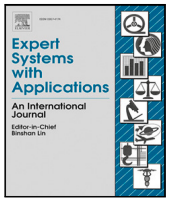
Expert Systems With Applications 231 (2023) 120761

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Multi-view imputation and cross-attention network based on incomplete longitudinal and multimodal data for conversion prediction of mild cognitive impairment

Tao Wang [a], Xiumei Chen [a], Xiaoling Zhang [a], Shuoling Zhou [a], Qianjin Feng [a,b,c,*], Meiyan Huang [a,b,c,*]

[a] School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[b] Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou 510515, China
[c] Guangdong Province Engineering Laboratory for Medical Imaging and Diagnostic Technology, Southern Medical University, Guangzhou 510515, China

ARTICLE INFO

Keywords:
Adversarial learning
Cross-attention
Conversion prediction
Longitudinal and multimodal data
Mild cognitive impairment
Multi-view imputation

ABSTRACT

Predicting whether subjects with mild cognitive impairment (MCI) can convert to Alzheimer's disease is significant for personalized treatment development and disease progression delay. Longitudinal and multimodal data have been recognized for their ability to capture longitudinal variations and provide complementary information for MCI conversion prediction. However, incomplete or missing data pose a persistent challenge in effectively utilizing such valuable information. Additionally, early-stage conversion prediction, particularly at baseline visit (BL), is crucial in clinical practice. Therefore, longitudinal data must only be incorporated during training to capture disease progression information. To address these challenges, we propose a multi-view imputation and cross-attention network (MCNet) to integrate data imputation and MCI conversion prediction in a unified framework. First, a multi-view imputation method combined with adversarial learning is presented to handle various missing data scenarios and reduce imputation errors. Second, two cross-attention blocks are introduced to exploit the potential associations in longitudinal and multimodal data. Finally, a multi-task learning model is established for data imputation, longitudinal classification, and conversion prediction. By appropriately training the model, disease progression information learned from longitudinal data improves the MCI conversion prediction that only uses BL data. To verify its effectiveness and flexibility in such MCI conversion prediction, we test MCNet on independent testing sets and single-modal BL data. Results show that MCNet outperforms competitive methods with an area under the receiver operating characteristic curve value of 86.0%. Furthermore, the interpretability of MCNet is demonstrated, indicating its potential as a valuable tool for incomplete data analysis in MCI conversion prediction.

1. Introduction

Alzheimer's disease (AD) is characterized by the irreversible impairment of cognitive functions and is one of the most common neurodegenerative diseases among the elderly people (Gaugler et al., 2022). Mild cognitive impairment (MCI) is commonly regarded as a prodromal stage of AD and a critical period for early diagnosis of this disease (Scheltens et al., 2021). As reported, approximately 9.6% of subjects who have MCI are expected to progress to AD annually, whereas other MCI subjects maintain a stable clinical condition with time (Abdelnour et al., 2022; Ganguli et al., 2019). Therefore, the accurate identification of MCI subjects who may progress to AD is crucial in providing support for delaying the disease progression and developing new clinical therapies (Abdelnour et al., 2022). Generally, based on the criteria of whether MCI subjects convert to AD within three years, subjects can be classified into stable MCI (sMCI) and progressive MCI (pMCI) groups (Arco et al., 2021; Bucholc, Titarenko, Ding, Canavan, & Chen, 2023). Such classification is denoted as MCI conversion prediction, which is the focus in this study. In clinical practice, longitudinal and multimodal data are increasing and have attracted our attention (El-Sappagh, Abuhmed, Islam, & Kwak, 2020).

On the one hand, the integration of longitudinal and multimodal data provides comprehensive information for early AD diagnosis and MCI conversion prediction, surpassing the utilization of multimodal or longitudinal data alone (Jung et al., 2021; Lee, Kang, Nho, Sohn, & Kim,

* Corresponding authors at: School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China.
E-mail addresses: wangtao_9802@sina.com (T. Wang), chenxiumei97@163.com (X. Chen), zhangxiaoling9911@163.com (X. Zhang), zslandsouling@163.com (S. Zhou), fengqj99@smu.edu.cn (Q. Feng), huangmeiyan16@163.com (M. Huang).

https://doi.org/10.1016/j.eswa.2023.120761
Received 24 March 2023; Received in revised form 26 May 2023; Accepted 6 June 2023
Available online 10 June 2023

2019). Multimodal data, such as magnetic resonance imaging (MRI) and positron emission tomography (PET) images, can provide complementary structural and functional information (Arco et al., 2021; Chen et al., 2022). Given that AD is a progressive disease, early pathological changes and structural abnormalities over time within longitudinal variations can be captured by longitudinal data (Bucholc et al., 2023; Huang, Lai, et al., 2021; Zhang, Wu, et al., 2021). Nevertheless, research has primarily focused on using multimodal neuroimages at baseline visit (BL) for early AD diagnosis and MCI conversion prediction (Kikuchi et al., 2022; Ko, Jung, Jeon, & Suk, 2022; Zhu, Sun, Huang, Han, & Zhang, 2021). Moreover, most studies used only a simple concatenation of multimodal features, which may bring redundant information and fail to exploit potential associations among different modalities (El-Sappagh et al., 2020; Venugopalan, Tong, Hassanzadeh, & Wang, 2021; Zhou, Thung, Zhu, & Shen, 2019). Although several strategies for integrating multimodal data have been proposed (Arco et al., 2021; Luo et al., 2023), fusion methods that jointly analyze and cohesively combine multimodal and longitudinal relationships remain scarce.

On the other hand, in the use of longitudinal and multimodal data, missing data remains a common but great challenge, limiting its direct usage in most conventional machine learning or deep learning methods. Previous studies proposed to handle this issue by using three types of strategies: (a) exclude the subjects with missing data (Khan & Zubair, 2022; Lo & Jagust, 2012); (b) employ network characteristics and special strategies to leverage all available data (Chen et al., 2022; Zhou et al., 2019); and (c) impute the missing data (Che, Purushotham, Cho, Sontag, & Liu, 2018; Jung et al., 2021). For strategy (a), valuable information may be lost when only complete data are used in MCI conversion prediction (Chen et al., 2022; Pan, Chen, Shen, & Xia, 2021). The methods of type (b) show promising results in effectively utilizing missing data, but still require complete data as input in practical applications. For type (c), several studies initially imputed missing data and subsequently used the imputed data to train a model for prediction (El-Sappagh et al., 2020; Wang, Qiu, & Yu, 2018). However, this decoupling two-stage strategy may lead to sub-optimal results, and the chosen imputation method heavily influence the model performance (Ghazi et al., 2019; Ma, Li, & Cottrell, 2022). Various techniques can be used for imputing missing values, such as simple forward/backward filling (Nguyen et al., 2020), matrix factorization based on singular value decomposition (Anandkumar, Ge, Hsu, Kakade, & Telgarsky, 2014), and statistical (El-Sappagh et al., 2020) and machine learning (Zhang, Wu, et al., 2021) methods. Recently, recurrent neural networks (RNN) have shown advancements in data imputation (Jung et al., 2021; Nguyen et al., 2020) but necessitate modality-complete data at BL (Ghazi et al., 2019; Jung et al., 2021; Nguyen et al., 2020). Unfortunately, many subjects have no available PET images at BL due to various practical issues (e.g., high cost, poor image quality, and others) (Liu et al., 2022). Moreover, the estimated data may have errors that can affect the performance of subsequent tasks (Ghazi et al., 2019; Ma et al., 2022). Therefore, further reducing the errors of imputed data remains a problem to be solved. Additionally, previous studies tended to utilize all available longitudinal data to estimate the disease status beyond the last time point (Ghazi et al., 2019; Jung et al., 2021; Nguyen et al., 2020). In clinical practice, MCI conversion prediction must be achieved in the early stages of disease progression to facilitate timely intervention, specifically at BL (Albert et al., 2011; Petersen et al., 2014). On this basis, the model can focus on the prediction performance at BL without requiring longitudinal data as inputs during testing/usage phase. Therefore, how to effectively use disease progression information in longitudinal data while only BL data are required as inputs at the model testing phase is also a problem that need consideration.

To address the aforementioned challenges, an end-to-end multi-task deep learning framework, named multi-view imputation and cross-attention network (MCNet), is proposed to utilize incomplete longitudinal and multimodal data for MCI conversion prediction (i.e., classify subjects into sMCI and pMCI). The proposed method consists of data imputation and conversion prediction modules. These two modules share the same multimodal features extracted from the RNN-based network for multi-task learning, including data imputation, longitudinal classification, and conversion prediction tasks. First, in the data imputation module, a novel multi-view imputation strategy with adversarial learning is designed to utilize disease progression information to impute MRI/PET data from a longitudinal view and apply the associations between different modalities to impute the PET data from a multimodal view. Moreover, incorporating adversarial learning is conducive to increase the realities and reduce the errors of imputed data. Second, the features obtained from the imputation module are fused by two unique feature fusion blocks, named cross-attention blocks, for final MCI conversion prediction. Various missing data scenarios exist in incomplete longitudinal and multimodal data, which leads to differences in importance among the information contained in the features. Therefore, two cross-attention blocks are developed to weigh features and reflect the information's importance. Then, the fused features are used to accomplish longitudinal classification and conversion prediction tasks. With the data imputation strategy and cross-attention blocks, the proposed method extracts disease progression information from longitudinal data during training and directly applies the information on BL data to obtain prediction results without feeding longitudinal data in the testing phase. In other words, we aim to use disease progression information that is learned from longitudinal data to improve performance of MCI conversion prediction at a single time point. Based on previous reports, no research has combined disease progression information from longitudinal data and multimodal associations from multimodal data to achieve adversarial multi-view imputation at all time points with small errors and integrated classification and prediction tasks in the same framework to achieve joint optimization for MCI conversion prediction. In summary, the contributions of this work are as follows:

- Based on incomplete longitudinal and multimodal data, data imputation and MCI conversion prediction are integrated into a unified network, and a multi-task learning strategy is introduced to achieve joint optimization and improve the prediction performance.
- A multi-view imputation strategy is designed for different modalities and time points to achieve data imputation that can cope with various missing data scenarios. Moreover, the adversarial learning is incorporated into the imputation strategy to make an imputed data distribution that is close to the real distribution, thereby further reducing imputation errors.
- Two cross-attention blocks are proposed to fuse multimodal features at different time points to capture information importance at different time points and modalities, thereby further improving prediction performance.
- With well training, disease progression information learned from longitudinal data can be leveraged by our proposed method to complete MCI conversion prediction with BL data in the testing phase. Hence, our proposed method meets the requirement of early prediction in clinical practice. Additionally, our model can still perform well when only single-modal data (e.g., MRI) are available at BL. The proposed method is trained on two datasets provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (i.e., ADNI-1 and ADNI-2) and tested on two external independent datasets (i.e., ADNI-3 and Open Access Series of Imaging Studies-3 [OASIS-3]). Competitive results are achieved using the proposed method, which further demonstrates the well generalized ability of the proposed method.

For clarification and readability of the paper, Table 1 summarizes the acronyms that appear in the main text.

Table 1
Acronyms and corresponding explanations.

Acronyms	Explanations	Acronyms	Explanations
ACC	Accuracy	MAE	Mean absolute error
AD	Alzheimer's disease	MCI	Mild cognitive impairment
ADNI	Alzheimer's disease neuroimaging initiative	MCNet	Multi-view imputation and cross-attention network
AJRNN	Adversarial joint-learning recurrent neural network	MLP	Multilayer perceptron
ANTs	Advanced normalization tools	MMSE	Mini-mental state examination
AUC	Area under receiver operating characteristic curve	MNI	Montreal neurological institute
BAC	Balanced accuracy	MRI	Magnetic resonance imaging
BL	Baseline visit	OASIS	Open access series of imaging studies
BLS	Board learning system	PET	Positron emission tomography
CN	Cognitive normal	RMSE	Root mean square error
DRM	Deep recurrent model	RNN	Recurrent neural networks
DTI	Diffusion tensor imaging	ROI	Region-of-interest
FC	Fully connected	SVM	Support vector machine
FDG-PET	Fluorodeoxyglucose positron emission tomography	pMCI	Progressive mild cognitive impairment
GM	Gray matter	fMRI	Functional magnetic resonance imaging
GRU	Gated recurrent unit	sMCI	Stable mild cognitive impairment
LSTM	Long short-term memory	std	Standard deviation

2. Related work

2.1. Multimodal AD-related analysis

Many studies have focused on the applications of multimodal neuroimages, which contain complementary information, for AD diagnosis (Behrad & Abadeh, 2022; Shi, Zheng, Li, Zhang, & Ying, 2017). For example, different deep neural networks were first used to learn high-level features from multimodal data (e.g., PET, MRI, genetic data, and others) at BL. Then, the features learned from the different modalities were concatenated directly for final AD detection (Venugopalan et al., 2021). Zhou et al. (2019) proposed a three-stage deep feature learning and fusion framework to accomplish multi-scale feature fusion for AD prediction. Better prediction performance was achieved using multimodal data than single-modal data. However, the direct concatenation of multimodal features may fail to take full advantage of the information from different modalities (Ning, Xiao, Feng, Chen, & Zhang, 2021). Zu, Wang, Zhou, Wang, and Zhang (2018) utilized multi-kernel learning to combine multimodal data for AD classification. Furthermore, Leng et al. (2023) developed a cross enhanced fusion mechanism to emphasize the correlation and complementarity between multimodal features for AD diagnosis. Although these strategies can be used to effectively combine multimodal data, their effects on the feature fusion of longitudinal and multimodal data need further investigation.

