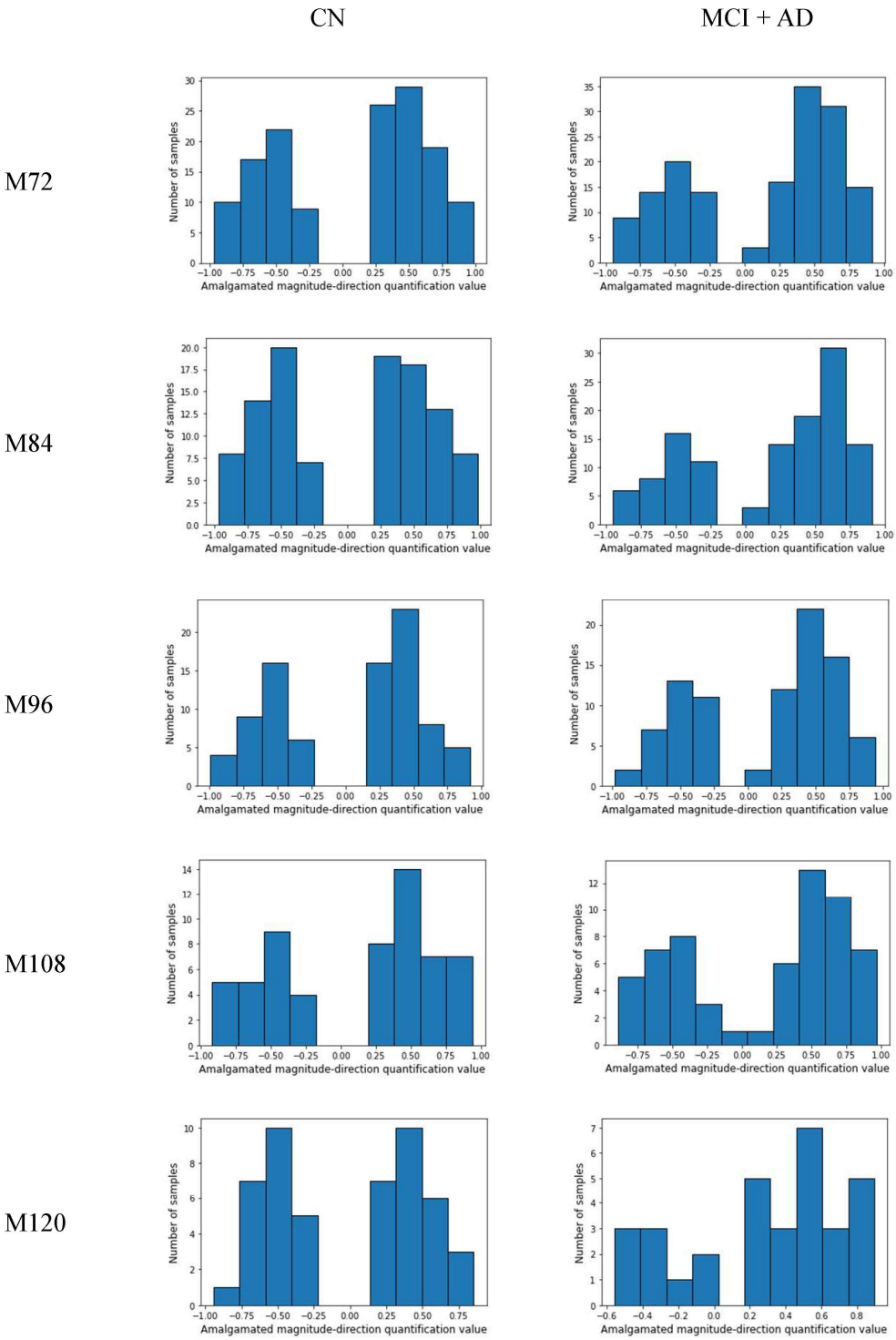


Figure E.6: Continuation of Figure E.5, differences in the distribution of early stage (BL-M06) AMDQ quantitative values for CTA. of R.Precuneus - Vol(C). of R.Supramarginal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

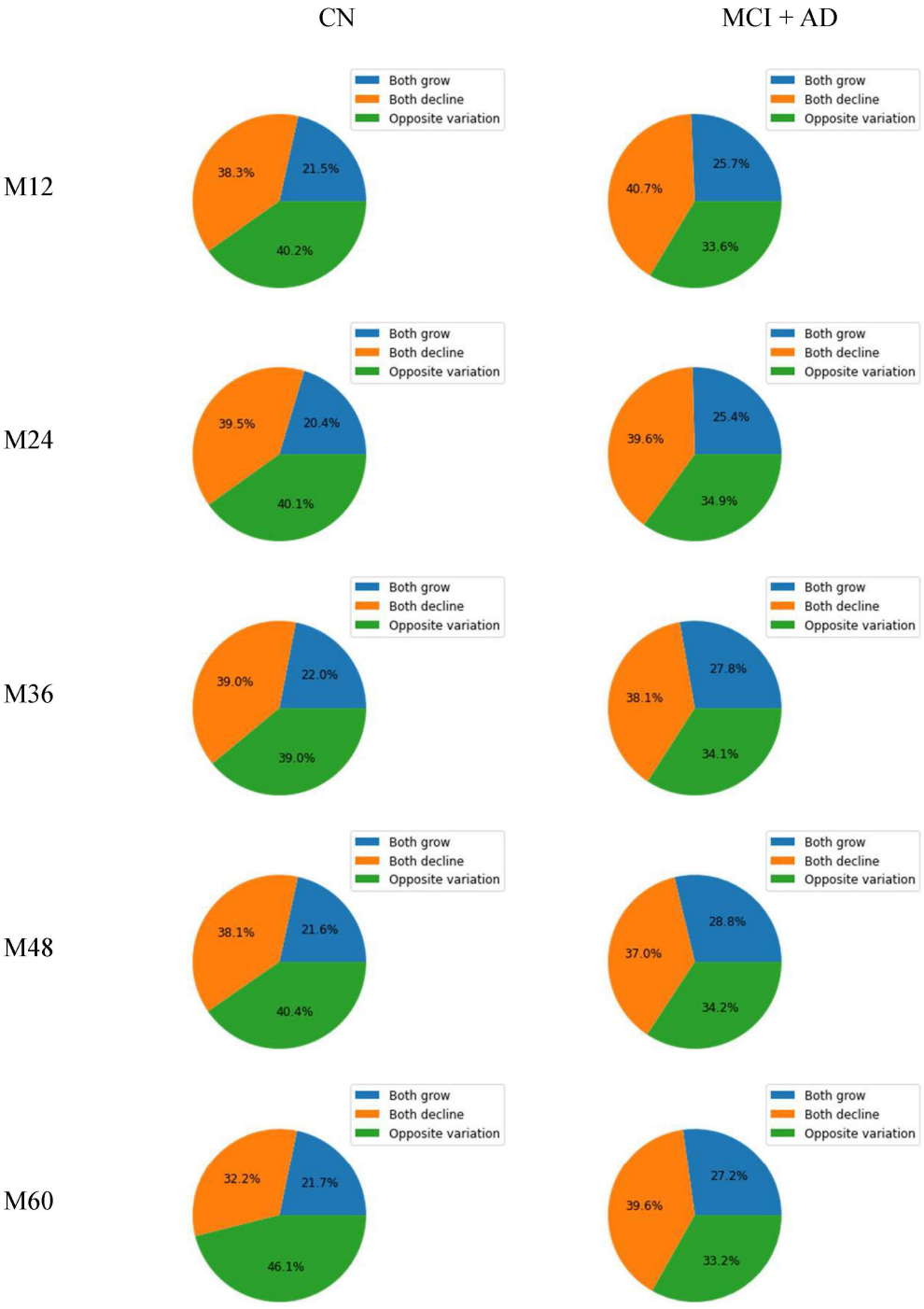


Figure E.7: Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for CTA. of R.Precuneus - Vol(C). of R.Supramarginal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

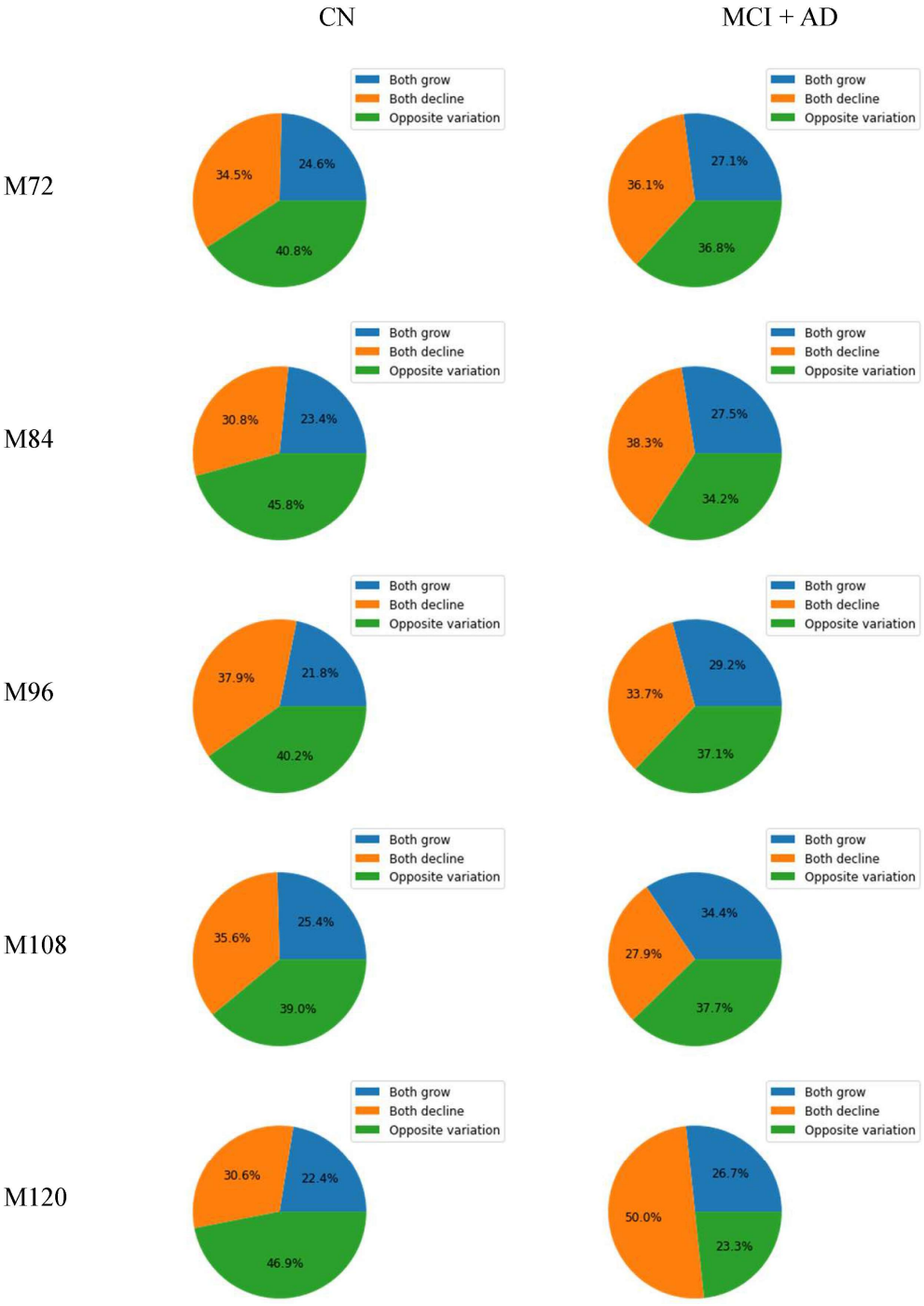


Figure E.8: Continuation of Figure E.7, Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for CTA. of R.Precuneus - Vol(C). of R.Supramarginal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

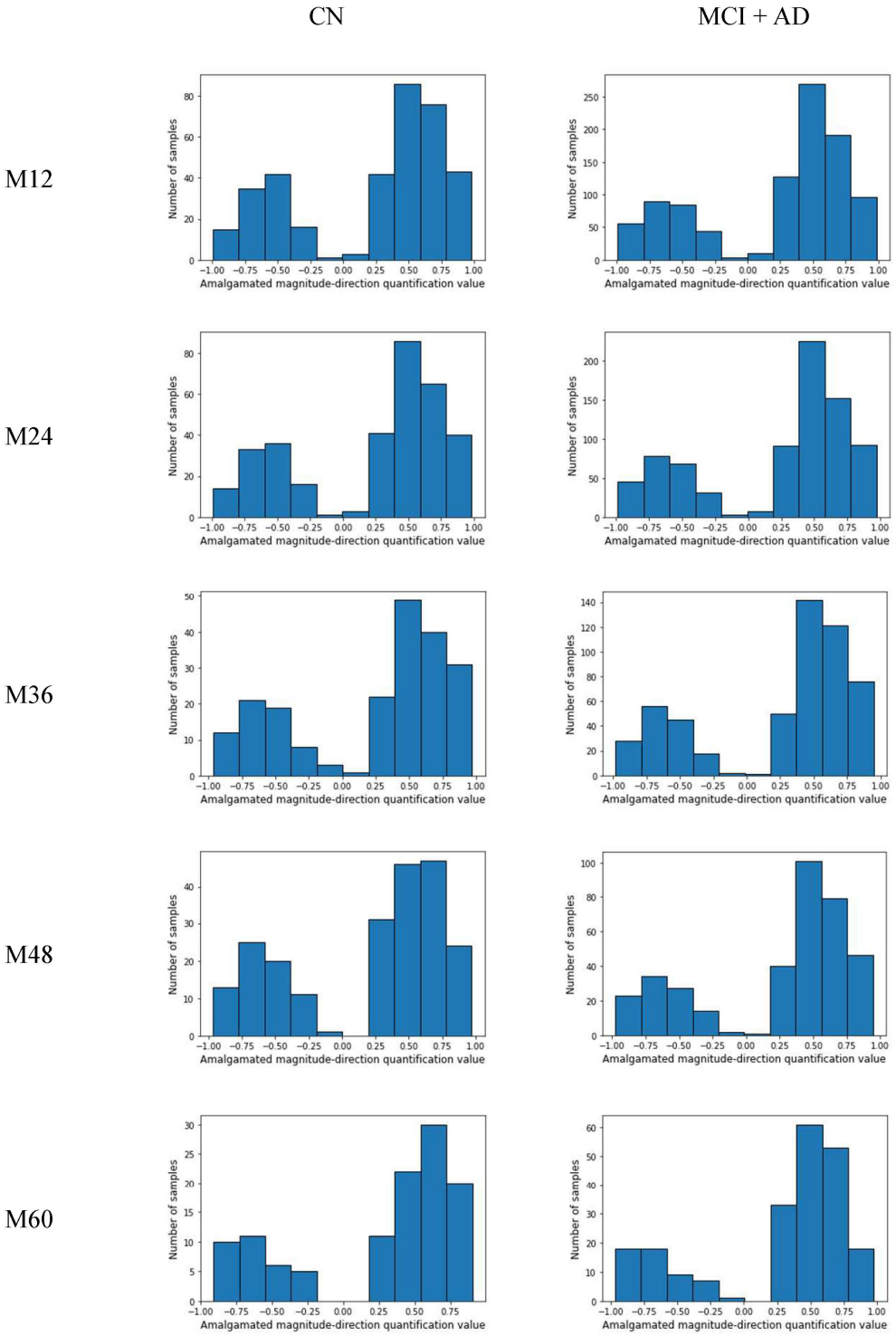


Figure E.9: Differences in the distribution of early stage (BL-M06) AMDQ quantitative values for CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

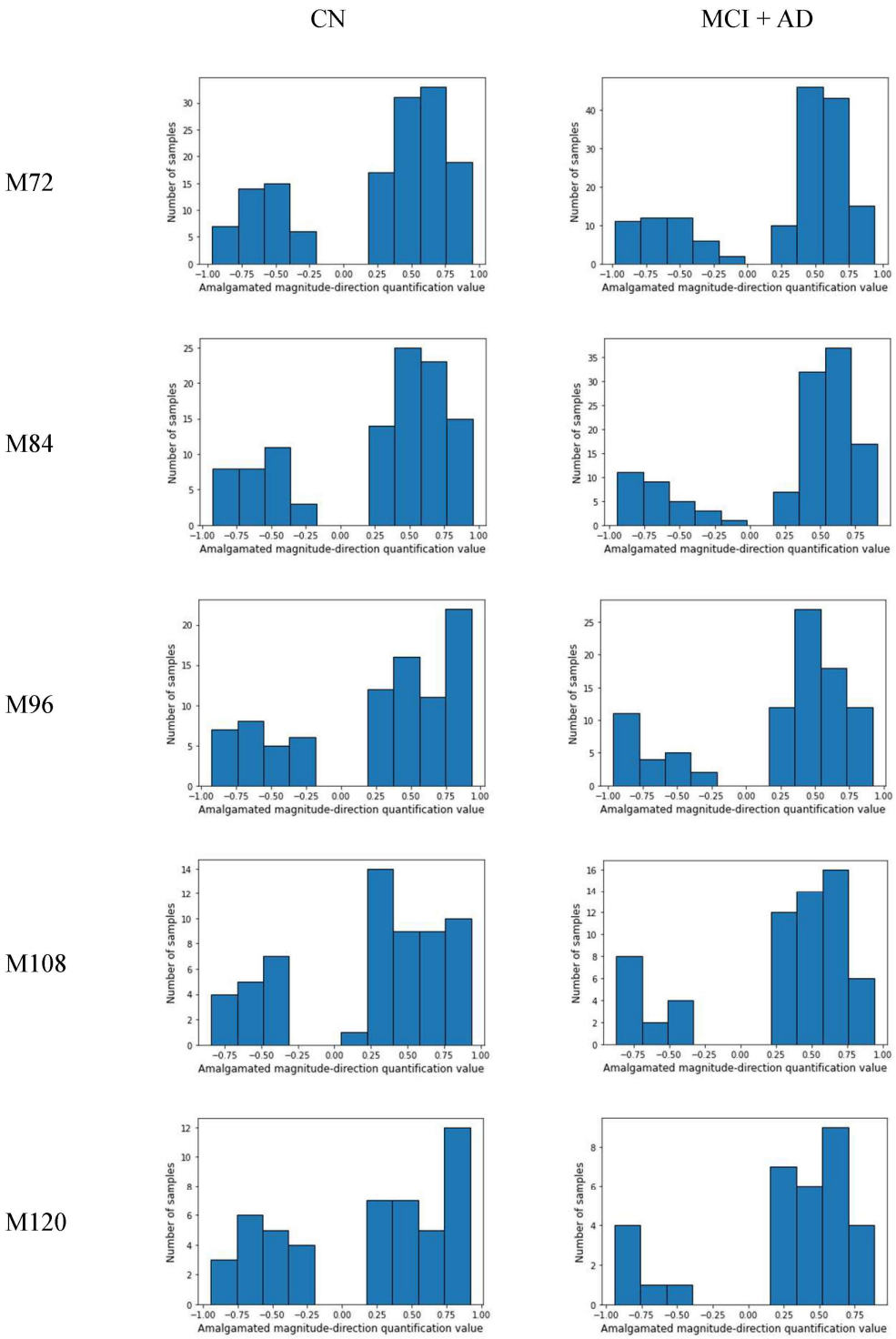


Figure E.10: Continuation of Figure E.9, differences in the distribution of early stage (BL-M06) AMDQ quantitative values for CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

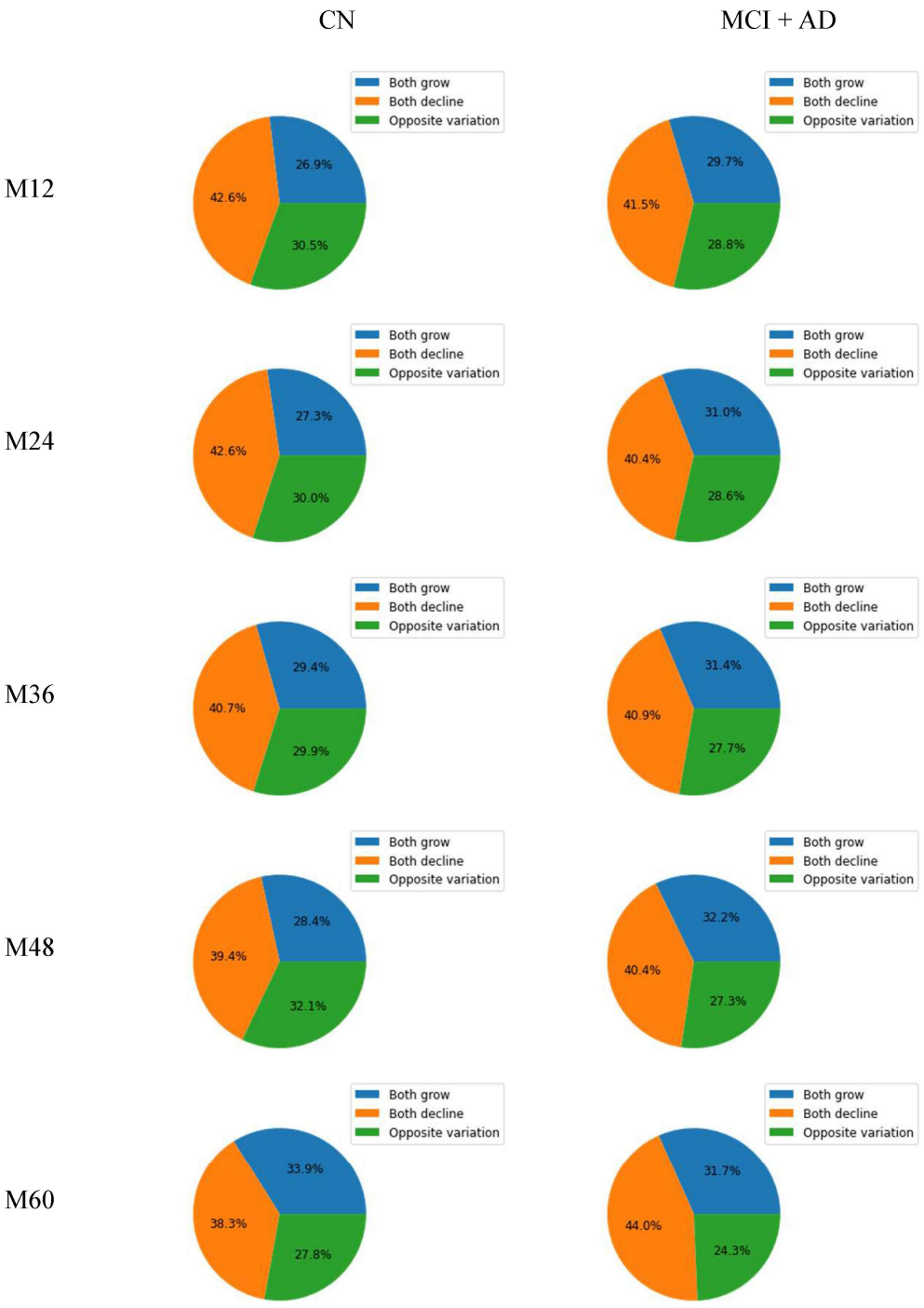


Figure E.11: Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

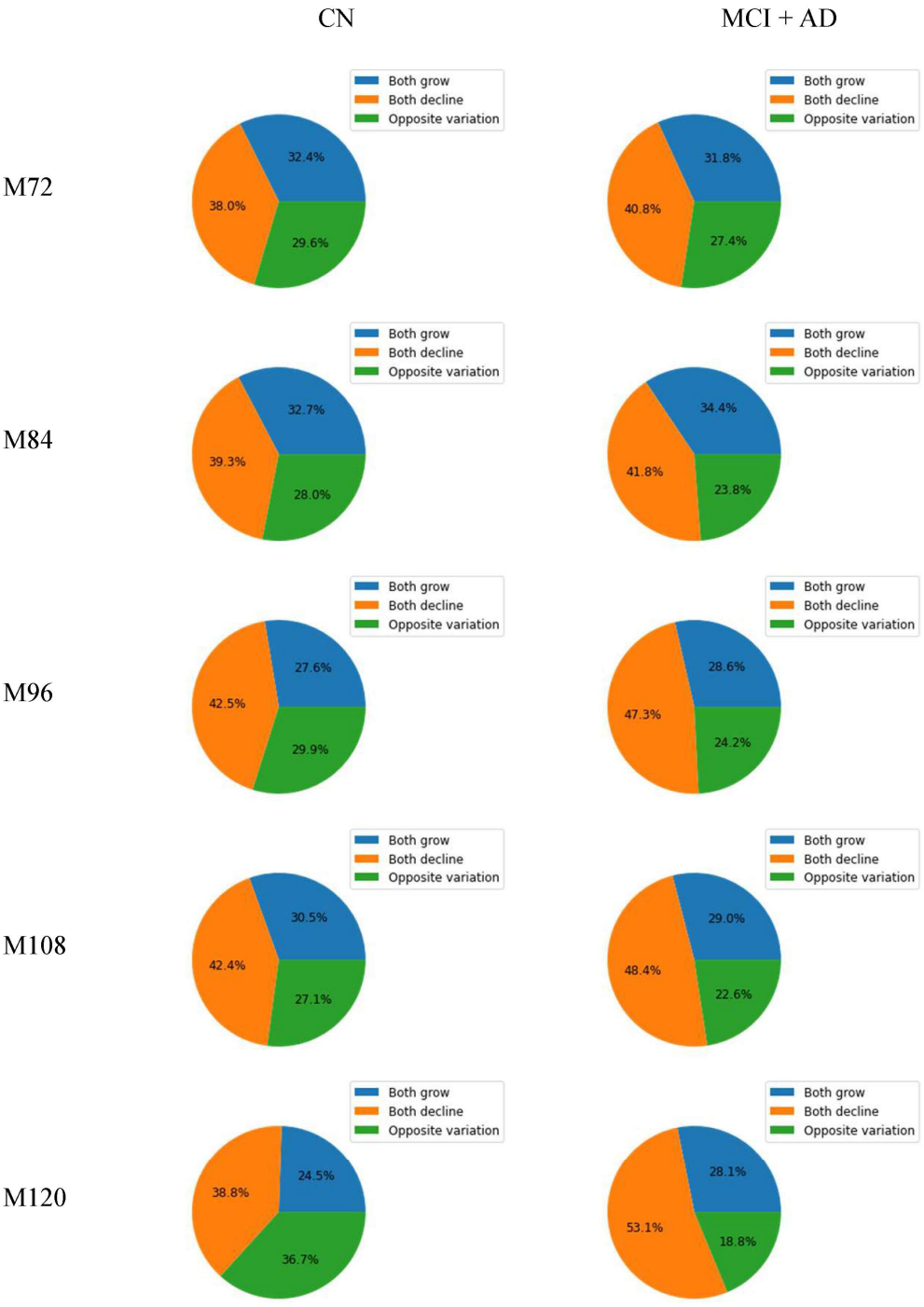


Figure E.12: Continuation of Figure E.11, Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for CTA. of L.Postcentral - Vol(C). of L.SuperiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

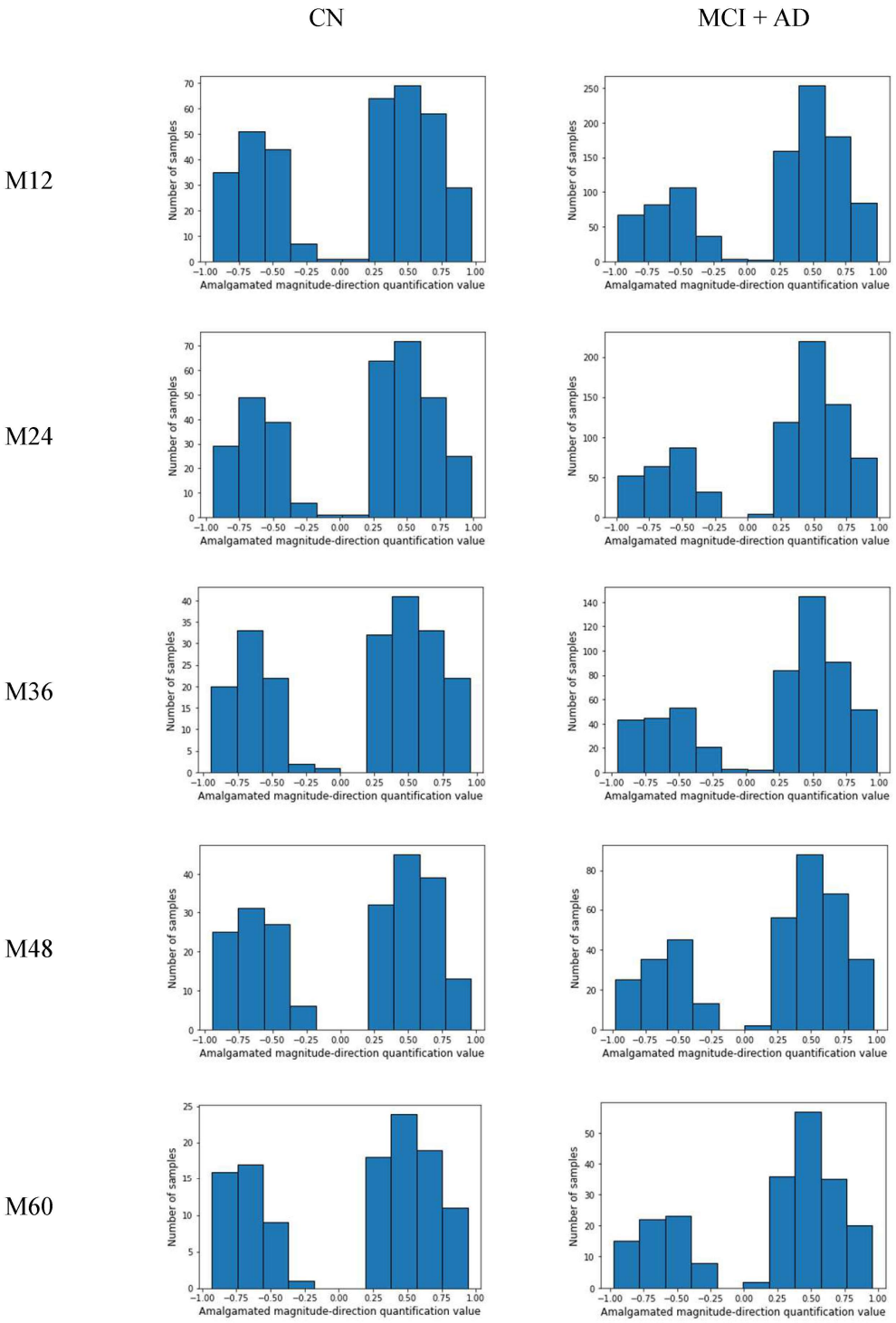


Figure E.13: Differences in the distribution of early stage (BL-M06) AMDQ quantitative values for CTA. of L.InferiorParietal - CTA. of L.Precentral correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