In the longitudinal and multimodal study, associations among different modalities and different time points should be considered. Inspired by the self-attention mechanism (Vaswani et al., 2017), we introduce two cross-attention blocks to exploit the importance of the features extracted from different modalities at different time points and to enhance the performance of MCI conversion prediction.

2.2. Longitudinal AD-related analysis

An increasing number of studies attempted to utilize these data for AD-related analysis with the increasing amount of available longitudinal data collected at follow-up time points (Abdelaziz, Wang, & Elazab, 2021; Brand, Nichols, Wang, Shen, & Huang, 2019; Huang, Yang, Feng, & Chen, 2017; Zhang, Wu, et al., 2021). Some studies explored the use of traditional machine learning methods. Huang, Chen, Yu, Lai, and Feng (2021) proposed a novel temporal group sparsity regression and additive model to identify the associations between longitudinal imaging and genetic data for the detection of potential AD biomarkers. More recently, deep learning methods have shown great potential in AD analysis and have been applied to related classification and regression tasks with promising performance (El-Sappagh et al., 2020; Jung et al., 2021). Among them, RNN-based deep learning methods are often used in longitudinal studies. Nevertheless, conventional RNNs are designed to be used with complete data; incomplete data still present serious problems for the applications of RNN. Some studies tried to alleviate the negative impact of this problem by taking advantage of RNN to deal with variable-length series data for imaging feature extraction and AD diagnosis but directly ignored the missing data issue (Huang, Lai, et al., 2021). Che et al. (2018) designed a gated recurrent unit with decay (GRU-D) to introduce a decay mechanism using information on the interval and location of missing values. Then, they combined decay rates with the incomplete longitudinal data to accomplish classification. Moreover, Ghazi et al. (2019) proposed a generalized backpropagation through time algorithm for long short-term memory (LSTM), and this method can handle missing input and output values. All the missing values of longitudinal data were initialized with zeros.

Although certain attempts have been made in these methods, missing data issue is only considered in longitudinal view, whereas discarded in multimodal view. Hence, further exploration is still needed to reduce the impact of missing data issue on the final prediction task.

2.3. Data imputation

Missing data is a common issue for longitudinal and multimodal data and may decrease the accuracy of MCI conversion prediction. El-Sappagh et al. (2020) tried to solve this issue by discarding the subjects with serious missing data conditions and using the *k*-nearest neighbor algorithm to impute missing values for remaining subjects. However, this kind of data imputation method is a decouple two-stage methodology, which may lead to sub-optimal results (Ma et al., 2022). Therefore, some studies trained a unified model to accomplish data imputation and prediction or classification simultaneously. Nguyen et al. (2020) used the temporal dependencies of RNN to impute a set of longitudinal data and performed classification in a unified model. However, only temporal associations were considered in this method, and the correlations between different modalities at a time point may be ignored. Jung et al. (2021) integrated data imputation and longitudinal data classification into a unified framework by utilizing information on the interval between missing values, the location of missing values, and multivariate relations, where the relations of different modalities can be reflected by the multivariate relations and reasonable data imputation, and classification results can be achieved by this method. Additionally, advanced data imputation/prediction methods have been developed involving the application of deep learning techniques in other fields (Zhang, Zhao, & He, 2021). For instance, Zhang et al. proposed a method that combines generalized learning systems with LSTM for predicting battery capacity and its remaining useful life, which can be extended and adapted for medical data analysis (Zhao, Zhang, & Wang, 2022). However, estimation errors of imputation data may accumulate in these unified training methods during the feedforward of the RNN (Bengio, Vinyals, Jaitly, & Shazeer, 2015). Moreover, the missing data issue at BL cannot be addressed in previous studies.

Generative adversarial network has unparalleled advantages in data generation. Therefore, Ma et al. (2022) introduced the adversarial learning strategy into the data imputation and classification framework to solve the accumulated errors problem, which can further improve the classification performance. Inspired by this idea, we also introduce an adversarial loss in the proposed method. This study uses such a strategy for the first time in AD longitudinal and multimodal data imputation. Moreover, we design a novel multi-view imputation method according to the specific scenarios of missing data to effectively impute missing data and solve the missing data issue at BL.

Additionally, these methods were not designed to consider how to complete the prediction using only BL data during the model testing/usage phase. Specifically, most existing methods still require longitudinal data as inputs in the testing/usage phase (El-Sappagh et al.,

2020; Nguyen et al., 2020). Instead of using longitudinal data during testing phase, the proposed method tries to learn helpful disease progression information from longitudinal data when training and then uses the learned information to improve MCI conversion prediction at BL when testing. We hypothesize that even if longitudinal data are limited, the underlying information of disease trajectories can be leveraged by deep learning approaches. When the model is properly trained, the model is able to judge the trend of the disease from BL data through the learned pattern of disease progression and utilized this auxiliary knowledge to enhance the performance of MCI conversion prediction. Therefore, only multimodal data at BL or even single modal data at BL are needed in the testing phase in the proposed method.

3. Materials

The brain imaging data used in this paper were obtained from the ADNI (https://www.adni.loni.usc.edu/) and OASIS-3 databases (https://www.oasis-brains.org/). A total of 1387 subjects with T1-weighted MRI and fluorodeoxyglucose positron emission tomography (FDG-PET) images in the three ADNI subsets, namely, ADNI-1, ADNI-2, and ADNI-3, were collected in this study. For ADNI-1 and ADNI-2, images at BL and at 6, 12, 24, and 36 months were included when available, whereas only images at BL provided by ADNI-3 were included as the independent testing set. Specifically, subjects that had MRI images at BL and more than two other time points were included, and the available status of PET images was not considered as a criterion for selecting subjects for ADNI-1 and ADNI-2. Moreover, an additional 143 subjects obtained from OASIS-3 were used as another independent testing set. In terms of mini-mental state examination (MMSE) scores and clinical dementia rating, the clinical status of a subject at a time point can be divided into three categories, i.e., cognitive normal (CN), MCI, and AD. In this study, all subjects were further categorized into four groups based on the individual clinical status at BL and future time points, as follows: (a) CN: the subjects were diagnosed as CN at BL and remained CN afterwards; (b) sMCI: the subjects were diagnosed as MCI at all time points; (c) pMCI: the subjects were diagnosed as MCI at BL and then converted to AD within three years; and (d) AD: the subjects had a clinical status of AD at all time points. The number of enrolled subjects and more demographic information are shown in Table 2. Subjects with reverse conversion of clinical status were removed. The details of the subject number with different image modalities at different time points are listed in Table 3.

Raw MRI images acquired by 1.5T/3T scanners and PET images preprocessed by ADNI were downloaded. Then, all MRI images were processed through the following procedures (Chen et al., 2022): (a) anterior commissure-posterior commissure correction by using MIPAV software (https://mipav.cit.nih.gov/); (b) image intensity inhomogeneity correction by using N4 algorithm; (c) skull stripping via a robust brain extraction network named HD-BET (Isensee et al., 2019); (d) registering images to Montreal Neurological Institute (MNI) space via advanced normalization tools (ANTs) (https://github.com/ANTsX/ANTs); (e) segmentation of three main tissues, i.e., gray matter (GM), white matter, and cerebrospinal fluid, by using Atropos algorithm in ANTs; (f) labeling 90 regions-of-interest (ROIs) on all registered images based on the automated anatomical label atlas of MNI space; and (g) computing the GM tissue volume of each ROI in the labeled images. Subsequently, PET images obtained from OASIS-3 were preprocessed in the same way as those obtained from ADNI (Jagust et al., 2015). Finally, preprocessed PET images were aligned to their corresponding MRI by using co-registration strategy and the average intensity value of each ROI was calculated as a PET feature. Thus, 90-dimensional ROI features were separately extracted from the MRI and PET data for each subject.

Table 2
Demographic information of enrolled subjects. The age and MMSE scores are presented by mean ± standard deviation (std).

Dataset	Total	Category	Number	Male/Female	Age	MMSE
ADNI-1	543	CN	165	90/75	75.3 ± 5.2	29.0 ± 1.1
		sMCI	144	91/53	74.6 ± 7.5	27.3 ± 1.6
		pMCI	116	68/48	73.8 ± 6.9	26.8 ± 1.8
		AD	118	60/58	75.1 ± 7.8	23.4 ± 1.9
ADNI-2	758	CN	255	129/126	73.7 ± 5.8	29.1 ± 1.2
		sMCI	286	163/123	71.5 ± 7.5	28.2 ± 1.6
		pMCI	104	57/47	73.4 ± 6.7	27.6 ± 1.9
		AD	113	69/44	74.3 ± 7.8	23.8 ± 2.5
ADNI-3	86	sMCI	73	44/29	75.5 ± 7.6	28.5 ± 1.2
		pMCI	13	8/5	74.0 ± 6.6	26.8 ± 2.8
OASIS-3	143	CN	65	37/28	73.7 ± 8.6	29.2 ± 1.0
		sMCI	67	27/21	74.9 ± 5.6	28.4 ± 1.9
		pMCI	37	19/4	75.6 ± 7.9	27.3 ± 2.3
		AD	7	4/3	76.1 ± 5.6	24.4 ± 1.3

Table 3
Number of available subjects for different modalities at different time points in ADNI-1 and ADNI-2, where M06, M12, M24, and M36 represent 6, 12, 24, and 36 months, respectively.

Dataset	BL	M06	M12	M24	M36
ADNI-1 (MRI/PET)	543/292	534/274	527/267	451/225	284/134
ADNI-2 (MRI/PET)	758/606	534/0	745/111	580/297	396/1

4. Method

A multi-task learning framework, named MCNet, is proposed for joint data imputation, longitudinal classification, and MCI conversion prediction. Fig. 1 presents an overview of the proposed method consisting of two modules. In the first module (Fig. 1(a)), data imputation was conducted on the MRI and PET ROI features from the multi-views (i.e., longitudinal and multimodal views) using a MinimalRNN-based network. Moreover, an adversarial learning block was proposed to reduce imputation errors and increase the realities of the imputed data. Therefore, longitudinal and multimodal features (hidden features $\boldsymbol{H}^{\mathrm{MRI}}$ and $\boldsymbol{H}^{\mathrm{PET}}$ in Fig. 1) can be well explored with the imputed data. In the second module (Fig. 1(b)), two cross-attention blocks were applied to effectively fuse the multimodal and longitudinal features shared from the data imputation module for longitudinal classification and MCI conversion prediction. On the one hand, multimodal features at each time point were fed into the first cross-attention block to fuse multimodal features and capture multimodal associations for longitudinal classification. On the other hand, the fused multimodal features at all time points were fed into the second cross-attention block to exploit potential disease progression information for MCI conversion prediction. The model was trained on all available longitudinal and multimodal data in the training phase, and only data at BL were used as inputs in the testing phase. The proposed method still performed well when PET data was missing at BL.

4.1. Notations

In this study, matrices, vectors, and scalars are denoted as bold uppercase letters, boldface lowercase letters, and normal italic letters, respectively. Moreover, all scalars about time points are enclosed in parentheses. The ROI features of MRI and PET data can be represented as $\boldsymbol{X} = \{\boldsymbol{X}^S\}_{S=\mathrm{MRI,PET}}$, where S represents the image modality and $\boldsymbol{X}^S = \left\{\boldsymbol{x}^S_{(1)}, \ldots, \boldsymbol{x}^S_{(t)}, \ldots, \boldsymbol{x}^S_{(T)}\right\} \in \mathbb{R}^{N\times T\times D}$. N, T, and D are the subject number, number of time points, and dimension of ROI features, respectively. Missing time points often appear in longitudinal MRI and PET data, and mask vectors $\boldsymbol{M}^S = \left\{\boldsymbol{m}^S_{(1)}, \ldots, \boldsymbol{m}^S_{(t)}, \ldots, \boldsymbol{m}^S_{(T)}\right\}$ are applied to indicate whether data exist in a time point, where $\boldsymbol{m}^S_{(t)} \in \mathbb{R}^{N\times 1}$. In particular, the value in $\boldsymbol{m}^S_{(t)}$ is 1 when $\boldsymbol{x}^S_{(t)}$ exists and 0 when $\boldsymbol{x}^S_{(t)}$ is missing. Moreover, the longitudinal and the clinical status labels are denoted as $\{\boldsymbol{y}_{(t)}\}^T_{t=1}$ and $\boldsymbol{C}$, respectively, where $\boldsymbol{y}_{(t)} \in \mathbb{R}^{N\times 1}$ is

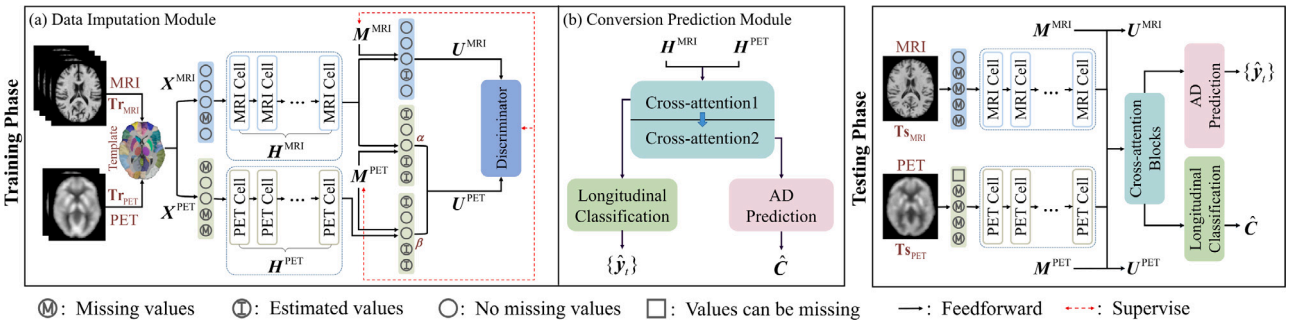


Fig. 1. Overview of the proposed framework: (a) Data imputation module combined with adversarial learning; (b) Conversion prediction module for longitudinal classification and MCI conversion prediction. In the training phase, MRI and PET data at different time points were trained using modules (a) and (b), where $\mathbf{Tr}_{\mathrm{MRI}}$ and $\mathbf{Tr}_{\mathrm{PET}}$ represent the MRI and PET images of training subjects, respectively, with a total of 1171 training subjects. In the testing phase, only data at BL were used as inputs, where PET data at BL can either be available or missing; $\mathbf{Ts}_{\mathrm{MRI}}$ and $\mathbf{Ts}_{\mathrm{PET}}$ represent the MRI and PET images of testing subjects, respectively, with 130 subjects for ADNI-1/2, 86 subjects for ADNI-3, and 104 subjects for OASIS-3.

Table 4
Main notations used in this study.

Symbol	Description
Notations of inputs	
$\mathbf{Tr}_{\mathrm{MRI}}$, $\mathbf{Tr}_{\mathrm{PET}}$	MRI and PET images of training subjects
$\mathbf{Ts}_{\mathrm{MRI}}$, $\mathbf{Ts}_{\mathrm{PET}}$	MRI and PET images of testing subjects
Notations of modality features	
K	Image modality
N	Total number of subjects
T	Total number of time points
D	Dimension of ROI features
$\boldsymbol{x}_{(t)}^{S}$	ROI features of modality S at tth time point
$\boldsymbol{X}^{S}=\left\{\boldsymbol{x}_{(1)}^{S},\ldots,\boldsymbol{x}_{(t)}^{S},\ldots,\boldsymbol{x}_{(T)}^{S}\right\}$	ROI features of modality S at all time points
$\boldsymbol{X}=\left\{\boldsymbol{X}^{S}\right\}_{S=\mathrm{MRI,PET}}$	ROI features of all modalities at all time points
$\boldsymbol{m}_{(t)}^{S}$	Mask vector of modality S at tth time point
$\boldsymbol{M}^{S}=\left\{\boldsymbol{m}_{(1)}^{S},\ldots,\boldsymbol{m}_{(t)}^{S},\ldots,\boldsymbol{m}_{(T)}^{S}\right\}$	Mask vectors of modality S at all time points
$\boldsymbol{y}_{(t)}$	Longitudinal label at tth timepoint
$\boldsymbol{C}$	Conversion label of a subject
Notations in data imputation module	
(a) Notations in MinimalRNN	
$\boldsymbol{\Phi}(\cdot)$	A network for mapping ROI features into a latent representation
$\boldsymbol{z}_{(t)}^{S}$	Latent representation of modality S at tth time point
$\boldsymbol{g}_{(t)}^{S}$	Update gate of modality S
$\boldsymbol{h}_{(t)}^{S}$	Hidden feature of modality S at tth time point, i.e., output of MinimalRNN
$\boldsymbol{H}^{S}=\left\{\boldsymbol{h}_{(1)}^{S},\ldots,\boldsymbol{h}_{(t)}^{S},\ldots,\boldsymbol{h}_{(T)}^{S}\right\}$	Hidden features of modality S at all time points
$\boldsymbol{W}_{x}^{S}$, $\boldsymbol{W}_{h}^{S}$, $\boldsymbol{W}_{z}^{S}$, $\boldsymbol{b}_{x}^{S}$	Learnable weight matrices and bias vectors of modality S in minimalRNN
$\sigma(\cdot)$	Sigmoid activation function
$\tanh(\cdot)$	Hyperbolic tangent function
$\odot$	Element-wise product
Notations in data imputation module	
(b) Notations in multi-view imputation	
$\tilde{\boldsymbol{x}}_{\mathrm{cs},(t)}^{\mathrm{PET}}$	Estimated ROI features of PET at tth time point from the multimodal view
$\tilde{\boldsymbol{x}}_{\mathrm{lg},(t)}^{\mathrm{PET}}$	Estimated ROI features of PET at tth time point from the longitudinal view
$\hat{\boldsymbol{x}}_{(t)}^{S}$	Final estimated ROI features of modality S at tth time point
$\boldsymbol{u}_{(t)}^{S}$	Imputed ROI features of modality S at tth time point
$\boldsymbol{U}^{S}=\left\{\boldsymbol{u}_{(1)}^{S},\ldots,\boldsymbol{u}_{(t)}^{S},\ldots,\boldsymbol{u}_{(T)}^{S}\right\}$	Imputed ROI features of modality S at all time points
α, β	Learnable weighted coefficients
$\boldsymbol{W}_{\mathrm{cs}}^{\mathrm{PET}}$, $\boldsymbol{W}_{\mathrm{lg}}$, $\boldsymbol{b}_{\mathrm{cs}}^{\mathrm{PET}}$, $\boldsymbol{b}_{\mathrm{lg}}$	Learnable weight matrices and bias vectors in multi-view imputation

Symbol	Description
(b) Notations in multi-view imputation	
Concat $(\cdot)$	The operation of feature concatenation
$\boldsymbol{W}_{\mathrm{cs}}^{\mathrm{PET}}$, $\boldsymbol{W}_{\mathrm{lg}}$	Learnable weighted coefficients
$\boldsymbol{F}_{\mathrm{mini}}(\cdot)$	The update function of MinimalRNN
$\boldsymbol{I}(\cdot)$	The process of data imputation
$\|\cdot\|$	Mean absolute error
$\mathcal{L}_{est}$	Estimation loss function
Notations in data imputation module	
(c) Notations in adversarial learning	
$\log(\cdot)$	Logarithmic function
$p_{\mathrm{imp}}(\boldsymbol{u})$	Imputed data distribution
$p_{\mathrm{real}}(\boldsymbol{u})$	Real data distribution
$\mathrm{Ds}(\cdot)$	Discriminator
$\mathcal{L}_{\mathrm{D}}$	Discriminator loss function
$\mathcal{L}_{\mathrm{adv}}$	Adversarial loss function
Notations in conversion prediction module	
J	The number of heads in first cross-attention blocks
J'	The number of heads in second cross-attention blocks
D'	Dimension of hidden features
$\boldsymbol{H}_{(t)}$	Concatenated hidden features at tth time point
$\boldsymbol{Q}_{(t)}^{j}$, $\boldsymbol{K}_{(t)}^{j}$, $\boldsymbol{V}_{(t)}^{j}$	Three projection matrices of head j at tth time point in first cross-attention block
$\boldsymbol{A}_{(t)}^{j}$	Attention matrix of head j at tth time point in first cross-attention block
$\tilde{\boldsymbol{H}}_{(t)}$	Fused features of first cross-attention block
$\tilde{\boldsymbol{H}}$	Concatenation of fused features of first cross-attention block
$\hat{\boldsymbol{Q}}^{j'}$, $\hat{\boldsymbol{K}}^{j'}$, $\hat{\boldsymbol{V}}^{j'}$	Three projection matrices of head j' in second cross-attention block
$\hat{\boldsymbol{A}}^{j'}$	Attention matrix of head j' in second cross-attention block
$\hat{\boldsymbol{H}}$	Final features for MCI conversion prediction
$\boldsymbol{W}_{q}^{j}$, $\boldsymbol{W}_{k}^{j}$, $\boldsymbol{W}_{v}^{j}$, $\boldsymbol{W}_{at1}$, $\boldsymbol{W}_{at2}$, $\boldsymbol{W}_{cls}$, $\boldsymbol{W}_{c}$, $\boldsymbol{b}_{at1}$, $\boldsymbol{b}_{at2}$, $\boldsymbol{b}_{cls}$, $\boldsymbol{b}_{c}$	Learnable weight matrices and bias vectors in conversion prediction module
$\hat{\boldsymbol{y}}_{(t)}$	Longitudinal prediction result at tth time point
$\hat{\boldsymbol{C}}$	Conversion prediction results
$\mathrm{Softmax}(\cdot)$	Softmax activation function
$\mathcal{L}_{\mathrm{cls}}$	Longitudinal classification loss function
$\mathcal{L}_{\mathrm{pred}}$	Conversion prediction loss function
$\mathcal{L}$	Overall loss function
λ, ζ, ξ	The hyperparameters in the overall loss function

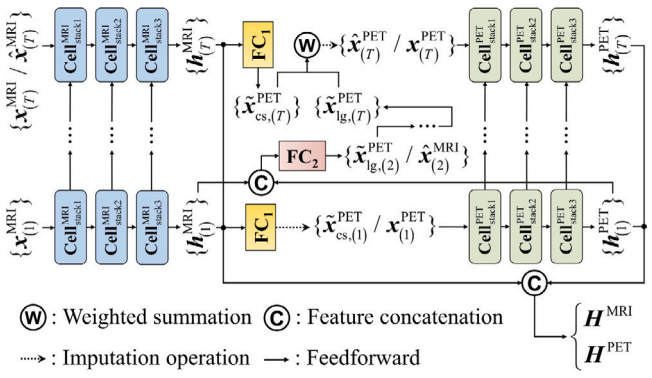


Fig. 2. Illustration of data imputation module. The parameters in all time points are shared, i.e., the same memory cells in figure share the same parameters.

longitudinal label at tth time point. Specifically, if the clinical status is unchanged, $\boldsymbol{y}_{(t)} = 0$; otherwise, $\boldsymbol{y}_{(t)} = 1$. The longitudinal labels are used in the longitudinal classification task. $\boldsymbol{C} \in \mathbb{R}^{N\times 1}$ indicates the clinical status of a subject. For instance, if the clinical status of a subject is sMCI, then $\boldsymbol{C} = 0$. Otherwise, if it is pMCI, then $\boldsymbol{C} = 1$. Moreover, $\boldsymbol{C}$ is applied to MCI conversion prediction. The main notations used in this study are listed in Table 4.

4.2. Data imputation module

4.2.1. Minimal recurrent neural network

We briefly considered MinimalRNN, which is used as the backbone network of the proposed method. MinimalRNN is a distinctive RNN architecture that adopts the minimum number of operations within RNN without sacrificing performance (Chen, 2017). Moreover, MinimalRNN can be used to capture disease progression information, the trainability of which can also be guaranteed.

A single MinimalRNN is a chain structure composed of several memory cells, where each cell corresponds to a time point. Moreover, all cells share the same parameters. For each cell, the input $\boldsymbol{x}^S_{(t)}$ is first fed into a fully connected (FC) network $\boldsymbol{\Phi}(\cdot)$ to generate a latent representation $\boldsymbol{z}^S_{(t)}$. Through this operation, the representations are confined to move within this latent space (Chen, 2017). Given $\boldsymbol{z}^S_{(t)}$, the weight of update gate $\boldsymbol{g}^S_{(t)}$ can be simply learned with a sigmoid function $\sigma(\cdot)$. Moreover, the update gate $\boldsymbol{g}^S_{(t)}$ weighs the contributions of the previous hidden feature $\boldsymbol{h}^S_{(t-1)}$ and the latent representation $\boldsymbol{z}^S_{(t)}$ toward the current hidden feature $\boldsymbol{h}^S_{(t)}$. Disease progression information can be continuously captured from the data during the forward process from the hidden feature$\boldsymbol{h}^S_{(t-1)}$ to $\boldsymbol{h}^S_{(t)}$. Therefore, the update process can be formulated as:

$$\boldsymbol{z}^S_{(t)} = \tanh\left(\boldsymbol{\Phi}\left(\boldsymbol{x}^S_{(t)}\right)\right) = \tanh\left(\boldsymbol{W}^S_x \boldsymbol{x}^S_{(t)} + \boldsymbol{b}^S_x\right) \tag{1}$$

$$\boldsymbol{g}^S_{(t)} = \sigma\left(\boldsymbol{W}^S_h \boldsymbol{h}^S_{(t-1)} + \boldsymbol{W}^S_z \boldsymbol{z}^S_{(t)}\right) \tag{2}$$

$$\boldsymbol{h}^S_{(t)} = \boldsymbol{g}^S_{(t)} \odot \boldsymbol{h}^S_{(t-1)} + \left(1 - \boldsymbol{g}^S_{(t)}\right) \odot \boldsymbol{z}^S_{(t)} \tag{3}$$

where $\boldsymbol{W}^S_x$ and $\boldsymbol{b}^S_x$ denote the parameters of the embedding operation, $\boldsymbol{W}^S_h$ and $\boldsymbol{W}^S_z$ denote the parameters related to update gate $\boldsymbol{g}^S_{(t)}$, and $\odot$ is the element-wise product. Finally, the hidden features at all time points can be denoted as $\boldsymbol{H}^S = \left\{\boldsymbol{h}^S_{(1)}, \ldots, \boldsymbol{h}^S_{(t)}, \ldots, \boldsymbol{h}^S_{(T)}\right\}$.