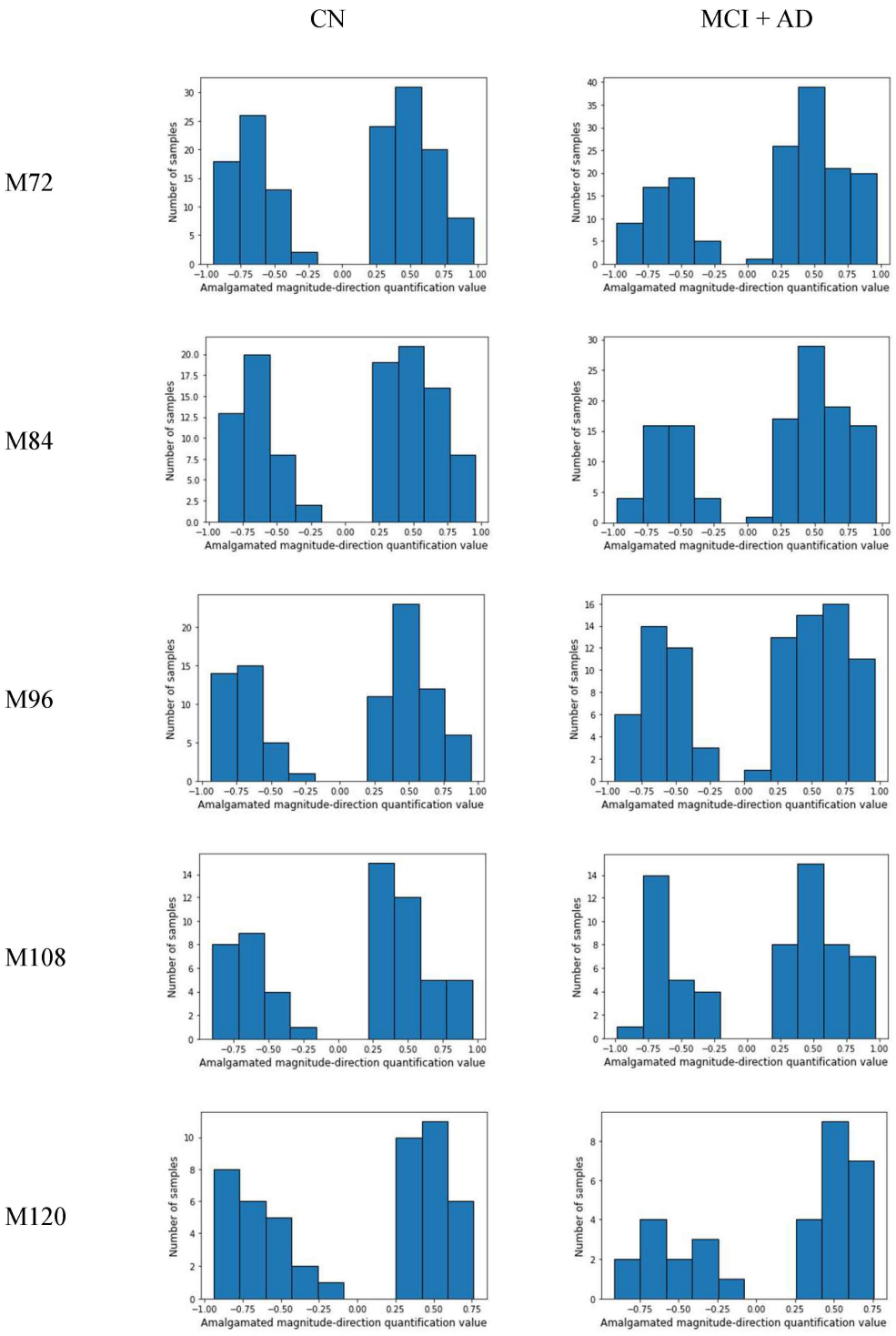


Figure E.14: Continuation of Figure E.13, differences in the distribution of early stage (BL-M06) AMDQ quantitative values for CTA. of L.InferiorParietal - CTA. of L.Precentral correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

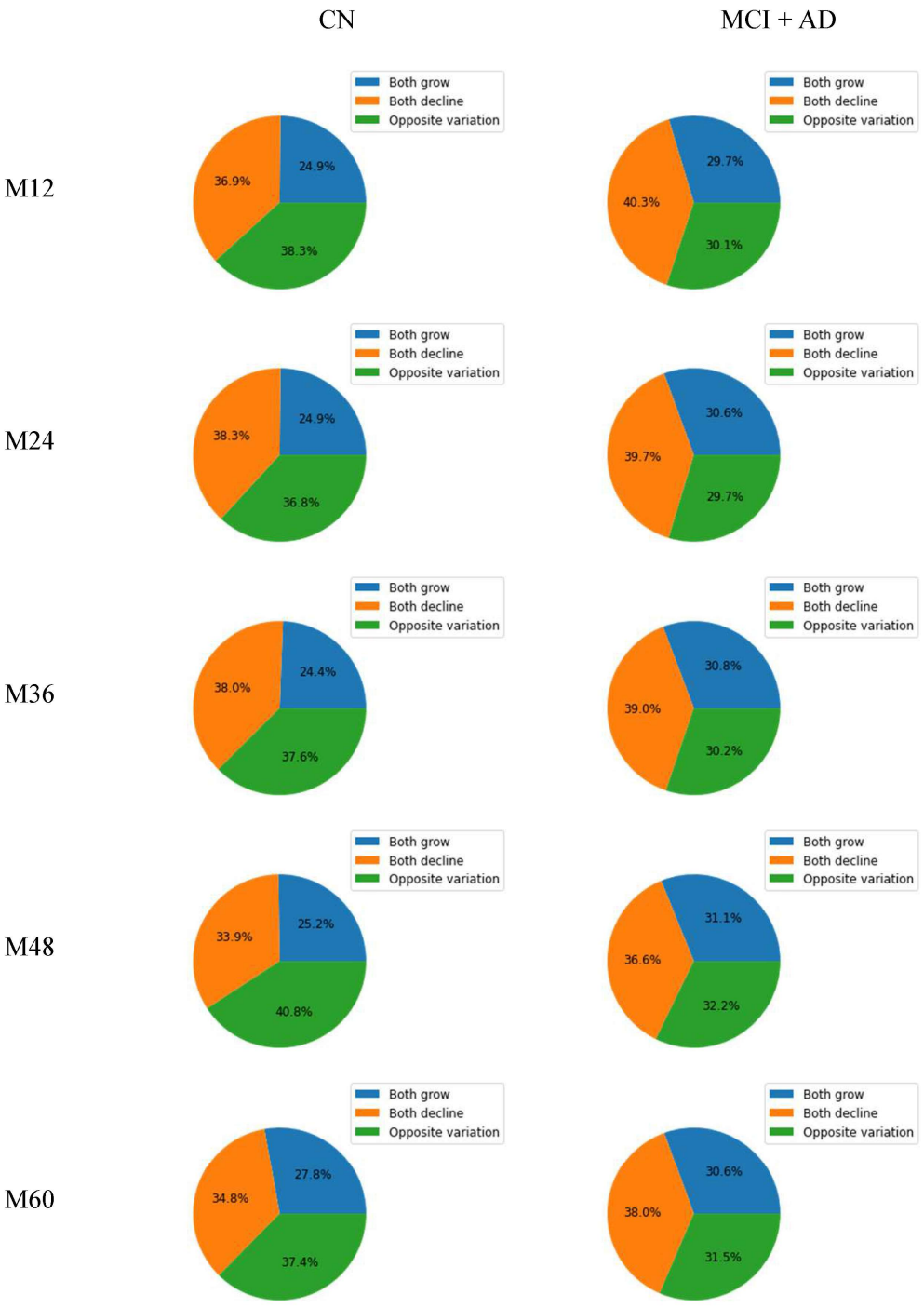


Figure E.15: Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for CTA. of L.InferiorParietal - CTA. of L.Precentral correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

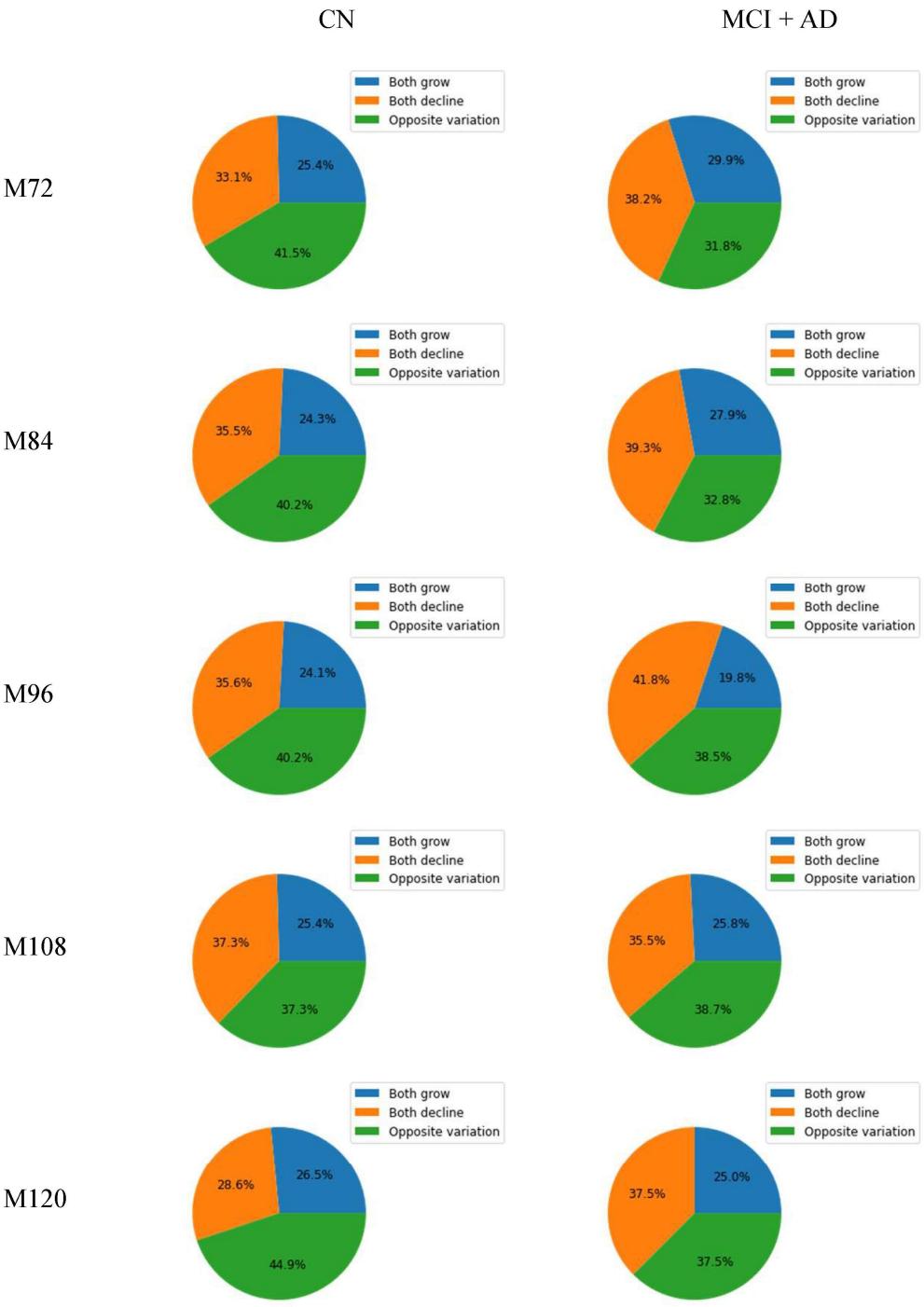


Figure E.16: Continuation of Figure E.15, Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for CTA. of L.InferiorParietal - CTA. of L.Precentral correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