4.2.2. Multi-view imputation

In this study, longitudinal and multimodal imaging data were used for accurate MCI conversion prediction. However, the missing data issue from longitudinal and multimodal views is the main limitation when using this type of data. Based on our observations on the ADNI datasets, the cases of missing data can be divided into three main scenarios: (a) PET data missing at BL; (b) MRI data missing at tth time point; and (c) PET data missing at tth time point. We designed a novel multi-view data imputation method that is different from traditional data imputation methods to impute missing values in different cases and address the complexity of the missing data issue.

Compared with longitudinal MRI images, more serious missing data issue appeared in longitudinal PET images. Therefore, as shown in Fig. 2, we adopted two separate stacked multi-layer MinimalRNNs to capture the different longitudinal features of different modalities and then impute data. MinimalRNN cannot perform feedforward when data are unavailable at BL. Therefore, the missing data issue of PET data at BL (scenario (a)) was considered first. We first imputed the PET data at BL from the multimodal view by using a FC network (FC_1 in Fig. 2) with hyperbolic tangent activation function due to the nonlinear relationship between PET and MRI data. We utilized the hidden feature $\boldsymbol{h}^{\mathrm{MRI}}_{(t)}$ encoded from the MRI data at BL to estimate the PET data at BL:

$$\tilde{\boldsymbol{x}}^{\mathrm{PET}}_{\mathrm{cs},(t)} = \tanh\left(\boldsymbol{W}^{\mathrm{PET}}_{\mathrm{cs}} \boldsymbol{h}^{\mathrm{MRI}}_{(t)} + \boldsymbol{b}^{\mathrm{PET}}_{\mathrm{cs}}\right) \tag{4}$$

where $\boldsymbol{W}^{\mathrm{PET}}_{\mathrm{cs}}$, and $\boldsymbol{b}^{\mathrm{PET}}_{\mathrm{cs}}$ are learnable parameters of FC network for PET estimation at BL.

For MRI data missing at tth time point (scenario (b)), the potential relationship (i.e., disease progression information) between adjacent time points was utilized for the imputation of the longitudinal view. Previous disease progression information up to $(t-1)$th time point was contained in the hidden features $\boldsymbol{h}^{\mathrm{MRI}}_{(t-1)}$ and $\boldsymbol{h}^{\mathrm{PET}}_{(t-1)}$. Accordingly, MRI data can be estimated from the longitudinal view. Specifically, the MRI data $\hat{\boldsymbol{x}}^{\mathrm{MRI}}_{(t)}$ of the tth time point were estimated using the concatenation of hidden features $\boldsymbol{h}^{\mathrm{MRI}}_{(t-1)}$ and $\boldsymbol{h}^{\mathrm{PET}}_{(t-1)}$. Meanwhile, the PET data $\tilde{\boldsymbol{x}}^{\mathrm{PET}}_{\mathrm{lg},(t)}$ of the same time point can also be estimated simultaneously from the longitudinal view. This process was implemented through a FC network (FC_2 in Fig. 2) and can be formulated as follows:

$$\hat{\boldsymbol{x}}^{\mathrm{MRI}}_{(t)}, \tilde{\boldsymbol{x}}^{\mathrm{PET}}_{\mathrm{lg},(t)} = \boldsymbol{W}_{\mathrm{lg}} \mathrm{Concat}\left(\boldsymbol{h}^{\mathrm{MRI}}_{(t-1)}, \boldsymbol{h}^{\mathrm{PET}}_{(t-1)}\right) + \boldsymbol{b}_{\mathrm{lg}} \tag{5}$$

where $\boldsymbol{W}_{\mathrm{lg}}$ and $\boldsymbol{b}_{\mathrm{lg}}$ are learnable parameters for MRI/PET longitudinal estimation, and $\mathrm{Concat}(\cdot)$ represents feature concatenation. In this study, our focus was solely on imputation from the longitudinal view for MRI data. In clinical practice, an MRI image is typically available at a given time point, whereas the corresponding PET image may be missing from the multimodal view.

For PET data missing at tth time point (scenario (c)), we adaptively combined the two estimated values obtained from the longitudinal and multimodal views to calculate the final estimated value at tth time point. In this way, we leveraged the complementary information from different modalities at the current time point and disease progression information from previous time points for multi-view imputation. Specifically, we can take advantage of the hidden features $\boldsymbol{h}^{\mathrm{MRI}}_{(t-1)}$ and $\boldsymbol{h}^{\mathrm{PET}}_{(t-1)}$ propagated from the previous $(t-1)$ time points as Eq. (5) and utilize the hidden feature encoded from the MRI data at the tth $(t > 1)$ time point as Eq. (4) to estimate the PET data (weighted summation operation in Fig. 2) except for BL:

$$\begin{cases} \hat{\boldsymbol{x}}^{\mathrm{PET}}_{(t)} = \alpha\left(\tilde{\boldsymbol{x}}^{\mathrm{PET}}_{\mathrm{cs},(t)}\right) + \beta\left(\tilde{\boldsymbol{x}}^{\mathrm{PET}}_{\mathrm{lg},(t)}\right), & \text{if} \quad t > 1 \\ \hat{\boldsymbol{x}}^{\mathrm{PET}}_{(t)} = \tilde{\boldsymbol{x}}^{\mathrm{PET}}_{\mathrm{cs},(t)}, & \text{if} \quad t = 1. \end{cases} \tag{6}$$

where α and β are learnable weighted coefficients, and $\alpha + \beta = 1$.

With the estimated PET and MRI data $\hat{\boldsymbol{x}}^S_{(t)}$ and mask vector $\boldsymbol{m}^S_{(t)}$, we can impute missing data of both modalities at all time points. Specifically, the imputed data $\boldsymbol{u}^S_{(t)}$ at tth time point can be defined as:

$$\boldsymbol{u}^S_{(t)} = \boldsymbol{m}^S_{(t)} \odot \boldsymbol{x}^S_{(t)} + \left(1 - \boldsymbol{m}^S_{(t)}\right) \odot \hat{\boldsymbol{x}}^S_{(t)} \tag{7}$$

After the imputation steps, the update equation of MinimalRNN can be formulated as:

$$\begin{aligned} \boldsymbol{h}^S_{(t)} &= F_{\mathrm{mini}}\left(\boldsymbol{h}^S_{(t-1)}, I\left(\boldsymbol{x}^S_{(t)}, \boldsymbol{m}^S_{(t)}, \boldsymbol{h}^S_{(t-1)}\right)\right) \\ &= F_{\mathrm{mini}}\left(\boldsymbol{h}^S_{(t-1)}, \boldsymbol{u}^S_{(t)}\right) \end{aligned} \tag{8}$$

where $I(\cdot)$ represents the process of data imputation, and $F_{\text{mini}}(\cdot)$ represents the update function of MinimalRNN as listed in Eqs. (1)–(3).

Therefore, we can obtain the complete longitudinal MRI and PET data $\boldsymbol{U}^S = \left\{\boldsymbol{u}^S_{(1)}, \ldots, \boldsymbol{u}^S_{(t)}, \ldots, \boldsymbol{u}^S_{(T)}\right\}$ and corresponding hidden features $\boldsymbol{H}^S$ through the imputation module. Finally, mean absolute error (MAE) was used to measure the loss between the estimated data and the real data, which can be defined as:

$$\mathcal{L}_{est} = \sum_{t=1}^{T} \sum_{S}^{\{\text{MRI,PET}\}} \left(\left\| \boldsymbol{x}^S_{(t)} - \hat{\boldsymbol{x}}^S_{(t)} \right\|_1 \odot \boldsymbol{m}^S_{(t)} \right) \tag{9}$$

4.2.3. Adversarial learning

Although multi-view estimation was applied to the estimation of longitudinal and multimodal ROI features, some estimation errors still existed. These errors may be accumulated in the feedforward of MinimalRNN. Thus, an adversarial learning strategy was incorporated into the proposed method to alleviate this dilemma. The adversarial learning block can be defined as a minimax game. Our goal was to learn an imputed data distribution $p_{\text{imp}}(\boldsymbol{u})$ that matched the real data distribution $p_{\text{real}}(\boldsymbol{u})$.

Specifically, we added a discriminator consisting of a multilayer perceptron (MLP) with a sigmoid function. The primary objective of this addition was to enforce the close approximation of $p_{\text{imp}}(\boldsymbol{u})$ to $p_{\text{real}}(\boldsymbol{u})$ by fooling the discriminators, thereby mitigating the negative impact of missing values. The supervision signal was provided by mask vectors. Thus, the discriminator loss can be defined as follows:

$$\begin{aligned} \mathcal{L}_{\text{D}} &= -\left[\mathbb{E}_{\boldsymbol{x} \sim p_{\text{real}}(\boldsymbol{u})} \log\left(\text{Ds}\left(\boldsymbol{x}\right)\right) + \mathbb{E}_{\hat{\boldsymbol{x}} \sim p_{\text{imp}}(\boldsymbol{u})} \log\left(1 - \text{Ds}(\hat{\boldsymbol{x}})\right) \right] \\ &= -\sum_{t=1}^{T} \sum_{S}^{\{\text{MRI,PET}\}} \boldsymbol{m}^S_{(t)} \odot \log\left(\text{Ds}\left(\boldsymbol{u}^S_{(t)}\right)\right) \\ &\quad -\sum_{t=1}^{T} \sum_{S}^{\{\text{MRI,PET}\}} \left(1 - \boldsymbol{m}^S_{(t)}\right) \odot \log\left(1 - \text{Ds}\left(\boldsymbol{u}^S_{(t)}\right)\right) \end{aligned} \tag{10}$$

where $\text{Ds}(\cdot)$ denotes the discriminator function and its output is the estimated mask probability. Therefore, the estimated probability for the real data should be maximized to 1, and the estimated probability for the imputed data should be minimized to 0. Then, we introduced an adversarial loss in the data imputation stage to let MinimalRNN maximize the probability of the discriminator output, which will be backpropagated to further optimize the parameters of the MinimalRNN:

$$\mathcal{L}_{\text{adv}} = \sum_{t=1}^{T} \sum_{S}^{\{\text{MRI,PET}\}} \left(1 - \boldsymbol{m}^S_{(t)}\right) \odot \log\left(1 - \text{Ds}\left(\boldsymbol{u}^S_{(t)}\right)\right) \tag{11}$$

Thus, our model first updated the discriminators $\text{Ds}(\cdot)$ to distinguish the real data from the imputed data with $\mathcal{L}_{\text{D}}$ and then updated MinimalRNNs with $\mathcal{L}_{\text{adv}}$. Notably, we considered that the case of missing data in the testing phase as an extreme scenario with only BL data. By leveraging the disease progression information and multimodal correlations learned during training phase, as well as the proposed imputation strategy, we can obtain longitudinal and multimodal features for prediction based solely on the available BL data.

4.3. Conversion prediction module

Our main goal was to perform MCI conversion prediction, that is, to classify subjects into sMCI and pMCI at BL using the proposed method. Thus, we designed a conversion prediction module to capture the longitudinal and multimodal associations and then developed two cross-attention blocks to fuse the longitudinal and multimodal features effectively. Data imputation module and conversion prediction modules share the same features extracted by MinimalRNNs. In this way, we can simply implement a multi-task learning strategy. We accomplished one of the tasks (i.e., data imputation task) in the data imputation module. In this module, besides MCI conversion prediction, we added another auxiliary task (i.e., longitudinal classification) to determine whether the clinical status of the subjects had changed at each time point. Specifically, this task can be used to exploit predictive representation $\boldsymbol{h}^S_{(t)}$ and help the training of the first cross-attention block at each time point, which can contribute to improving the performance of MCI conversion prediction. In conversion prediction module, we developed two cross-attention blocks for feature fusion to effectively combine longitudinal and multimodal information. Specifically, the first cross-attention block was mainly used to explore the relationships among different modalities and fuse the multimodal features of each time point, to determine the importance of different modalities for longitudinal classification/conversion prediction tasks; the second cross-attention block was mainly used to investigate the importance of fused multimodal features from different time points for MCI conversion prediction, and fuse the multimodal features of all time points.