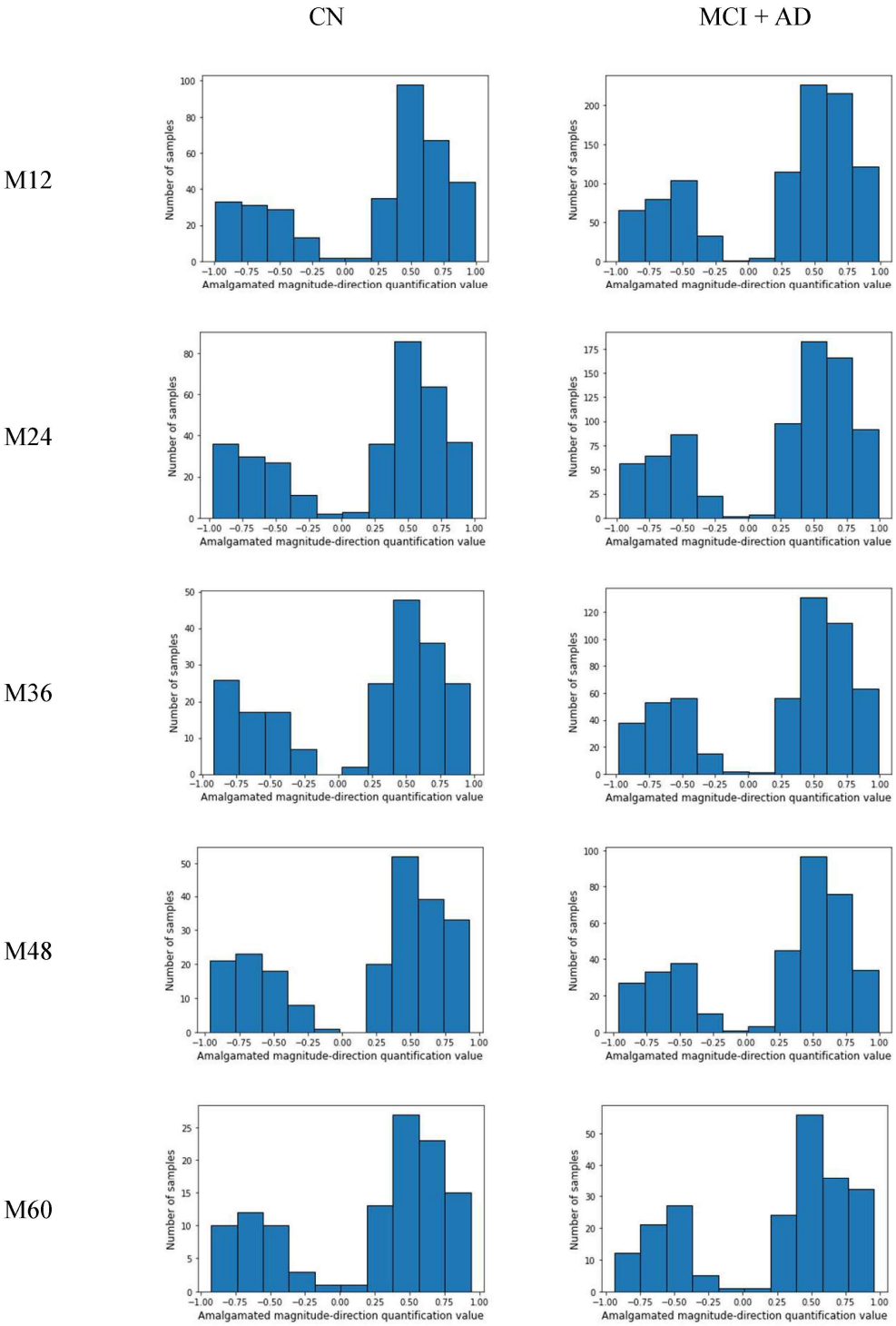


Figure E.17: Differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

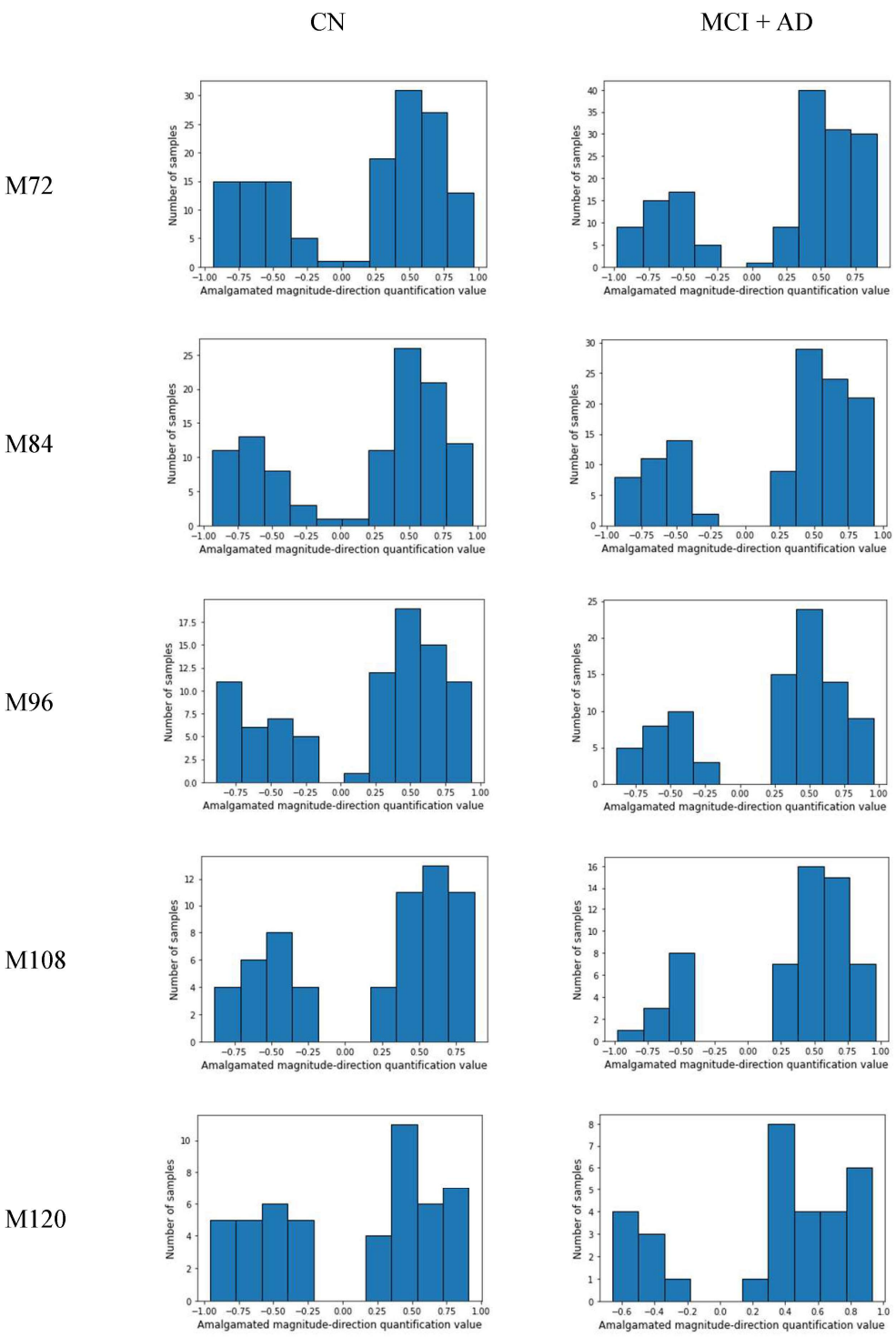


Figure E.18: Continuation of Figure E.17, differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

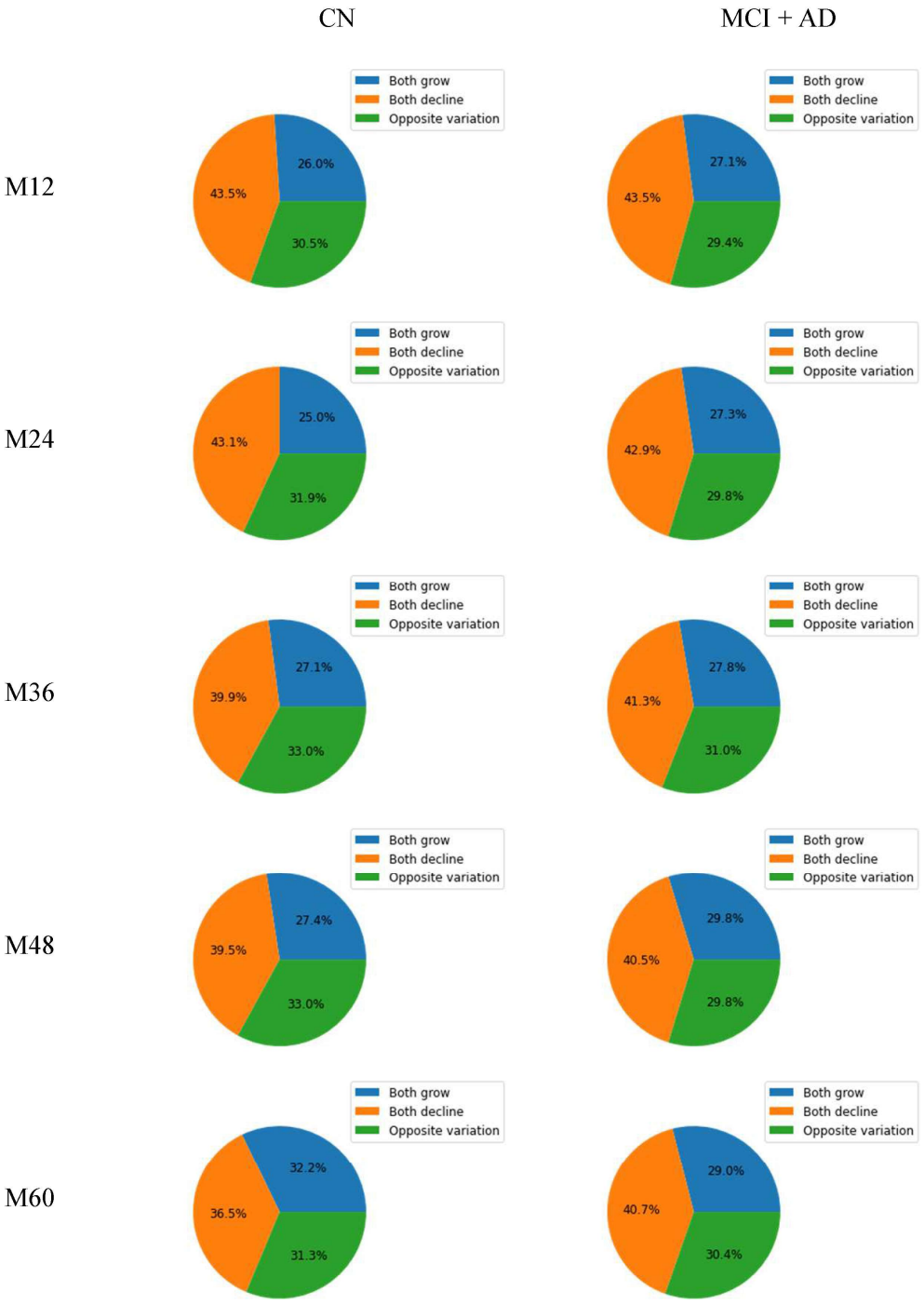


Figure E.19: Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

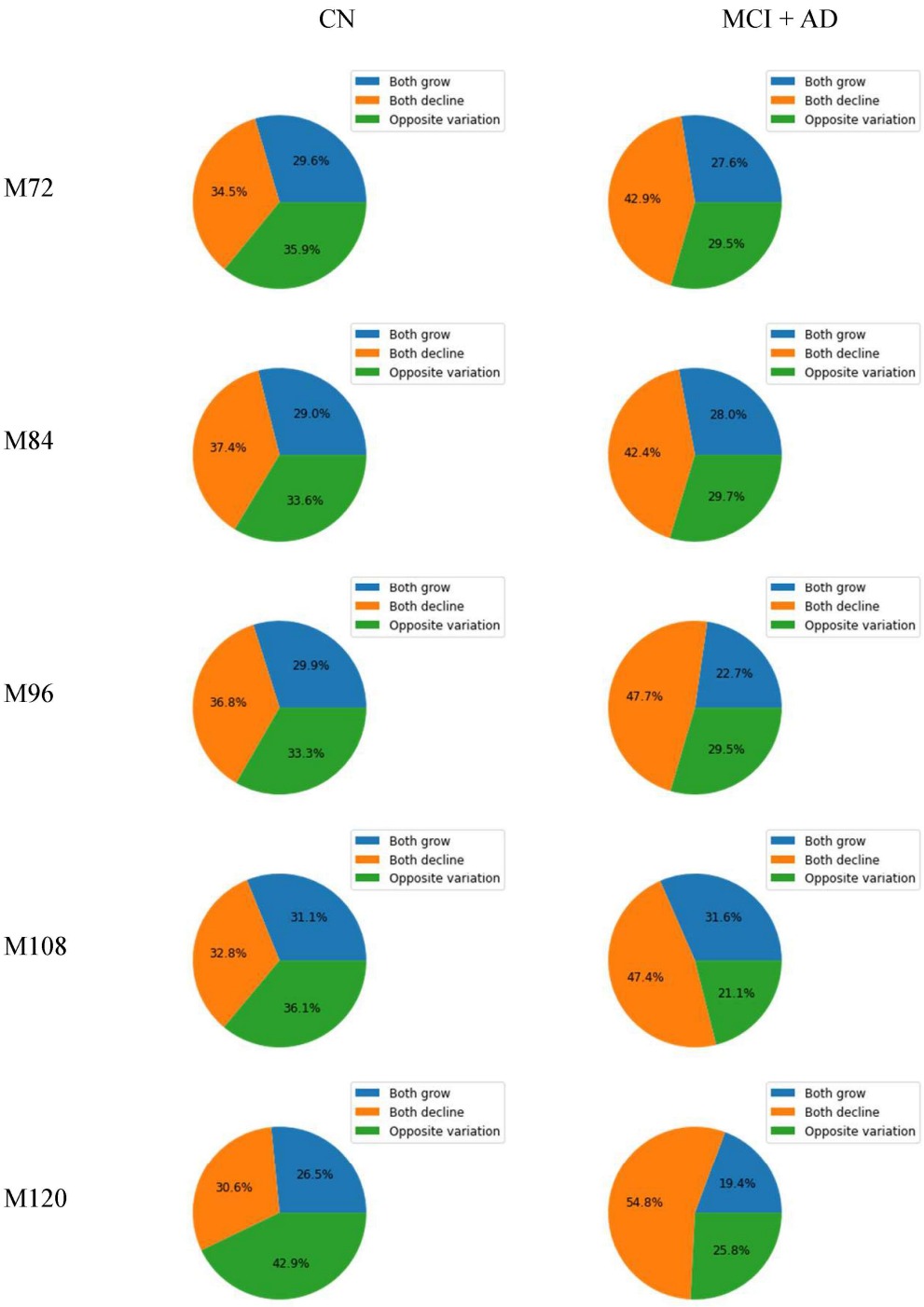


Figure E.20: Continuation of Figure E.19, Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of L.InferiorParietal - Vol(C). of L.SuperiorParietal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

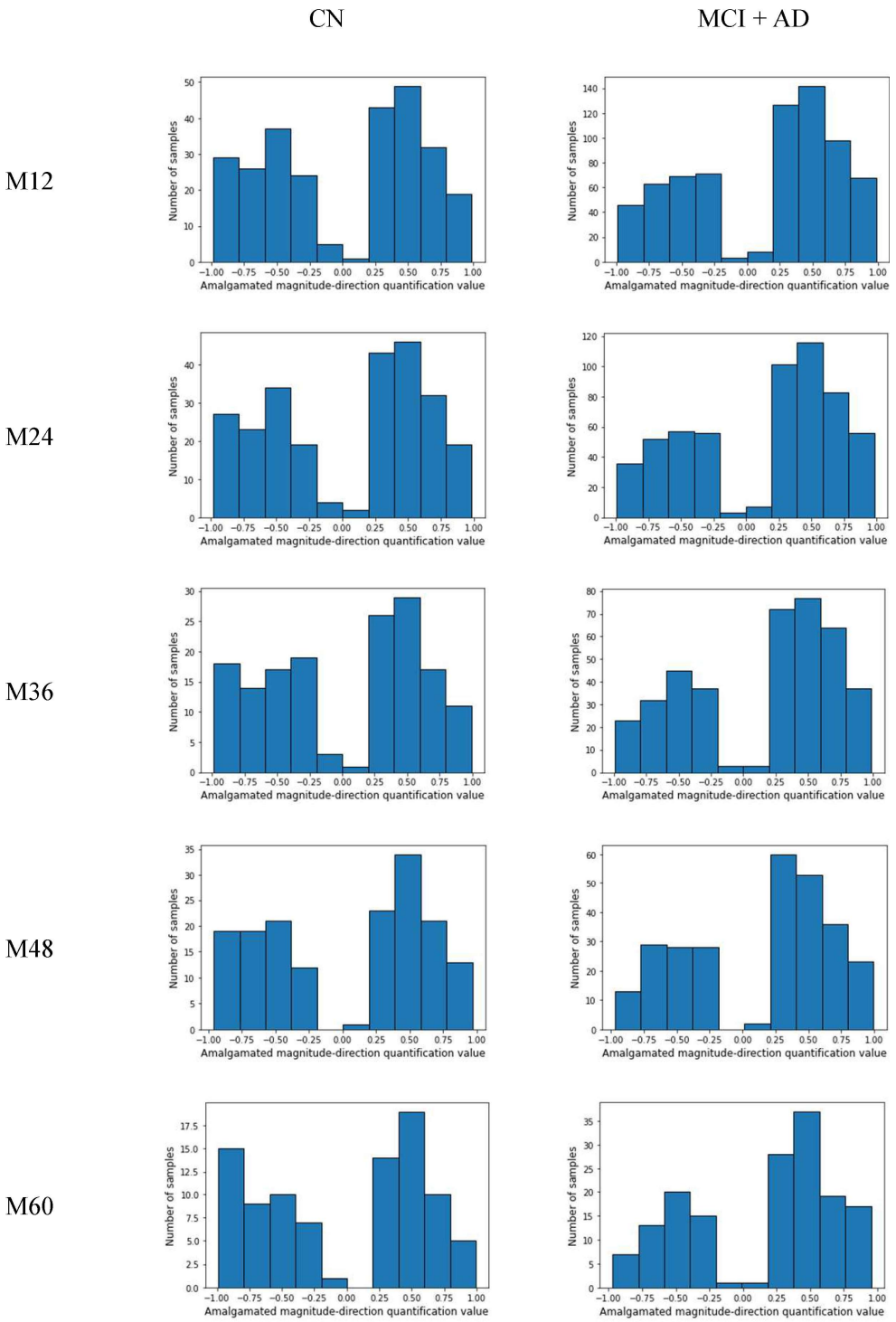


Figure E.21: Differences in the distribution of early stage (BL-M06) AMDQ quantitative values for CTA. of R.ParsTriangularis - CTA. of L.Postcentral correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

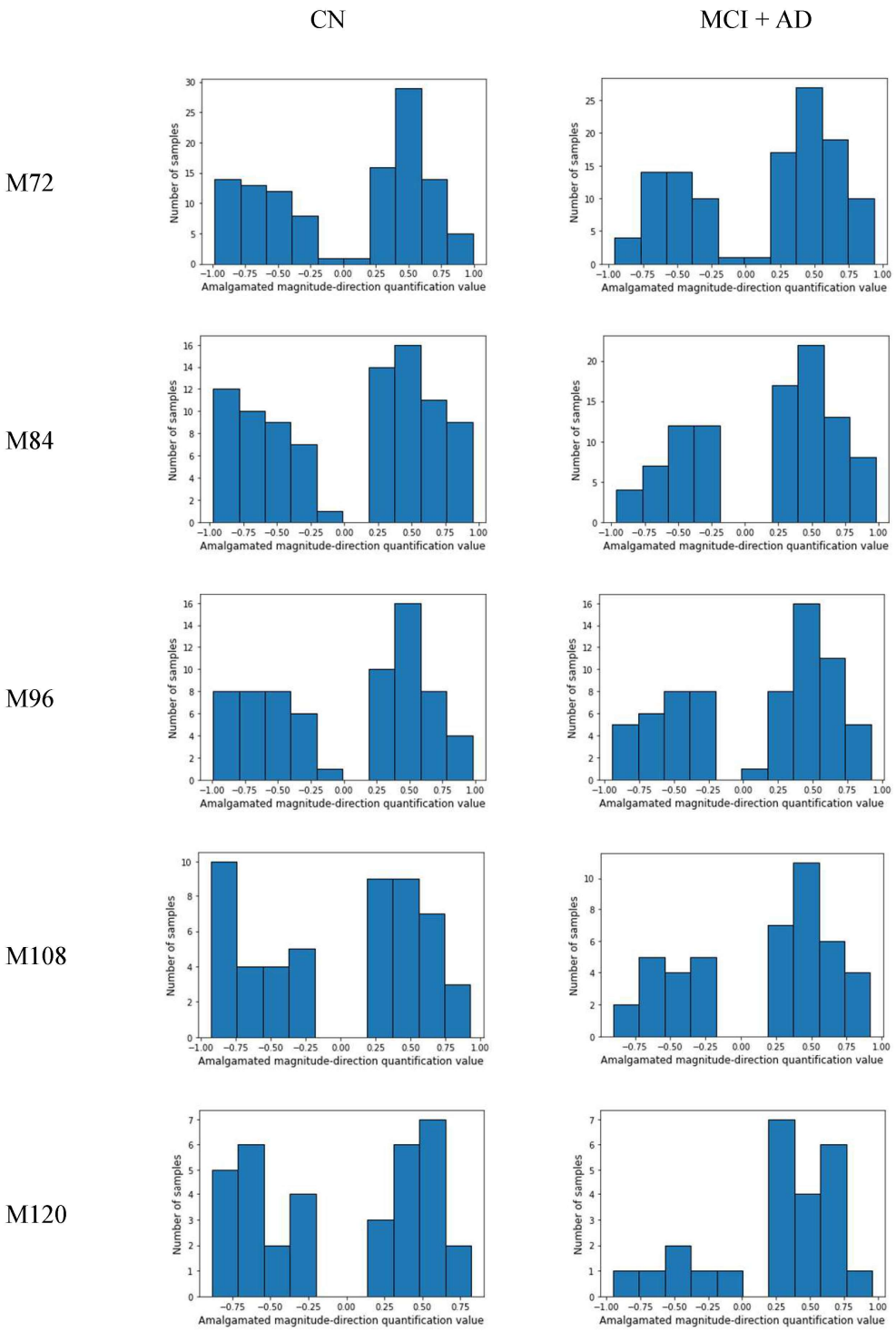


Figure E.22: Continuation of Figure E.21, differences in the distribution of early stage (BL-M06) AMDQ quantitative values for CTA. of R.ParsTriangularis - CTA. of L.Postcentral correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

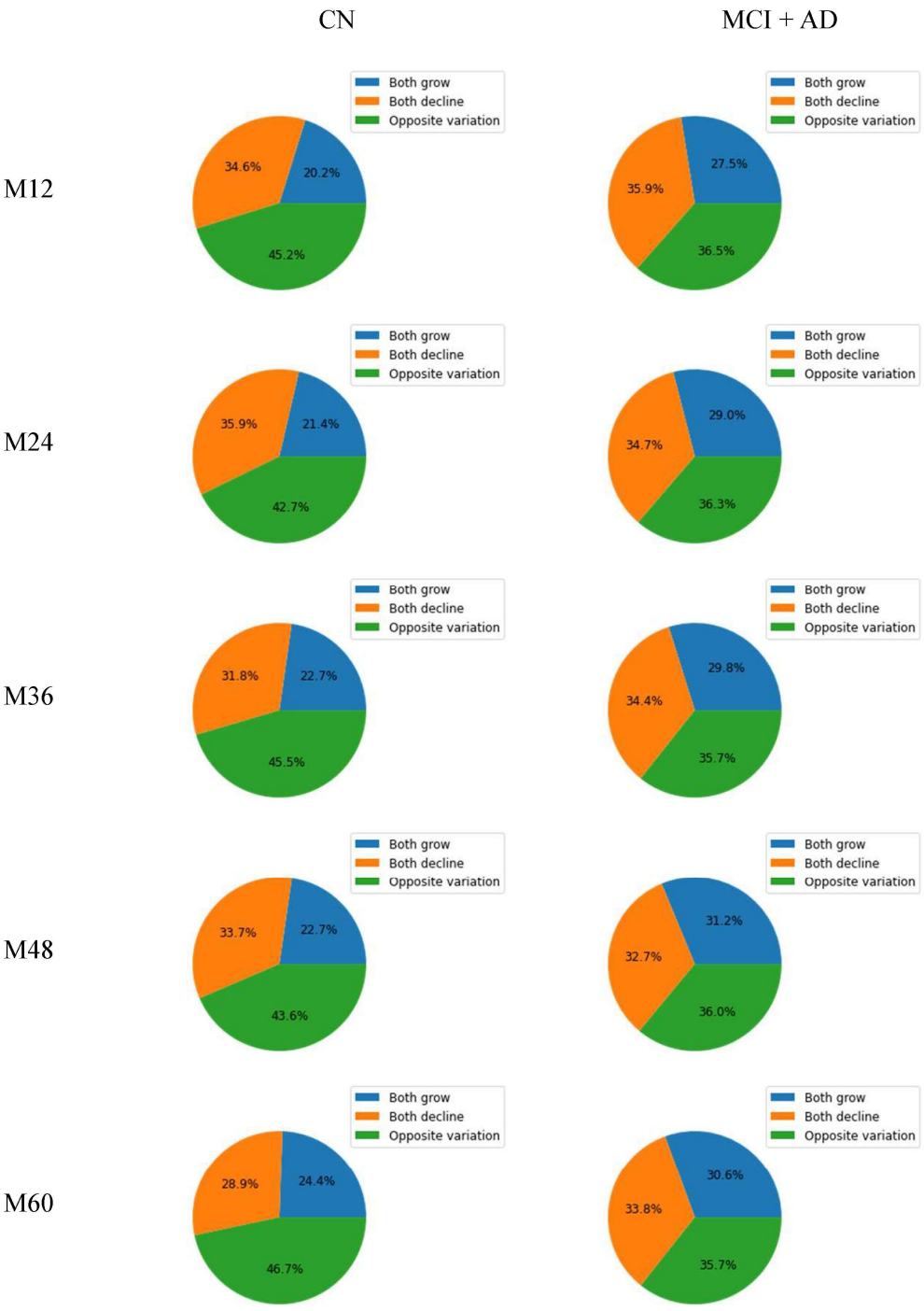


Figure E.23: Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for CTA. of R.ParsTriangularis - CTA. of L.Postcentral correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

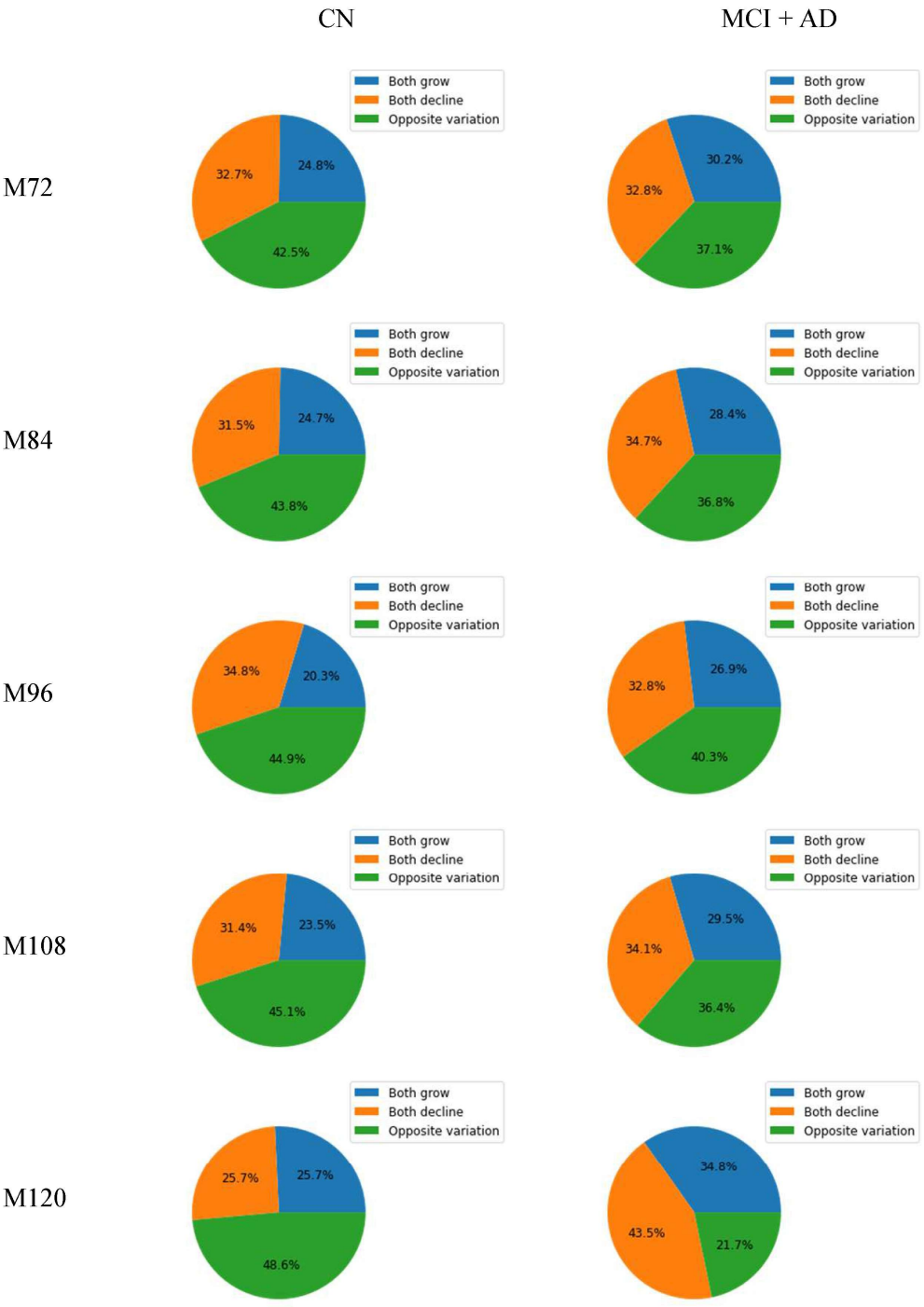


Figure E.24: Continuation of Figure E.23, Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for CTA. of R.ParsTriangularis - CTA. of L.Postcentral correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