In the first cross-attention block, the hidden features were fused from the multimodal view at each time point through the self-attention mechanism, which was based on a multi-head attention strategy. For head j, the concatenated hidden features $\boldsymbol{H}_{(t)} = \text{Concat}\left(\boldsymbol{h}^{\text{MRI}}_{(t)}, \boldsymbol{h}^{\text{PET}}_{(t)}\right) \in \mathbb{R}^{N\times 2\times D'}$ at tth time point was first translated to three matrices, namely, $\boldsymbol{Q}^j_{(t)} \in \mathbb{R}^{N\times 2\times (D'/J)}$, $\boldsymbol{K}^j_{(t)} \in \mathbb{R}^{N\times 2\times (D'/J)}$, and $\boldsymbol{V}^j_{(t)} \in \mathbb{R}^{N\times 2\times (D'/J)}$, with three projection matrices (i.e., $\boldsymbol{W}^j_q, \boldsymbol{W}^j_k, \boldsymbol{W}^j_v \in \mathbb{R}^{D'\times (D'/J)}$ for all subject), where J is the number of heads, and D' is dimension of hidden features. Then, the attention matrices $\boldsymbol{A}^j_{(t)}$ of different heads can be calculated with $\boldsymbol{Q}^j_{(t)}$, $\boldsymbol{K}^j_{(t)}$, and $\boldsymbol{V}^j_{(t)}$ at each head as follows:

$$\boldsymbol{A}^j_{(t)} = \text{softmax}\left(\boldsymbol{Q}^j_{(t)} {\boldsymbol{K}^j_{(t)}}^{\text{T}} \Big/ \left(\sqrt{D'/J} \right) \right) \boldsymbol{V}^j_{(t)} \tag{12}$$

Next, all heads were concatenated together on feature dimensions and fed into a FC layer to obtain the residual features, which were used to add to the original features $\boldsymbol{H}_{(t)}$ to obtain fused features:

$$\tilde{\boldsymbol{H}}_{(t)} = \left(\boldsymbol{W}_{\text{at}_1} \text{Concat}\left(\boldsymbol{A}^1_{(1)}, \ldots, \boldsymbol{A}^j_{(t)}, \ldots, \boldsymbol{A}^J_{(T)} \right) + \boldsymbol{b}_{\text{at}_1} \right) + \boldsymbol{H}_{(t)} \tag{13}$$

where $\boldsymbol{W}_{\text{at}_1} \in \mathbb{R}^{D'\times D'}$ and $\boldsymbol{b}_{\text{at}_1} \in \mathbb{R}^{D'\times 1}$ are the parameters of the FC layer for the concatenated attention matrices. Moreover, all subjects and both modalities shared the same $\boldsymbol{W}_{\text{at}_1}$ and $\boldsymbol{b}_{\text{at}_1}$. The fused features can be used to classify whether the clinical status changed according to $\tilde{\boldsymbol{H}}_{(t)}$ at t^{th} time point:

$$\hat{\boldsymbol{y}}_{(t)} = \text{softmax}\left(\boldsymbol{W}_{\text{cls}} \tilde{\boldsymbol{H}}_{(t)} + \boldsymbol{b}_{\text{cls}} \right) \tag{14}$$

where $\boldsymbol{W}_{\text{cls}}$ and $\boldsymbol{b}_{\text{cls}}$ are the learnable parameters for MRI and PET longitudinal classification, and $\text{Softmax}(\cdot)$ denotes softmax activation function. The universal cross-entropy loss for this task is defined as:

$$\mathcal{L}_{\text{cls}} = -\sum_{t=1}^{T} \boldsymbol{m}^{\text{MRI}}_{(t)} \odot \left(\boldsymbol{y}_{(t)} \log\left(\hat{\boldsymbol{y}}_{(t)}\right) + \left(1 - \boldsymbol{y}_{(t)}\right) \log\left(1 - \hat{\boldsymbol{y}}_{(t)}\right) \right) \tag{15}$$

where $\boldsymbol{y}_{(t)}$ and $\hat{\boldsymbol{y}}_{(t)}$ are the true and estimated probabilities of a clinical status' change at tth time point.

Then, conversion prediction task was incorporated into the proposed method. The head number of the second cross-attention block was set as J'. Specifically, the features in all time points were concatenated to prevent the loss of disease progression information at the early time points, where the concatenated features $\tilde{\boldsymbol{H}} = \text{Concat}\left(\tilde{\boldsymbol{H}}_{(1)}, \ldots, \tilde{\boldsymbol{H}}_{(t)}, \ldots, \tilde{\boldsymbol{H}}_{(T)}\right) \in \mathbb{R}^{N\times T\times 2D'}$ were fed into the second cross-attention block to integrate longitudinal and multimodal information. Similar to the first cross-attention block, three matrices, namely, $\hat{\boldsymbol{Q}}^{j'} \in \mathbb{R}^{N\times T\times (2D'/J')}$, $\hat{\boldsymbol{K}}^{j'} \in \mathbb{R}^{N\times T\times (2D'/J')}$, and $\hat{\boldsymbol{V}}^{j'} \in \mathbb{R}^{N\times T\times (2D'/J')}$, were used to calculate the attention matrix $\hat{\boldsymbol{A}}^{j'} \in \mathbb{R}^{N\times T\times (2D'/J')}$ in head j'. After concatenating the attention matrices from J' heads, the concatenated attention matrix was fed into a FC layer with parameters $\boldsymbol{W}_{\text{at}_2} \in \mathbb{R}^{2D'\times 2D'}$ and $\boldsymbol{b}_{\text{at}_2} \in \mathbb{R}^{2D'\times 1}$, and then final features $\hat{\boldsymbol{H}} \in \mathbb{R}^{N\times T\times 2D'}$ for MCI conversion prediction were obtained. Hence, the prediction results were defined as:

$$\hat{\boldsymbol{C}} = \text{softmax}\left(\boldsymbol{W}_{\text{c}} \hat{\boldsymbol{H}} + \boldsymbol{b}_{\text{c}} \right) \tag{16}$$

Table 5
Details of different testing sets.

Denotation	Included Data	Usage	Subject number
ADNI-1/2-A	Longitudinal data (incomplete BL data)	Section 5.3.1	130 MCI
ADNI-1/2-C	Longitudinal data (complete BL data)	Section 5.3.2	81 MCI
ADNI-3-A	BL data (only MRI data)	Section 5.3.1	86 MCI
ADNI-3-C	BL data (complete MRI and PET data)	Section 5.3.2	86 MCI
OASIS-3-A	BL data (only MRI data)	Section 5.3.1	104 MCI
OASIS-3-C	Longitudinal data (complete BL data)	Section 5.3.2	65 CN/6 MCI/7 AD

where $\boldsymbol{W}_c$ and $\boldsymbol{b}_c$ are learnable parameters for MCI conversion prediction. The class imbalance in subjects was serious. Thus, focal cross-entropy loss was applied:

$$\mathcal{L}_{\text{pred}} = -\mu\left(1-\hat{\boldsymbol{C}}\right)^{\gamma}\log\left(\hat{\boldsymbol{C}}\right) \tag{17}$$

where μ and γ are set to 0.3 and 2, respectively.

The overall loss function of our proposed method can be defined as follows:

$$\mathcal{L} = \lambda\mathcal{L}_{\text{est}} + \zeta\mathcal{L}_{\text{adv}} + \xi(\mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{pred}}) \tag{18}$$

where λ, ζ, and ξ are hyperparameters. Hence, our model can be trained in an end-to-end manner, and joint optimization for data imputation, longitudinal classification, and MCI conversion prediction can be achieved.

5. Experiment

5.1. Experimental settings

In this study, longitudinal and multimodal ROI features were used to evaluate the prediction and imputation performances of the proposed method. As described in the "*6. Materials*" section, three ADNI subsets, as well as the OASIS-3 database, were enrolled in our experiments. Moreover, the proposed method was implemented using PyTorch, and all experiments were performed on a server with NVIDIA TITAN X (Pascal) GPU. Moreover, the code of MCNet is publicly available at https://github.com/Meiyan88/MCNET.

In the experiments, a hold-out method was used, and all subjects from the ADNI-1 and ADNI-2 datasets were partitioned into 10 non-overlapping subsets with the same proportion of each class. Among which, eight subsets were applied for training, one was utilized for validation, and one was used for testing. For subjects in the training set, data at all available time points were used to train the networks, whereas only data at BL were used to select the hyperparameters and evaluate the networks for subjects in the validation and testing sets. The data partitioning process was repeated five times, and the results of the validation and testing sets were achieved in each process. The final results for the ADNI-1 and ADNI-2 datasets were obtained from the average of five results in the testing set. The subjects in ADNI-3 had FDG-PET scans at BL but had PET scans with other tracers (e.g., Pittsburgh compound B) at subsequent time points, which means that only the PET data at BL were available in ADNI-3 for testing. Different from ADNI-3, OASIS-3 contained a certain amount of longitudinal FDG-PET data. However, in all subjects containing longitudinal MRI and PET data, subjects that belong to the MCI category were lacking. According to the characteristics of different datasets, they were used in the different experiments for performance assessment. See Table 5 for details.

5.2. Implementation details

To alleviate the computational burden associated with hyperparameter tuning, the optimal combinations of hyperparameters were selected from a pre-defined search range (Huang, Lai, et al., 2021; Zhou et al., 2019). Under the premise of fixing other hyperparameters, certain hyperparameters were adjusted in each iteration. Thus, the hyperparameters utilized in MCNet underwent meticulous tuning in three distinct steps, in each of which the hyperparameters were fixed upon selection and the optimal ones were selected according to the best average value of area under receiver operating characteristic curve (AUC) on the validation set. The steps are as follows: (a) First, the hyperparameters associated with the architecture of data imputation module were selected. Given that the architectures of different modules in MCNet mainly consisted of FC layers, the primary hyperparameters were the layer and node numbers in the hidden layer. After determining the hyperparameters associated with architecture, the data imputation module was trained initially to perform the data imputation task using the corresponding losses listed in Eqs. (9)–(11). Subsequently, relevant hyperparameters within the overall loss function were selected. Specifically, the pre-defined search range are presented as follows: the layer number of MinimalRNN was selected from 1, 2, 3, and 4; the layer number of discriminator was chosen from 2, 3, 4, and 5; the node numbers of the hidden layers were selected from 64 to 512 with interval accumulation as multiples of 2; λ and ζ varied from 0.1, 1, 2, 10, 10^2, 10^3; the number of iterations for the discriminator after each imputation ranged from 1 to 5 with the interval of 1. (b) Second, the hyperparameters of conversion prediction module were selected, involving the head numbers J and J' in two cross-attention blocks. The head numbers were selected from 2, 4, 6, and 8. Notably, the dimensions of projection matrices were obtained by dividing the head number with the node number in the hidden layer. Once the head numbers were determined, the hidden features $\boldsymbol{H}^S$ obtained from the trained data imputation module were fed into the conversion prediction module for pre-training. Consequently, the first cross-attention block was initially pre-trained using the longitudinal classification task and the corresponding loss listed in Eq. (14). Then, the second cross-attention block was pre-trained using conversion prediction task and loss listed in Eq. (16). (c) Third, the final hyperparameter ξ was selected from 1, 2, 10, 10^2, 10^3, and 10^4 to determine the overall model architecture and loss function. Subsequently, the hyperparameters of training settings were selected from pre-defined search range. Adam optimizer was used during the training, and a $\ell 2$-regularization was applied to avoid overfitting. Based on the results of AUC obtained from the validation set, the weight decay coefficient for $\ell 2$-regularization was selected from 5×10^{-6} to 5×10^{-3} with interval accumulation as multiples of 10^{-1}. Similarly, learning rate was selected within the range of 5×10^{-5} to 5×10^{-1} with interval accumulation as multiples of 10^{-1}. Lastly, an end-to-end fine-tuning of the model was conducted using only BL data to obtain the final model.

In summary, two separate stacked three-layer MinimalRNNs were used to extract hidden features, with the node numbers in the hidden layers set to 128. Furthermore, with node numbers set to 90, 128, 64, 2, the discriminator was trained alternately with the remaining parts of the data imputation module. Specifically, after every two discriminator iterations, the module completed one round of data imputation. The hyperparameters λ, ζ, and ξ in the overall loss function were set to 2, 10, and 10, respectively. The selected hyperparameters and corresponding search ranges are listed in Table 6.

Several quantitative metrics were used to evaluate the methods' performance in different tasks. Accuracy (ACC), AUC, and balanced accuracy (BAC) were applied for the prediction task, and MAE and root mean square error (RMSE) were used for the quantitative evaluation of the imputation task. AUC represents the probability that the predicted positive samples are ranked before the negative samples (Huang & Ling, 2005). BAC is the arithmetic mean of specificity and sensitivity (Brodersen, Ong, Stephan, & Buhmann, 2010). Both AUC and BAC are more informative than ACC in reflecting the model performance in class-imbalanced datasets (Brodersen et al., 2010; Lai et al., 2022). Moreover, MAE was used to calculate the average absolute difference between imputed and real data, whereas RMSE reflects the standard deviation of the differences. Although both MAE and RMSE measure the average error of imputed data, MAE presents an unbiased measure of

Table 6
Hyperparameter search space and selected hyperparameters.

Hyperparameters	Search space	Selected value
MinimalRNN layers number	[1, 2, 3, 4]	3
MinimalRNN hidden dimension	[64, 128, 256, 512]	128
Heads number J	[2, 4, 6, 8]	4
Heads number J'	[2, 4, 6, 8]	4
λ	$[0.1, 1, 2, 10, 10^2, 10^3]$	2
ζ	$[0.1, 1, 2, 10, 10^2, 10^3]$	10
ξ	$[1, 2, 10, 10^2, 10^3, 10^4]$	10
Learning rate	$[5\times10^{-5}, 5\times10^{-4}, 5\times10^{-3}, 5\times10^{-2}, 5\times10^{-1}]$	5×10^{-3}
Weight decay	$[5\times10^{-6}, 5\times10^{-5}, 5\times10^{-4}, 5\times10^{-3}]$	5×10^{-4}

Table 7
Ablation experiments of the proposed method on different datasets for MCI conversion prediction.