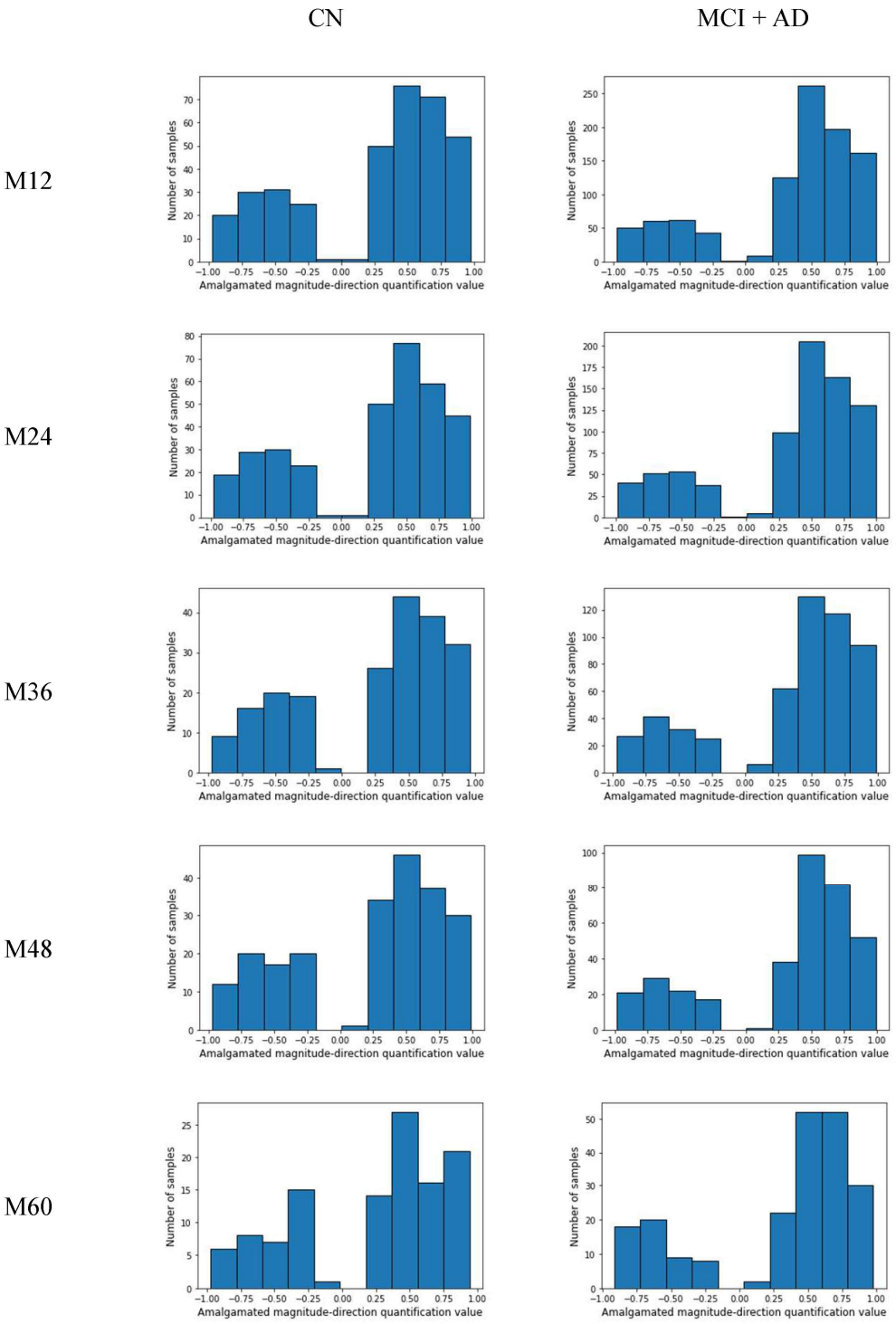


Figure E.25: Differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

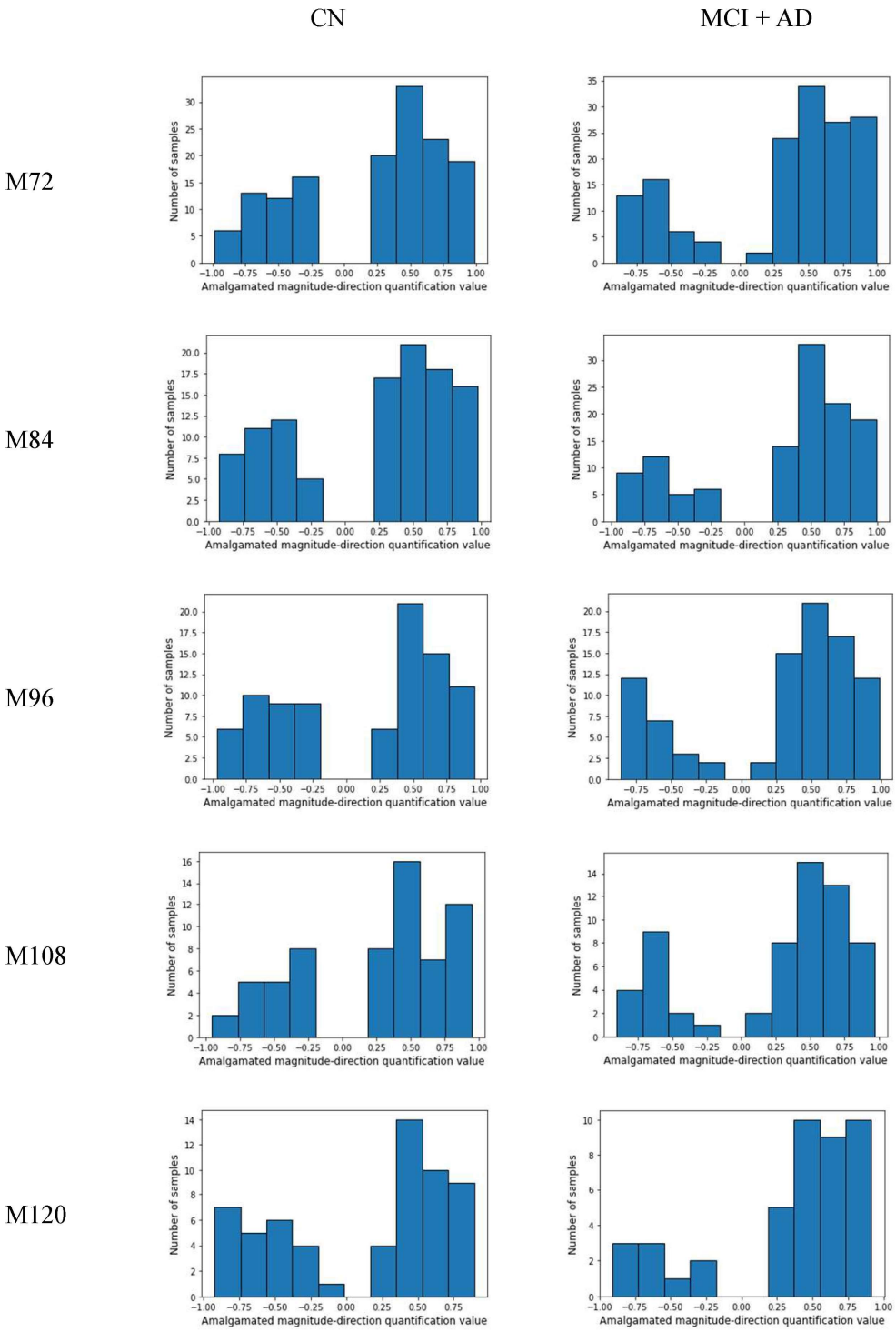


Figure E.26: Continuation of Figure E.25, differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

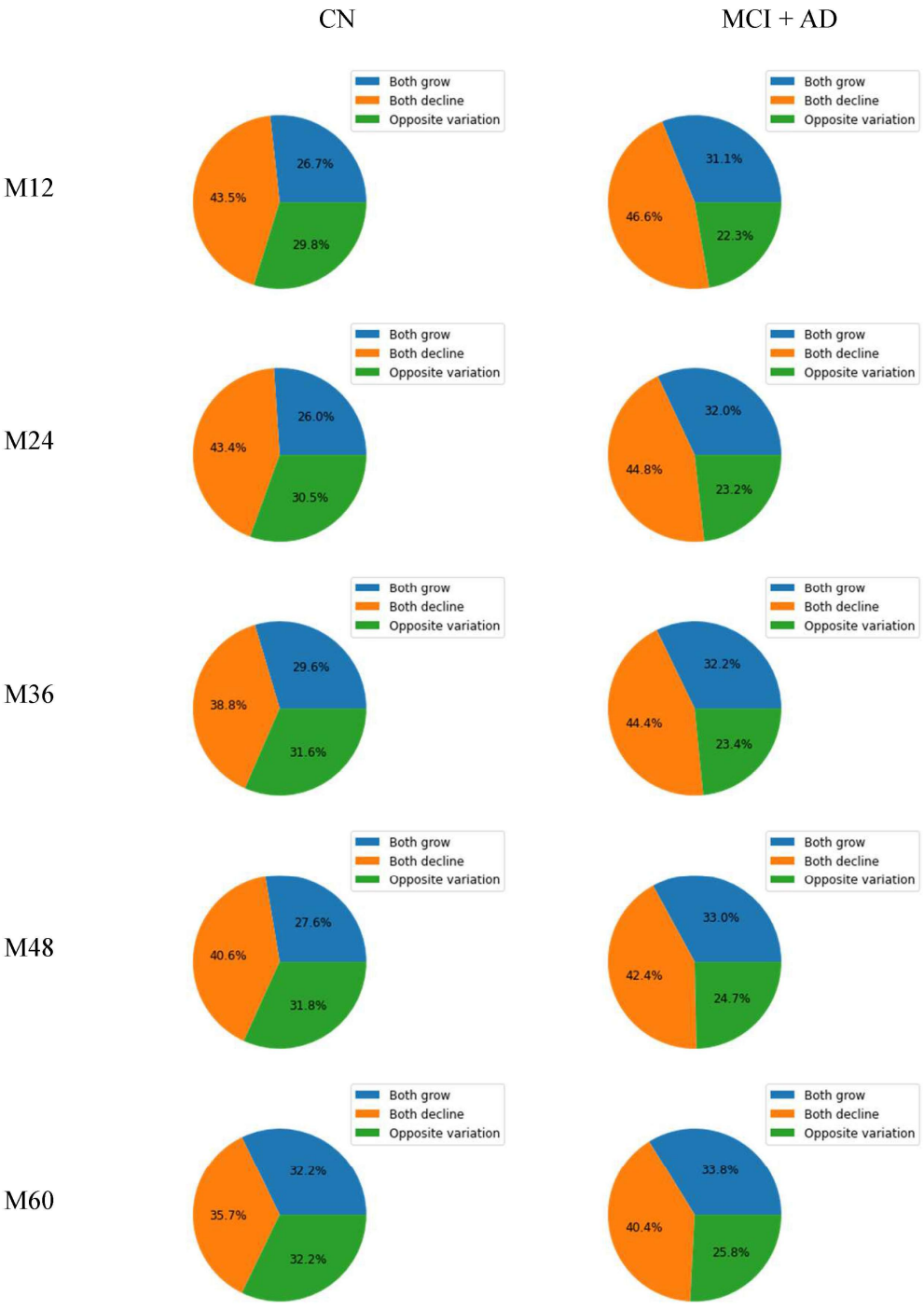


Figure E.27: Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

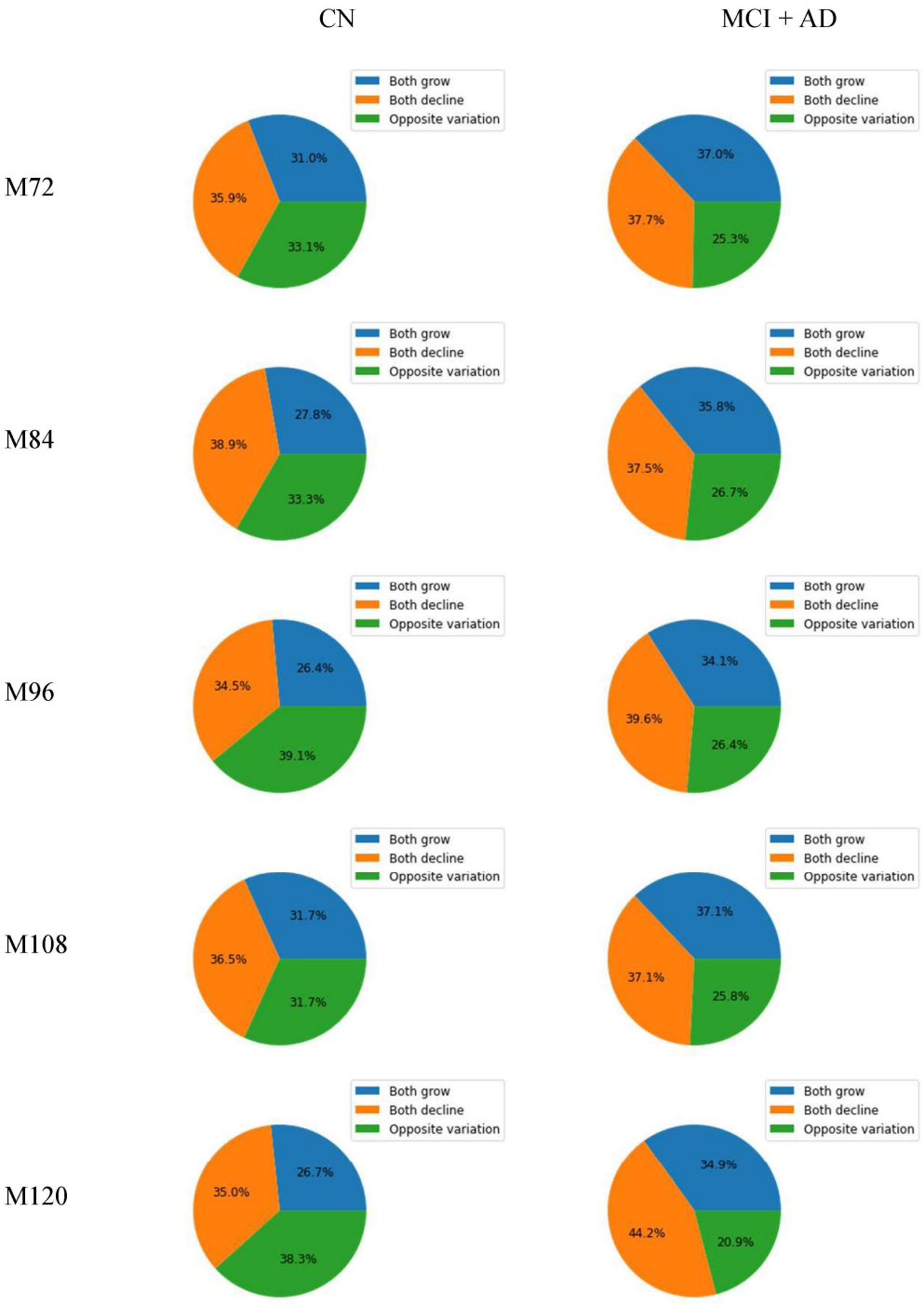


Figure E.28: Continuation of Figure E.27, Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of R.SuperiorFrontal - Vol(C). of L.SuperiorFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

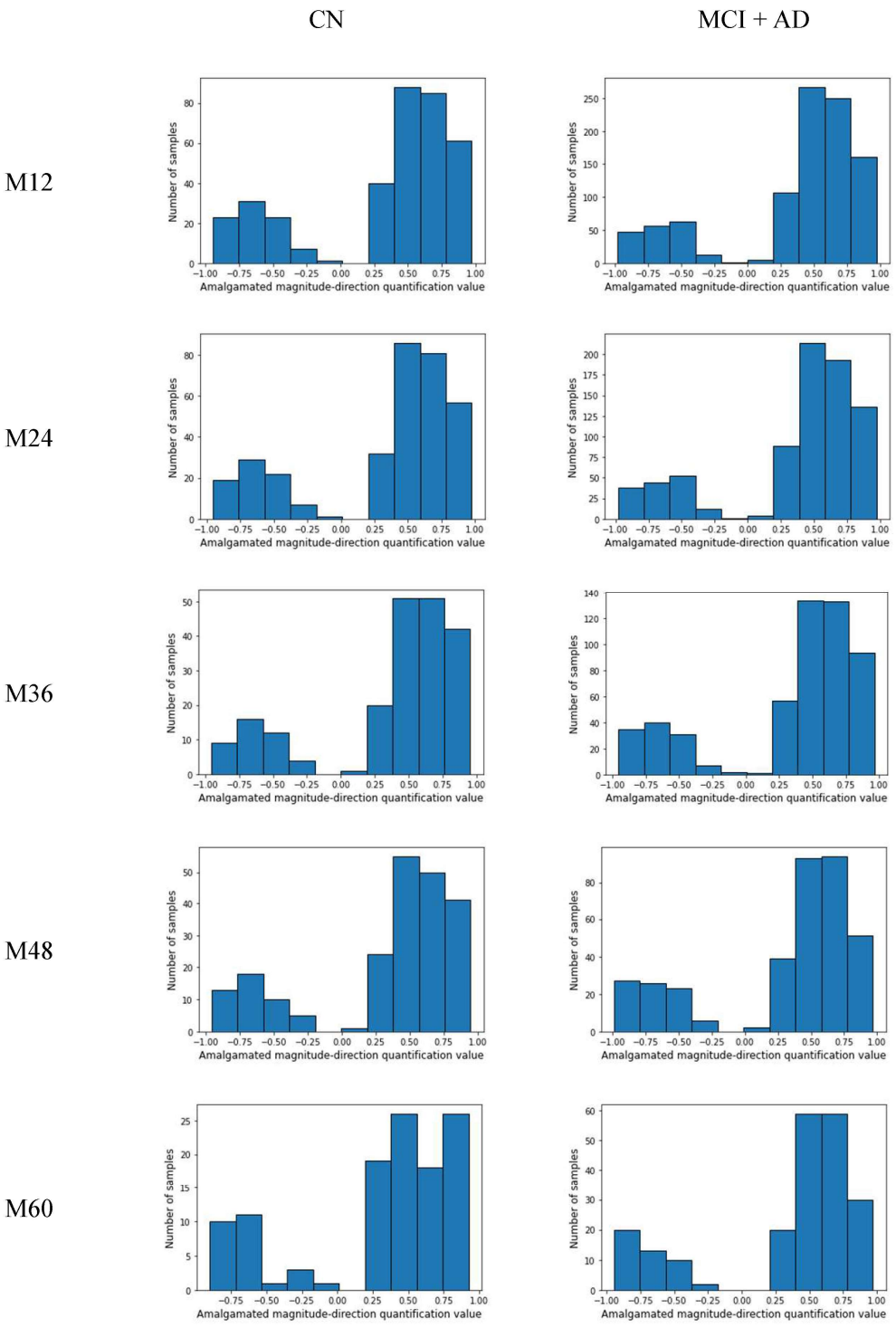


Figure E.29: Differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

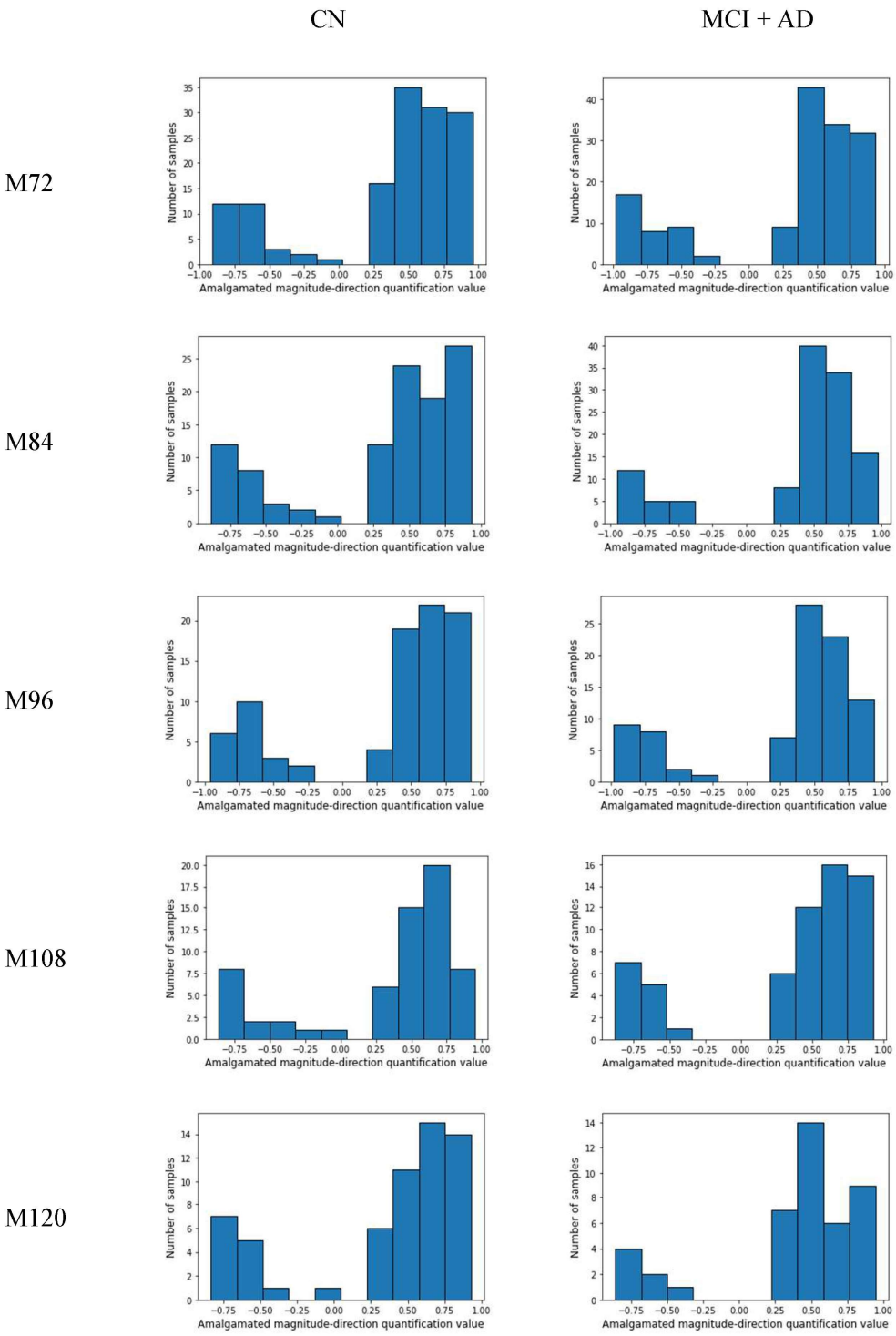


Figure E.30: Continuation of Figure E.29, differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

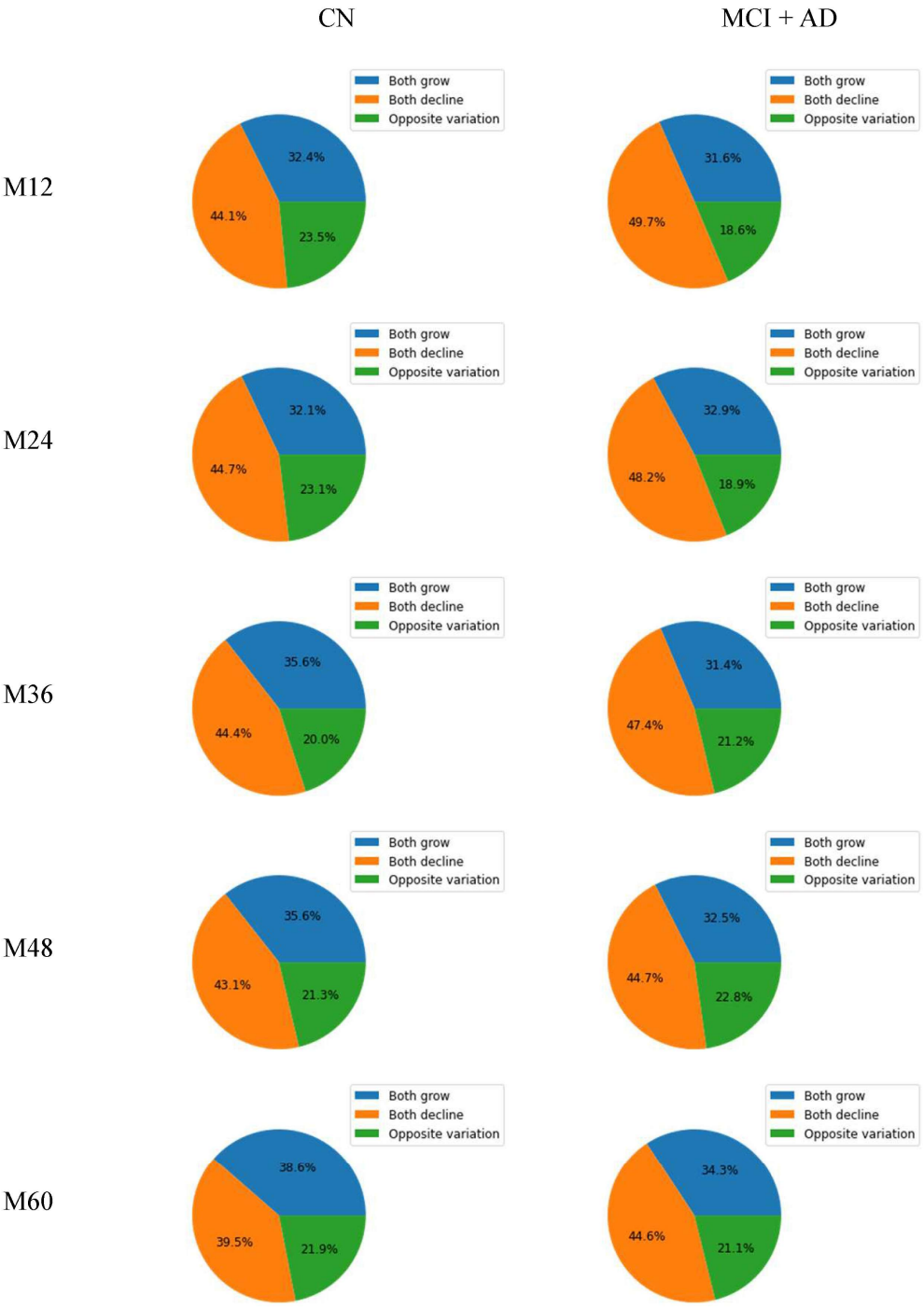


Figure E.31: Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

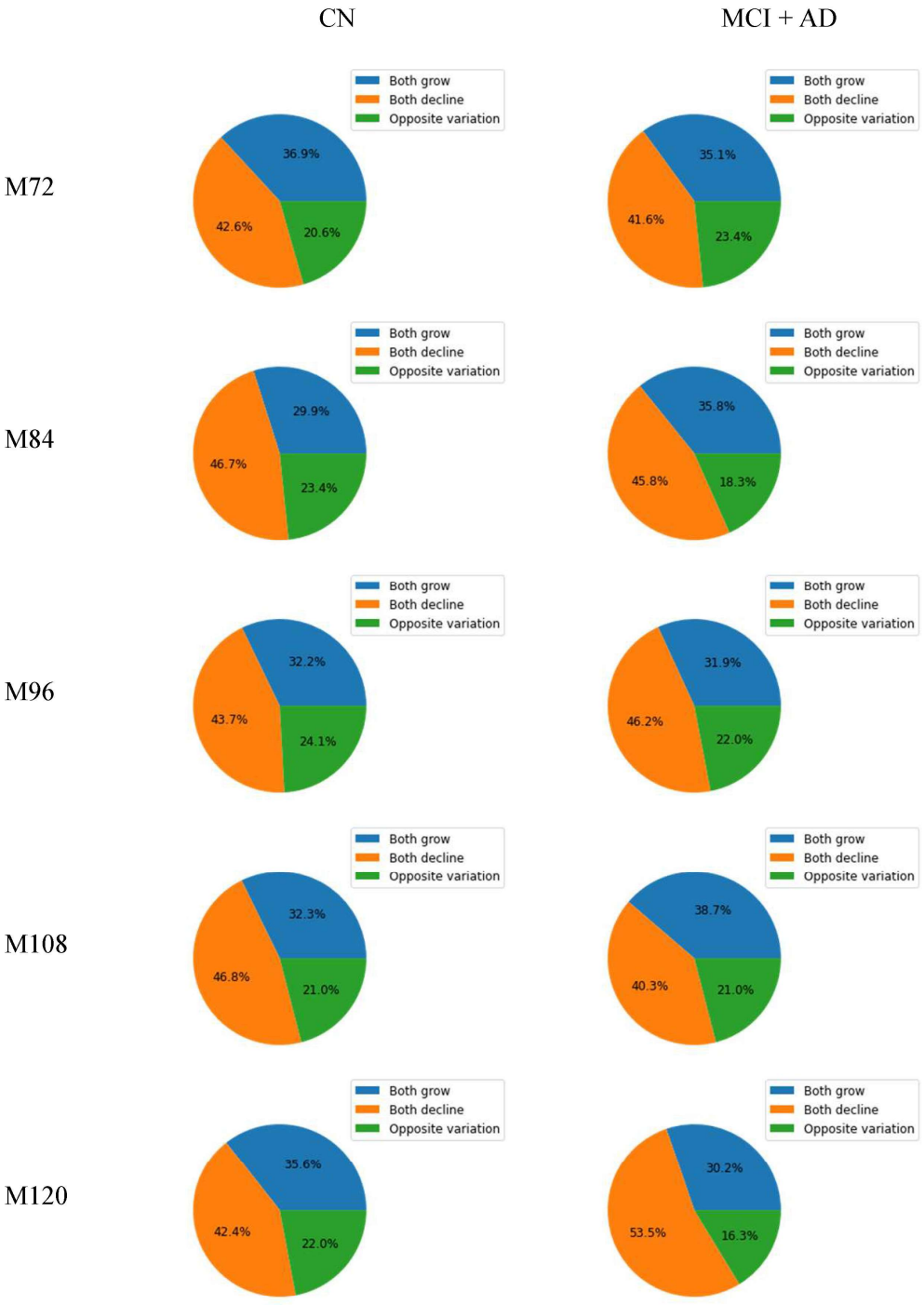


Figure E.32: Continuation of Figure E.31, Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of L.SuperiorFrontal - CTA. of L.SuperiorFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