Method	ADNI-1/2-A			ADNI-3-A			OASIS-3-A			Parameter size (M)	Inference time (ms)
	ACC	AUC	BAC	ACC	AUC	BAC	ACC	AUC	BAC		
MCNet-oLC	0.813 ± 0.023	0.832 ± 0.038	0.795 ± 0.024	0.802	0.806	0.789	0.808	0.825	0.772	2.603	0.131
MCNet-oCB	0.815 ± 0.014	0.826 ± 0.030	0.808 ± 0.011	**0.813**	0.819	0.802	0.817	0.820	0.798	1.337	0.084
MCNet-oDI	0.802 ± 0.029	0.818 ± 0.026	0.798 ± 0.027	0.756	0.765	0.761	0.817	0.806	0.786	2.372	0.057
MCNet-oImp	0.799 ± 0.035	0.794 ± 0.035	0.787 ± 0.037	0.733	0.739	0.748	0.779	0.784	0.750	1.589	0.049
MCNet-oAL	0.812 ± 0.030	0.824 ± 0.024	0.803 ± 0.026	0.791	0.821	0.782	0.798	0.838	**0.801**	2.604	0.134
MCNet	**0.830 ± 0.019**	**0.842 ± 0.032**	**0.813 ± 0.032**	0.802	**0.849**	**0.820**	**0.827**	**0.857**	0.799	2.604	0.134

average error whereas RMSE exhibits bias by assigning greater weight to imputed data with large errors over small ones (Fu, Wu, Ponnarasu, & Zhang, 2023). Paired t-test (at 95% significance level) was conducted for statistical significance test in the prediction task. Finally, parameter sizes and inference time of all compared methods were presented to provide a more comprehensive and deeper understanding of the proposed MCNet.

5.3. Experimental results and analysis

5.3.1. Ablation study

In this section, each of the components in the proposed method is removed separately to investigate its influence on the prediction performance. Table 7 shows the results of all ablation experiments, which are carried out on the ADNI-1 and ADNI-2 testing sets (ADNI-1/2-A) and two other independent testing sets (ADNI-3-A and OASIS-3-A). Five variants of MCNet are included for ablation study, which are denoted as MCNet-oLC, MCNet-oCB, MCNet-oDI, MCNet-oImp, MCNet-oAL, respectively. Specifically, MCNet-oLC discards the longitudinal classification; MCNet-oCB removes cross-attention blocks; MCNet-oDI removes the data imputation task by imputing mean values; MCNet-oImp removes the data imputation task without any imputation; and MCNet-oAL eliminates the adversarial learning strategy.

First, compared with MCNet-oLC, discarding longitudinal classification results in reduced performance in MCI conversion prediction, which indicates the usefulness of multi-task learning strategy. Moreover, removing the longitudinal classification task results in a more pronounced decrease in BAC (1% and 1.8% decrease in AUC and BAC, respectively). The possible reason is that our longitudinal task of predicting whether conversion occurs at each time point is particularly beneficial for identifying pMCI subjects, and thus removing the longitudinal classification task results in severely reduced BAC. Second, the removal of cross-attention blocks in MCNet-oCB leads to the direct concatenation of multimodal features at different time points. Although MCNet-oCB exhibits higher accuracy (81.3% vs. 80.2%) on ADNI-3, MCNet outperforms MCNet-oCB in other metrics, which is primarily due to class imbalance. The inferior performance of MCNet-oCB implies that considering the relationships among different modalities at different time points plays an important role in MCI conversion prediction. Third, when the data imputation task (i.e., the entire data imputation module) is omitted in MCNet-oDI, all missing data are imputed with mean values and the corresponding imputation loss in Eq. (9) is removed. The MCI conversion prediction clearly decreases in performance, proving the effectiveness of using a unified framework for imputation and prediction. Fourth, in MCNet-oImp, the data imputation task is also removed, and the missing data are processed through a masking layer without data imputation as mentioned in Cui et al. (2019). Given the variation in the number of hidden features obtained from different modalities, only first cross-attention block is applied to fuse the last hidden features and use them to achieve conversion prediction. The inferior results of MCNet-oImp compared with MCNet-oDI (ADNI-1/2-A: 79.9% vs. 80.2% in ACC, 79.4% vs. 81.8% in AUC, and 78.7% vs. 79.8% in BAC) demonstrate that the absence of data imputation leads to further performance degradation, which may be attributed to the fact that imputed data ensures alignment of all time-point data for all samples. Additionally, without data imputation, the effectiveness of our cross-attention strategy is partially compromised. Fifth, the performance of MCI conversion prediction decreases when the adversarial learning module is eliminated in MCNet-oAL. Moreover, after the removal of adversarial learning, the imputation errors are tested on ADNI-1/2-A and compared with those generated by MCNet, namely, 0.049 and 0.063 decrease in MAE and RMSE, respectively. Thus, the results prove that the adversarial learning module can help further reduce imputation errors and improve prediction performance. In terms of parameter size and inference time, both the data imputation and disease prediction modules occupy almost an equal portion of the parameters. Additionally, the imputation in the data imputation module and the self-attention mechanism of the cross-attention network in the disease prediction module result in the prolonged inference time of the model. However, overall, the current inference speed and parameter size remain within reasonable limits.

In summary, the proposed method achieves the best prediction performance, indicating that the proposed components are useful for MCI conversion prediction. Moreover, the proposed method achieves AUCs of 0.849 and 0.857 on two independent testing sets under the situation of using only MRI, which demonstrates that imputation from the multimodal view can effectively ensure the prediction performance when only MRI is used at BL. Moreover, the results also prove that our method is flexible in data requirements and can achieve reasonable performance without PET data.

5.3.2. Comparison with other methods

To demonstrate the performance of MCNet, several methods are used for comparison. Among them, four methods that can also be used to deal with incomplete multimodal and longitudinal data are applied to compare with the prediction and imputation performances of MCNet. Furthermore, to prove that MCNet is more effective than the method of using data at BL or cross-sectional data alone, two other methods based on support vector machine (SVM) and MLP are included in the comparison. In addition to using models for imputing, several traditional missing data imputation methods can be used to handle missing data issue. Therefore, two variants of MCNet are included in the comparison (i.e., MCNet-Forward and MCNet-Linear). The brief introductions of different methods are as follows:

- SVM-based method: The classifier is implemented by SVM with linear kernel. All ROI features are first simply concatenated and directly fed into the classifier.
- MLP-based method: The MLP-based method for MCI conversion prediction consists of three MLPs. ROI features of MRI and PET are first fed into two separate MLPs to obtain modality-specific hidden features. Then, the modality-specific hidden features are concatenated and fed into another MLP to obtain the prediction results.
- GRU-D (Che et al., 2018): A GRU-based method that designs a decay mechanism using the information on the interval and location of missing values, and combines decay rates with longitudinal data containing missing values to accomplish classification.
- LSTM-Robust (Ghazi et al., 2019): A robust backpropagation is presented through time algorithm by initializing the missing values of inputs to zero and backpropagating zero errors corresponding to the missing values of outputs when training. This

Table 8
Imputation errors and prediction performance of different methods. The results of ADNI-1/2-C are reported as mean ± std, and * denotes significant difference with p-value < 0.05. Data imputation cannot be implemented in GRU-D, SVM-based method, and MLP-based method, and thus metrics related to imputation performance are represented by '—'.

Method	ADNI-1/2-C							OASIS-3-C				ADNI-3-C			Parameter size (M)	Inference time (ms)
	MAE(MRI)	RMSE(MRI)	MAE(PET)	RMSE(PET)	ACC	AUC	BCA	MAE(MRI)	RMSE(MRI)	MAE(PET)	RMSE(PET)	ACC	AUC	BCA		
SVM-based	–	–	–	–	0.732 ± 0.023	0.777 ± 0.024	0.723 ± 0.043	–	–	–	–	0.720	0.753	0.697	–	0.123
MLP-based	–	–	–	–	0.766 ± 0.027	0.790 ± 0.024	0.738 ± 0.033	–	–	–	–	0.773	0.761	0.755	0.334	0.037
LSTM-Robust	0.686 ± 0.050	1.010 ± 0.065	0.815 ± 0.101	1.049 ± 0.130	0.807 ± 0.021	0.774 ± 0.021	0.774 ± 0.021	0.807	1.027	1.016	1.321	0.760	0.768	0.768	2.979	0.132
GRU-D	–	–	–	–	0.803 ± 0.038	0.816 ± 0.031	0.795 ± 0.026	–	–	–	–	0.773	0.819	0.775	5.135	0.147
BLS-LSTM	0.377 ± 0.051	0.539 ± 0.063	0.483 ± 0.045	0.607 ± 0.097	0.831 ± 0.037	0.827 ± 0.025	0.805 ± 0.028	0.436	0.589	0.667	0.773	0.800	0.834	0.820	5.358	0.096
AJRNN	0.363 ± 0.043	0.511 ± 0.067	0.428 ± 0.036	0.551 ± 0.046	0.831 ± 0.024	0.830 ± 0.016	0.810 ± 0.017	0.398	**0.517**	0.652	0.746	0.800	0.829	**0.828**	4.479	0.153
DRM	0.403 ± 0.042	0.574 ± 0.042	0.530 ± 0.015	0.699 ± 0.015	0.821 ± 0.025	0.819 ± 0.030	0.819 ± 0.030	0.588	0.710	0.781	0.891	**0.813**	0.818	0.813	3.695	0.127
MCNet-Linear	0.379 ± 0.078	0.598 ± 0.077	0.471 ± 0.055	0.607 ± 0.048	0.828 ± 0.012	0.821 ± 0.012	0.803 ± 0.007	0.512	0.690	0.631	0.704	0.773	0.801	0.775	2.604	0.117
MCNet-Forward	0.370 ± 0.039	0.538 ± 0.043	0.516 ± 0.068	0.705 ± 0.029	0.826 ± 0.015	0.829 ± 0.028	0.806 ± 0.007	0.420	0.601	0.677	0.788	0.787	0.804	0.783	2.604	0.117
MCNet*	**0.322 ± 0.034**	**0.468 ± 0.062**	**0.415 ± 0.031**	**0.513 ± 0.043**	**0.842 ± 0.012**	**0.860 ± 0.024**	**0.830 ± 0.011**	**0.372**	0.519	**0.621**	**0.733**	**0.813**	**0.845**	0.821	2.604	0.134

algorithm is used in the missing data estimation of longitudinal data, and a two-stage method is used by performing imputation first and then classification.

- Adversarial Joint-learning RNN (AJRNN) (Ma et al., 2022): An end-to-end model is trained in an adversarial and joint learning manner, which can directly perform classification with missing values and greatly reduce the error propagation from imputation to classification.
- Deep Recurrent Model (DRM) (Jung et al., 2021): A unified framework that applies multivariate and temporal relations inherent in longitudinal and multimodal data to achieve missing value imputation and model disease progression. The prediction result of each subject is obtained using the longitudinal predicted labels acquired from the disease progression task.
- Board learning system and long short-term memory neural network (BLS-LSTM): A method that utilizes board learning system (BLS) to enhance features and then feeds them into an LSTM network for data prediction/imputation. BLS-LSTM can be directly applied to impute ROI features, and hidden features from all time points are concatenated to achieve conversion prediction.
- MCNet-Forward: A forward filling strategy (Nguyen et al., 2020) is utilized, missing values are imputed with the available data of the last previous time point. Specifically, other components except for the imputation strategy proposed in our method are retained in MCNet-Forward.
- MCNet-Linear: A linear filling strategy (Nguyen et al., 2020) is applied to impute missing data by using the available data between the previous and the next time point. Moreover, if there is no future observed data for linear filling, then forward filling is utilized. Other settings are consistent with MCNet-Forward.

Different from the proposed method, complete multimodal data at BL are required for the compared methods (i.e., GRU-D, LSTM-Robust, AJRNN, and DRM). Therefore, the proposed method was also performed on the same subject number (i.e., ADNI-1/2-C, ADNI-3-C, and OASIS-3-C) used in the compared methods for a fair comparison. Similar to the proposed method, only multimodal data at BL were included in the validation and testing sets for the compared methods. Specially, same longitudinal data as other methods were adopted for the training and testing of SVM- and MLP-based methods, and each time point of each subject can be treated as a separate subject when training. Moreover, the missing PET data were filled by the mean value at each category at the training process of SVM and MLP. For all compared methods, a hold-out method was used, and the hyperparameters were turned carefully according to their corresponding papers to make a fair comparison. Furthermore, a similar training strategy was adopted across all methods to ensure consistency. This strategy primarily involves utilizing training sets that exclusively retain BL data for fine-tuning in the final stage.

Data imputation may affect the extracted features for final MCI conversion prediction. Thus, the imputation errors are quantitatively analyzed on different data imputation methods (i.e., LSTM-Robust, AJRNN, DRM, MCNet, MCNet-Forward, and MCNet-Linear). During the process of data imputation, all data were estimated regardless of whether the data were missing or not; thus, the data without any missing data can be used as ground truth to evaluate imputation errors. As shown in Table 8, AJRNN and DRM have better imputation results than LSTM-Robust, which indicates that the model based on zero initialization may not be suitable for our data. Methods with better imputation performance also exhibit better performance in conversion prediction, indicating that imputation performance is a critical factor in determining prediction performance. In comparison with the compared methods, a multi-view imputation combined with adversarial learning strategy is designed in the proposed method to handle missing data issues in various scenarios. The most important advantage of the proposed method is its ability to deal with missing PET data at BL. Moreover, the inclusion of adversarial learning makes the data distribution close to the real one, thereby further reducing imputation errors and improving the prediction performance. Compared with two variants of MCNet, our proposed method achieves better results, which indicates that the strategy of data imputation via model learning is more effective than traditional missing data imputation methods. Therefore, the proposed method achieves significantly accurate imputation values compared with the other methods (p-value < 0.05).

The main goal of this study is MCI conversion prediction (i.e., pMCI vs. sMCI) and the performance of all methods are presented in Table 8 with the following findings: First, the results of SVM and MLP are generally lower than those of other RNN-based methods, which demonstrates the feasibility of using longitudinal data during training phase to capture disease progression information and to improve prediction performance at BL. Second, except for SVM and MLP, the lowest MCI conversion prediction results are found in LSTM-Robust. The possible reason is that LSTM-Robust is a two-stage method, in which data imputation and MCI conversion prediction are performed separately, and thus suboptimal results may be obtained. Moreover, poor imputation performance results in degraded prediction performance in LSTM-Robust. Third, the prediction performances of AJRNN and DRM are better than that of GRU-D, which indicates the effectiveness of incorporating data imputation and conversion prediction into a unified framework. Fourth, BLS-LSTM achieves a higher prediction accuracy than DRM, in which the interval information of missing data is applied to assist in imputation. However, when only BL data are present, the interval information is lacking and may lead to poor imputation results and inferior prediction performance in DRM. Fifth, a comparison with AJRNN and MCNet reveals that BLS-LSTM performs worse in both imputation and prediction. This finding underscores the potential of adversarial learning in further enhancing data imputation and prediction, particularly in situations involving solely BL data. Sixth, results show that the proposed method achieves the best performance. The proposed and benchmark methods show a significant difference (p-value < 0.05), which implies that the proposed method can be used as a practical and general learning framework for MCI conversion prediction on incomplete longitudinal and multimodal data. Furthermore, the results of MCNet on ADNI-3-A (Table 7) and ADNI-3-C (Table 8) shows only minor differences, confirming that MCNet can achieve reasonable performance when only MRI data are available at BL. In terms of parameter size, MCNet holds an advantage over most of the compared methods (except for MLP-based method) due to its MinimalRNN-based backbone network with an inherently small parameter size. This reduction in

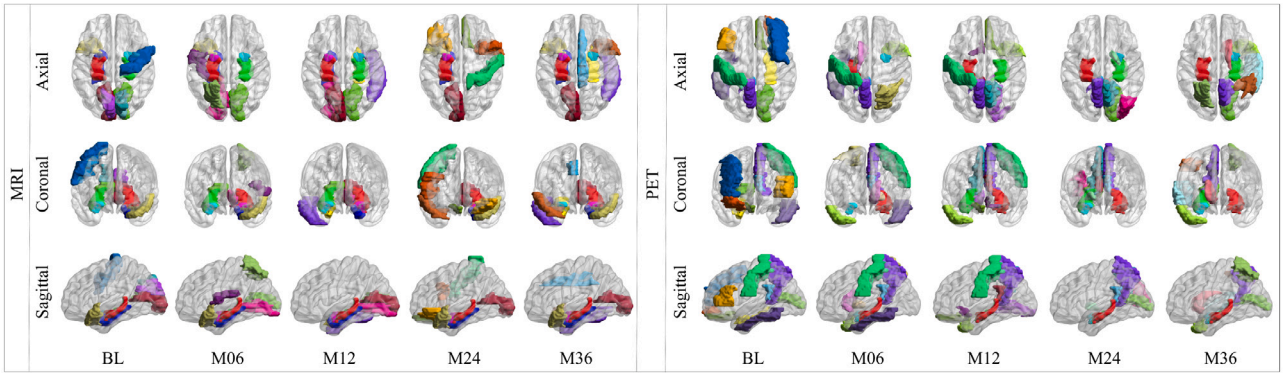


Fig. 3. Visualization results of the top-10 ROIs of different modalities at each time point, where M06, M12, M24, and M36 represent 6, 12, 24, and 36 months, respectively.

parameters helps prevent model overfitting. In terms of inference time, our approach remains comparable to other methods and maintains a similar level of efficiency.

5.3.3. Interpretability of the proposed method

The interpretability of the model is crucial for clinical prediction and can help us discover some potential information associated with AD. Therefore, the 10 most discriminative ROIs of different modalities at different time points are illustrated, and their corresponding interpretations are provided.

We introduce a gradient-based computation strategy that compute the contribution of each ROI to longitudinal classification at each time point to locate the most discriminative ROIs (Huang, Lai, et al., 2021). Therefore, based on the contribution values, we screen out the top-10 ROIs of different modalities at different time points. Specifically, for the rth ROI of modality S at the tth time point $\boldsymbol{X}^{S}_{(t)}(r) \in \mathbb{R}^{N\times 1\times 1}$, the derivatives of the predicted probability $\hat{\boldsymbol{y}}_{(t)}$ of the subjects with AD with regard to $\boldsymbol{X}^{S}_{(t)}(r)$ is obtained from the longitudinal classification task, and the absolute value of the derivative among all subjects is averaged as the contribution. The visualization results of the top-10 ROIs of different modalities at each time point are shown in Fig. 3. For MRI, the hippocampus, parahippocampal gyrus, and amygdala, which are highly correlated with memory, are detected at each time point. Higher contribution values are achieved at 6 and 12 months than at other time points. Moreover, the volume atrophy of these three ROIs is associated with healthy aging and different stages of AD (Teipel et al., 2006). On the contrary, the parahippocampal gyrus and amygdala are detected at the first two time points, and the hippocampus is detected at the last four time points for PET. Besides, the temporal pole, which is linked to visual cognition (Herlin, Navarro, & Dupont, 2021), is also detected at most time points in MRI instead of PET. The posterior cingulate gyrus is an important area detected by PET, and the remarkable metabolism reduction in the region is associated with memory impairment, which is a feature of early AD (Minoshima et al., 1997). Furthermore, the precuneus, which is associated with a high level of cognitive function (Cavanna & Trimble, 2006), is explored by PET data at most time points. Moreover, the contributions of the detected ROIs varies across different time points and are also worthy of further study.

6. Discussion

6.1. Effect of hyperparameters

In this section, we carry out a comprehensive analysis of the influence of hyperparameters on the MCNet. Specifically, we examined the influence of the hyperparameters listed in Table 6 on the prediction performance. For each iteration, we systematically adjusted a pair of hyperparameters while keeping the others fixed at their optimal values. Moreover, the pre-defined ranges of the hyperparameters are elaborated upon in "*5.2. Implementation Details*" section. Fig. 4 shows the AUC of conversion prediction achieved by MCNet with different hyperparameters. On the basis of the results, we observed that excessively high values of λ (>10) and ζ (>10) significantly deteriorate the prediction performance. This decline can be attributed to the model overly prioritizing the data imputation task, thereby neglecting the optimization for the prediction task. Hence, to maintain an appropriate balance between the ratios of λ, ζ, and ξ is essential. Furthermore, in configuring an excessive head number for cross-attention blocks, a large learning rate or weight decay value can have an adverse effect on network optimization causing the reduced prediction performance. The optimal hyperparameters for MCNet have been summarized in Table 6.

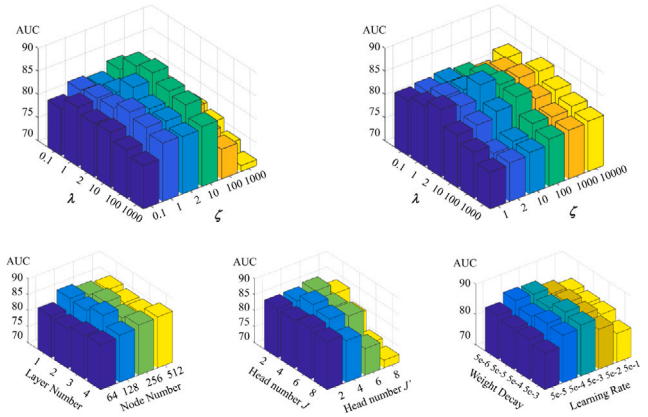


Fig. 4. Variations in performance of conversion prediction under different hyperparameters.

6.2. Effect of training set size

In this section, we investigate the performance variations of MCNet under different training set sizes. While keeping the testing sets unchanged, we evaluated the performance of MCNet using different proportions of the total training sets from 100% down to 30%. Specifically, we discuss the performance of different trained models in two scenarios; one including both MRI and PET data (ADNI-1/2-C) and the other including only MRI data (ADNI-3-A). Fig. 5 presents the corresponding results, where a decrease in the training set size leads to a consistent decline in the overall model performance. However, the reduction in the training set size results in a faster decline in model performance when using a dataset composed solely of MRI data (ADNI-3-A) compared with using datasets that simultaneously include both MRI and PET data (ADNI-1/2-C). The extreme case of having only PET data is more reliant on the assurance of data imputation

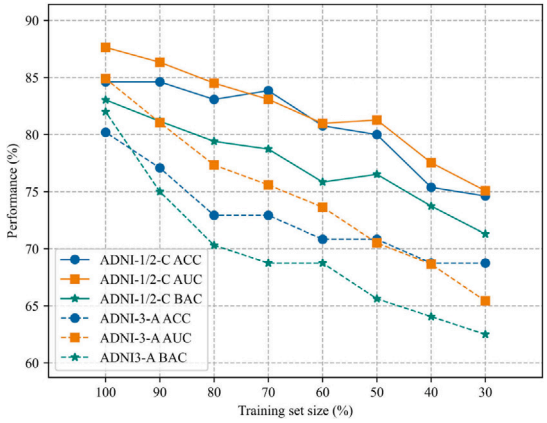


Fig. 5. Variations in performance of conversion prediction under different training set sizes.

capability, thereby potentially increasing the sensitivity to changes in the training set size. This observation may be ascribed to the fact that such extreme cases place much emphasis on the reliability of data imputation, which consequently exhibits a heightened sensitivity to variations in the training set size. Furthermore, the model trained with 50% of the training set size outperforms the one trained with 60% on certain metrics (AUC and BAC on ADNI-1/2-C, ACC on ADNI-3-C). This discrepancy in performance can be attributed to the randomness involved in the data removal during down-sampling, where certain samples with a negative effect on training may be unintentionally excluded. In summary, with the availability of MRI and PET data at BL, training the model using only 50% of the training set size yields an AUC performance exceeding 80%.

6.3. Comparison with other methods

Conventional methods that use longitudinal data modeling primarily concentrate on utilizing all available data to predict the clinical status at the subsequent time point. The subsequent time point refers to the next time point following the last time point in the longitudinal data. However, our specific emphasis lies in leveraging disease progression information from longitudinal data to facilitate early-stage conversion prediction. Thus, approaches such as GRU-D, DRM, and LSTM-Robust may not be well-suited for our specific emphasis. GRU-D and DRM utilize the time intervals between adjacent available data and locations of missing data to assist in imputation or prediction. However, this mechanism leads to uniformity of the above information across all samples when only BL data are available, rendering this strategy ineffective. For LSTM-Robust, the loss computation and gradient propagation are both dependent on the quantity of missing data and aim to mitigate its adverse effects. However, under the extreme scenario where only BL data are available, this strategy becomes nearly ineffective. In BLS-LSTM, the BLS technique can be used to enhance features from all time points, and thus also has certain effects when only BL data are available. However, BLS-LSTM does not incorporate any additional specialized strategies, resulting in its performance being lower than those of AJRNN and MCNet. By contrast, AJRNN and the proposed MCNet achieve higher AUC values than the previous methods (83.0% in AJRNN and 86.0% in MCNet vs. 77.4% in LSTM-Robust, 81.6% in GRU-D, 82.7% in BLS-LSTM, and DRM: 81.9%). This improvement can be attributed to the adversarial imputation strategy in AJRNN and MCNet, which effectively reduce the errors caused by continuous forward imputation when only BL data are present. Another advantage of MCNet lies in ensuring the maximum amount of training data, which is primarily achieved through the proposed multi-view imputation strategy. The quantity of data directly affects the model generalization and reliability. Given an equal amount of training data, MCNet outperforms AJRNN in handling multimodal data (ACC: 84.2% vs. 83.1%; AUC: 86.0% vs. 83.1%; BAC: 83.0% vs. 81.0%), primarily due to the incorporation of cross-attention blocks. Specifically, the self-attention mechanism used in the two cross attention blocks helps MCNet fully exploit the relationships between modalities and time points, respectively.