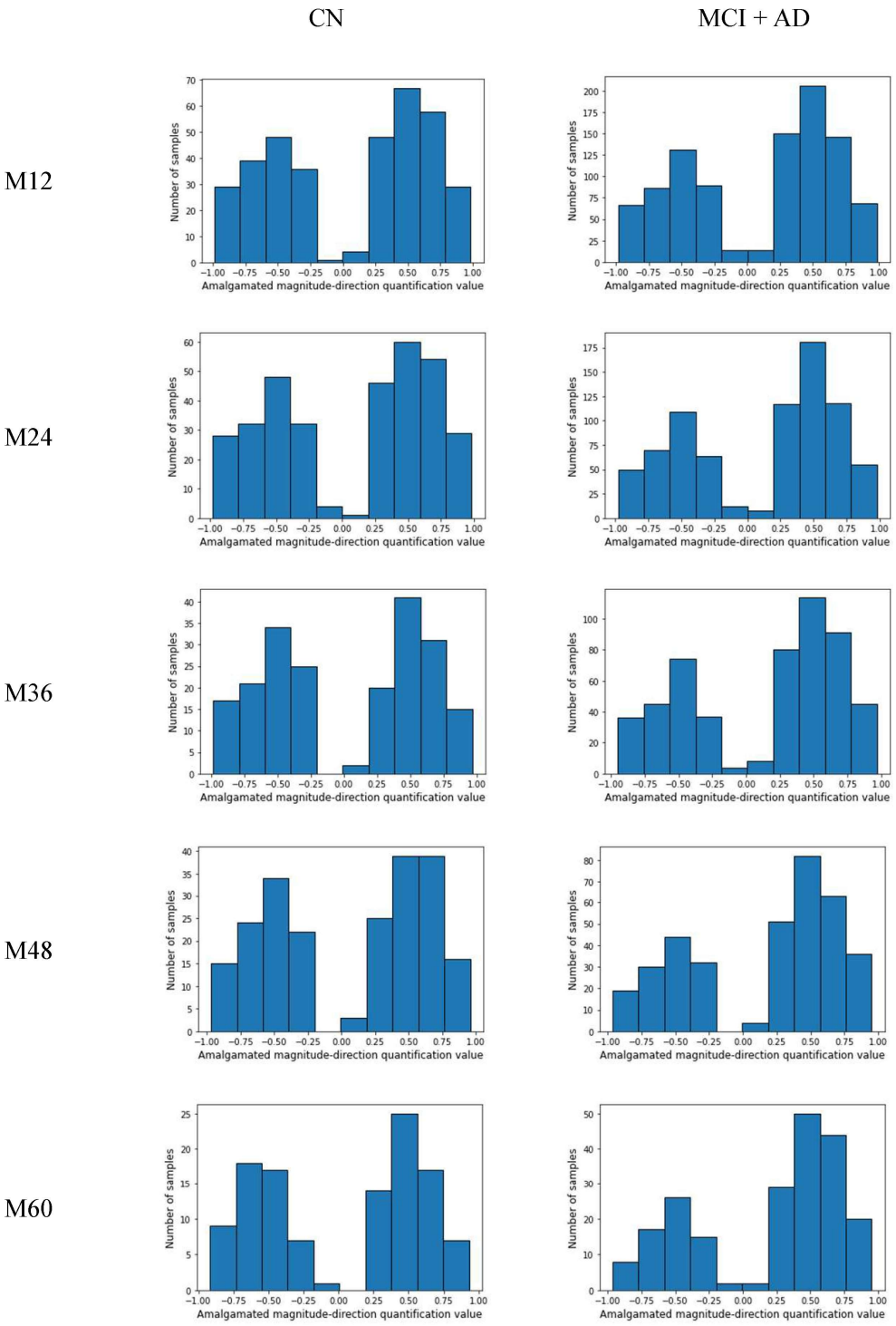


Figure E.33: Differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

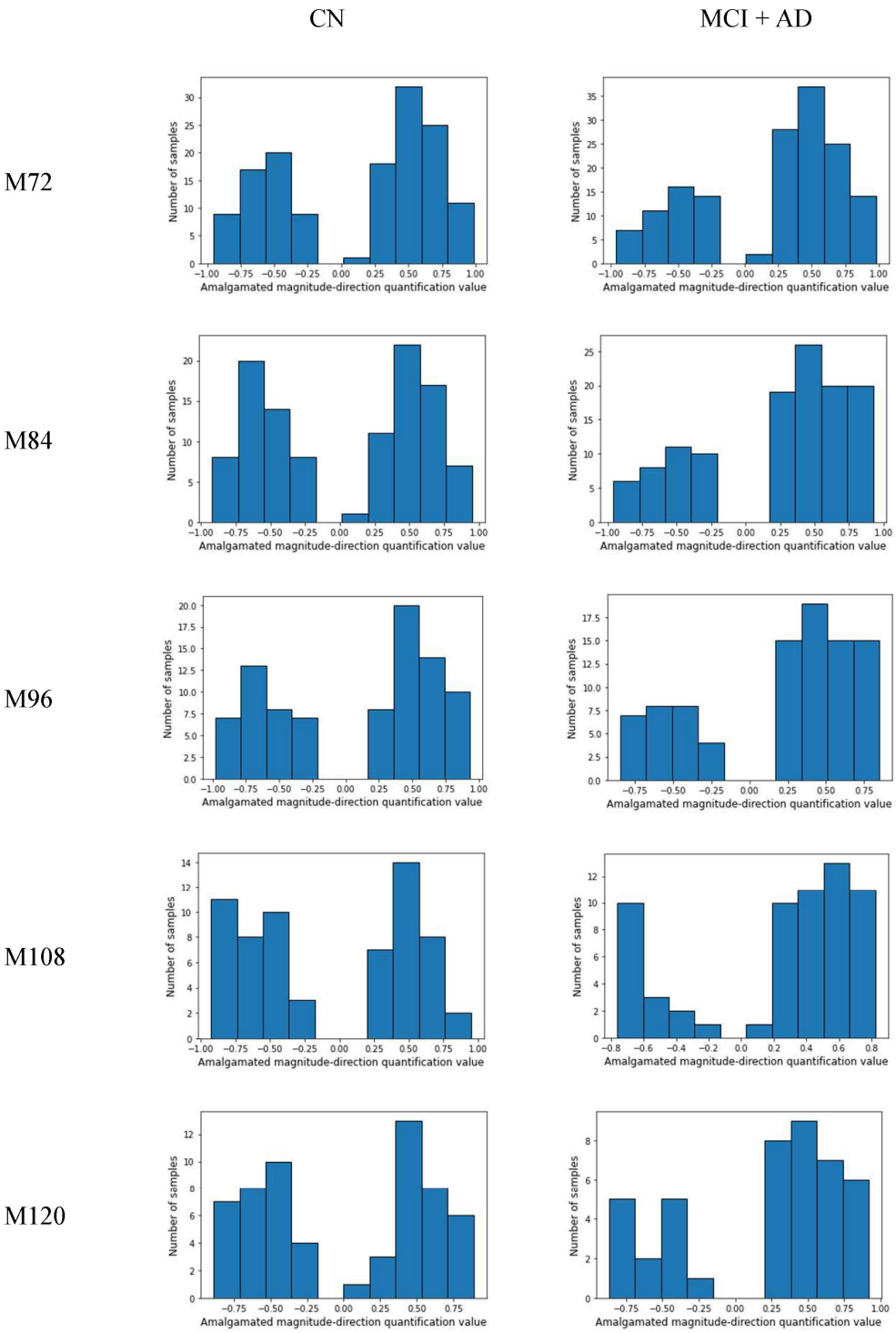


Figure E.34: Continuation of Figure E.33, differences in the distribution of early stage (BL-M06) AMDQ quantitative values for Vol(C). of Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.

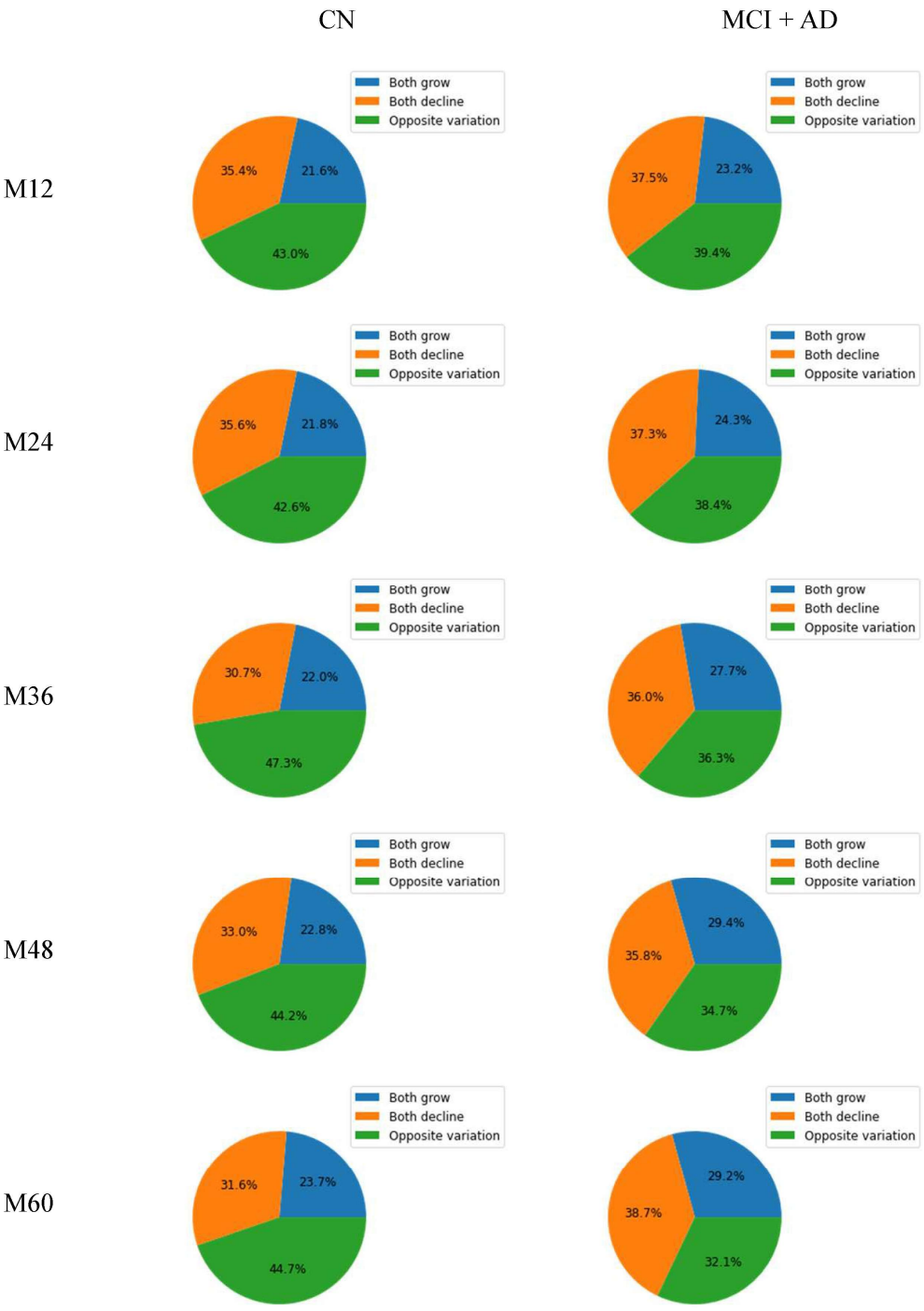


Figure E.35: Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M12 to M60.

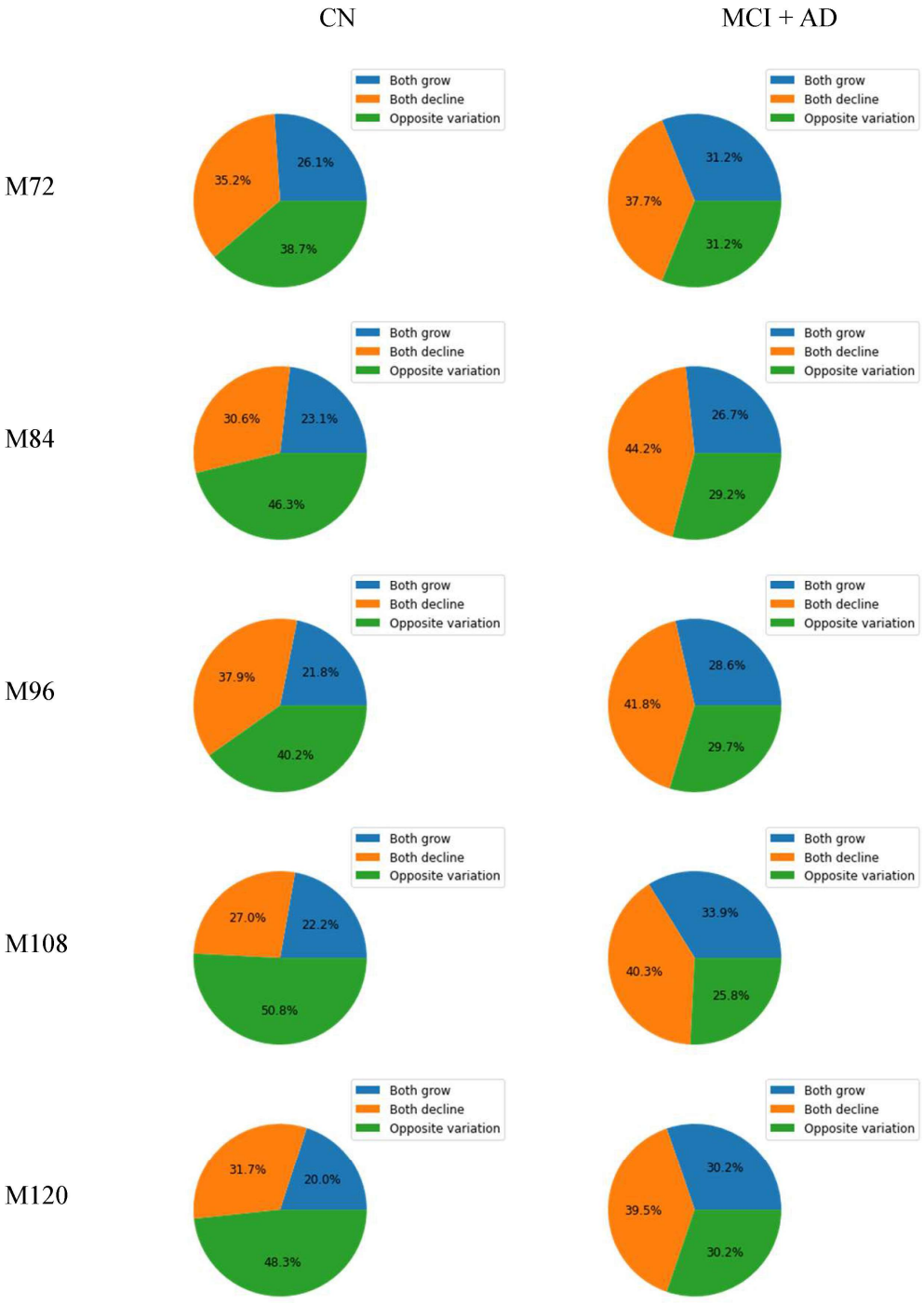


Figure E.36: Continuation of Figure E.35, Differences in the distribution of early stage (BL-M06) relative structure variation status between biomarkers for Vol(C). of R.Precuneus - CTA. of R.RostralMiddleFrontal correlations between cognitively impaired and non-cognitively impaired individuals at time points M72 to M120.