6.4. Limitations and future works

Several issues must be addressed in future research. First, although 1530 subjects are included for training in this study, this number remains insufficient to fully exploit the potential of deep learning. Moreover, a limited sample size may lead to model overfitting. In the future, more samples can be collected through collaboration with clinicians instead of using publicly available datasets. Second, our model is built based on the ROI features extracted from neuroimages, resulting in the loss of spatial information of ROIs. Such information must be introduced in future work.

Only PET and MRI data are included for analysis in this study. Increasing neuroimaging modalities, such as functional MRI (fMRI) and diffusion tensor imaging (DTI), are proven to be effective for AD diagnosis. In fMRI, the prevalent analysis approach involves constructing functional connectivity networks, and subjects undergoing the conversion to AD demonstrate aberrant alterations in specific functional connections (Liebe et al., 2022). In DTI, structural differences in white matter are observed, which can be effective in conversion prediction (Velazquez & Lee, 2022). Hence, an interesting topic for study is the efficient integration of information from other modalities into our framework. One straightforward approach is to extend MCNet with additional stacked MinimalRNNs to analyze more than two modalities. Another reasonable direction is to establish a graph-based analytical model. Functional connectivity networks constructed from fMRI data are highly compatible for graph analysis. ROI features generated based on brain regions are also well-suited as node features in the graph. Furthermore, the construction of graphs partially addresses the challenge of losing spatial information.

7. Conclusion

In this study, we propose an end-to-end multi-task deep learning framework for MCI conversion prediction. A multi-view imputation method combined with adversarial learning is developed for incomplete longitudinal and multimodal data to handle missing data. Moreover, cross-attention blocks are introduced to explore crucial information of different modalities at different time points, which can contribute to the achievement of accurate MCI conversion prediction. The proposed method is trained on two ADNI datasets with 1301 subjects. Moreover, two independent testing sets are applied to further evaluate the generalization ability of the proposed method. Based on the experiments, the proposed method achieves high accuracy in missing data imputation and MCI conversion prediction and performs well when only MRI data were available at BL. To our best knowledge, no research has combined longitudinal and multimodal associations to achieve multi-view adversarial imputation at different time points with small errors, performed AD classification and prediction in the same framework for the joint optimization of MCI conversion prediction, and achieved satisfying the results of MCI conversion prediction using only single-modal data at BL during testing. Therefore, the proposed method may be a crucial tool for MCI conversion prediction, diagnosis, and monitoring.

CRediT authorship contribution statement

Tao Wang: Investigation, Writing – original draft, Writing – review & editing, Methodology, Validation. **Xiumei Chen:** Investigation, Visualization, Validation, Writing – review & editing. **Xiaoling Zhang:** Investigation, Visualization, Data curation. **Shuoling Zhou:** Formal analysis, Data curation. **Qianjin Feng:** Resources, Conceptualization, Funding acquisition, Project administration, Supervision. **Meiyan Huang:** Methodology, Conceptualization, Funding acquisition, Project administration, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in the experiments were the ADNI and OASIS public databases.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 82272069, No. 81601562, No. 81974275, and No. 12126603), the Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515012011), and the Science and Technology Planning Project of Guangzhou (No. 201904010417). Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) and the Open Access Series of Imaging Studies (OASIS) database (oasis-brains.org). The investigators within the ADNI and OASIS contributed to the design and implementation of ADNI and OASIS, but did not participate in the analysis or writing of this paper.

References

Abdelaziz, M., Wang, T., & Elazab, A. (2021). Alzheimer's disease diagnosis framework from incomplete multimodal data using convolutional neural networks. *Journal of Biomedical Informatics*, *121*, Article 103863.

Abdelnour, C., Agosta, F., Bozzali, M., Fougère, B., Iwata, A., Nilforooshan, R., et al. (2022). Perspectives and challenges in patient stratification in Alzheimer's disease. *Alzheimer's Research & Therapy*, *14*(1), 1–12.

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279.

Anandkumar, A., Ge, R., Hsu, D., Kakade, S. M., & Telgarsky, M. (2014). Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, *15*, 2773–2832.

Arco, J. E., Ramírez, J., Górriz, J. M., Ruz, M., Alzheimer's Disease Neuroimaging Initiative, et al. (2021). Data fusion based on searchlight analysis for the prediction of Alzheimer's disease. *Expert Systems with Applications*, *185*, Article 115549.

Behrad, F., & Abadeh, M. S. (2022). An overview of deep learning methods for multimodal medical data mining. *Expert Systems with Applications*, *200*, Article 117006.

Bengio, S., Vinyals, O., Jaitly, N., & Shazeer, N. (2015). Scheduled sampling for sequence prediction with recurrent neural networks. In *2015 Advances in neural information processing systems 28* (pp. 1171—1179).

Brand, L., Nichols, K., Wang, H., Shen, L., & Huang, H. (2019). Joint multi-modal longitudinal regression and classification for Alzheimer's disease prediction. *IEEE Transactions on Medical Imaging*, *39*(6), 1845–1855.

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In *2010 20th International conference on pattern recognition* (pp. 3121–3124). IEEE.

Bucholc, M., Titarenko, S., Ding, X., Canavan, C., & Chen, T. (2023). A hybrid machine learning approach for prediction of conversion from mild cognitive impairment to dementia. *Expert Systems with Applications*, *217*, Article 119541.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564–583.

Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, *8*(1), 1–12.

Chen, M. (2017). Minimalrnn: Toward more interpretable and trainable recurrent neural networks. In *2017 Advances in neural information processing systems 30 symposium on interpretable machine learning*.

Chen, X., Wang, T., Lai, H., Zhang, X., Feng, Q., & Huang, M. (2022). Structure-constrained combination-based nonlinear association analysis between incomplete multimodal imaging and genetic data for biomarker detection of neurodegenerative diseases. *Medical Image Analysis*, *78*, Article 102419.

Cui, R., Liu, M., Alzheimer's Disease Neuroimaging Initiative, et al. (2019). RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Computerized Medical Imaging and Graphics*, *73*, 1–10.

El-Sappagh, S., Abuhmed, T., Islam, S. R., & Kwak, K. S. (2020). Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. *Neurocomputing*, *412*, 197–215.

Fu, X., Wu, M., Ponnarasu, S., & Zhang, L. (2023). A hybrid deep learning approach for dynamic attitude and position prediction in tunnel construction considering spatio-temporal patterns. *Expert Systems with Applications*, *212*, Article 118721.

Ganguli, M., Jia, Y., Hughes, T. F., Snitz, B. E., Chang, C.-C. H., Berman, S. B., et al. (2019). Mild cognitive impairment that does not progress to dementia: a population-based study. *Journal of the American Geriatrics Society*, *67*(2), 232–238.

Gaugler, J., James, B., Johnson, T., Reimer, J., Solis, M., Weuve, J., et al. (2022). 2022 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *18*(4), 700–789.

Ghazi, M. M., Nielsen, M., Pai, A., Cardoso, M. J., Modat, M., Ourselin, S., et al. (2019). Training recurrent neural networks robust to incomplete data: application to Alzheimer's disease progression modeling. *Medical Image Analysis*, *53*, 39–46.

Herlin, B., Navarro, V., & Dupont, S. (2021). The temporal pole: From anatomy to function—A literature appraisal. *Journal of Chemical Neuroanatomy*, *113*, Article 101925.

Huang, M., Chen, X., Yu, Y., Lai, H., & Feng, Q. (2021). Imaging genetics study based on a temporal group sparse regression and additive model for biomarker detection of Alzheimer's disease. *IEEE Transactions on Medical Imaging*, *40*(5), 1461–1473.

Huang, M., Lai, H., Yu, Y., Chen, X., Wang, T., Feng, Q., et al. (2021). Deep-gated recurrent unit and diet network-based genome-wide association analysis for detecting the biomarkers of Alzheimer's disease. *Medical Image Analysis*, *73*, Article 102189.

Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *17*(3), 299–310.

Huang, M., Yang, W., Feng, Q., & Chen, W. (2017). Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease. *Scientific Reports*, *7*(1), 1–13.

Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., et al. (2019). Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping*, *40*(17), 4952–4964.

Jagust, W. J., Landau, S. M., Koeppe, R. A., Reiman, E. M., Chen, K., Mathis, C. A., et al. (2015). The Alzheimer's disease neuroimaging initiative 2 PET core: 2015. *Alzheimer's & Dementia*, *11*(7), 757–771.

Jung, W., Jun, E., Suk, H.-I., Initiative, A. D. N., et al. (2021). Deep recurrent model for individualized prediction of Alzheimer's disease progression. *NeuroImage*, *237*, Article 118143.

Khan, A., & Zubair, S. (2022). An improved multi-modal based machine learning approach for the prognosis of Alzheimer's disease. *Journal of King Saud University-Computer and Information Sciences*, *34*(6), 2688–2706.

Kikuchi, M., Kobayashi, K., Itoh, S., Kasuga, K., Miyashita, A., Ikeuchi, T., et al. (2022). Identification of mild cognitive impairment subtypes predicting conversion to Alzheimer's disease using multimodal data. *Computational and Structural Biotechnology Journal*, *20*, 5296–5308.

Ko, W., Jung, W., Jeon, E., & Suk, H.-I. (2022). A deep generative–discriminative learning for multi-modal representation in imaging genetics. *IEEE Transactions on Medical Imaging*, *41*(9), 2348–2359.

Lai, H., Fu, S., Zhang, J., Cao, J., Feng, Q., Lu, L., et al. (2022). Prior knowledge-aware fusion network for prediction of macrovascular invasion in hepatocellular carcinoma. *IEEE Transactions on Medical Imaging*, *41*(10), 2644–2657.

Lee, G., Kang, B., Nho, K., Sohn, K.-A., & Kim, D. (2019). MildInt: deep learning-based multimodal longitudinal data integration framework. *Frontiers in Genetics*, *10*, 617.

Leng, Y., Cui, W., Peng, Y., Yan, C., Cao, Y., Yan, Z., et al. (2023). Multimodal cross enhanced fusion network for diagnosis of Alzheimer's disease and subjective memory complaints. *Computers in Biology and Medicine*, Article 106788.

Liebe, T., Dordevic, M., Kaufmann, J., Avetisyan, A., Skalej, M., & Müller, N. (2022). Investigation of the functional pathogenesis of mild cognitive impairment by localisation-based locus coeruleus resting-state fMRI. *Human Brain Mapping*, *43*(18), 5630–5642.

Liu, Y., Yue, L., Xiao, S., Yang, W., Shen, D., & Liu, M. (2022). Assessing clinical progression from subjective cognitive decline to mild cognitive impairment with incomplete multi-modal neuroimages. *Medical Image Analysis*, *75*, Article 102266.

Lo, R. Y., & Jagust, W. J. (2012). Predicting missing biomarker data in a longitudinal study of Alzheimer disease. *Neurology*, *78*(18), 1376–1382.

Luo, M., He, Z., Cui, H., Chen, Y.-P. P., Ward, P., Alzheimer's Disease Neuroimaging Initiative, et al. (2023). Class activation attention transfer neural networks for MCI conversion prediction. *Computers in Biology and Medicine*, *156*, Article 106700.

Ma, Q., Li, S., & Cottrell, G. W. (2022). Adversarial joint-learning recurrent neural network for incomplete time series classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(4), 1765–1776.

Minoshima, S., Giordani, B., Berent, S., Frey, K. A., Foster, N. L., & Kuhl, D. E. (1997). Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *42*(1), 85–94.

Nguyen, M., He, T., An, L., Alexander, D. C., Feng, J., Yeo, B. T., et al. (2020). Predicting Alzheimer's disease progression using deep recurrent neural networks. *NeuroImage*, *222*, Article 117203.

Ning, Z., Xiao, Q., Feng, Q., Chen, W., & Zhang, Y. (2021). Relation-induced multi-modal shared representation learning for Alzheimer's disease diagnosis. *IEEE Transactions on Medical Imaging*, *40*(6), 1632–1645.

Pan, Y., Chen, Y., Shen, D., & Xia, Y. (2021). Collaborative image synthesis and disease diagnosis for classification of neurodegenerative disorders with incomplete multi-modal neuroimages. In *2021 24th International conference on medical image computing and computer-assisted intervention* (pp. 480–489). Springer.

Petersen, R. C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., & Fratiglioni, L. (2014). Mild cognitive impairment: a concept in evolution. *Journal of Internal Medicine*, *275*(3), 214–228.

Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., et al. (2021). Alzheimer's disease. *The Lancet*, *397*(10284), 1577–1590.

Shi, J., Zheng, X., Li, Y., Zhang, Q., & Ying, S. (2017). Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, *22*(1), 173–183.

Teipel, S. J., Pruessner, J. C., Faltraco, F., Born, C., Rocha-Unold, M., Evans, A., et al. (2006). Comprehensive dissection of the medial temporal lobe in AD: measurement of hippocampus, amygdala, entorhinal, perirhinal and parahippocampal cortices using MRI. *Journal of Neurology*, *253*(6), 794–800.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *2017 Advances in neural information processing systems 30* (pp. 5998–6008).

Velazquez, M., & Lee, Y. (2022). Multimodal ensemble model for alzheimer's disease conversion prediction from early mild cognitive impairment subjects. *Computers in Biology and Medicine*, *151*, Article 106201.

Venugopalan, J., Tong, L., Hassanzadeh, H. R., & Wang, M. D. (2021). Multimodal deep learning models for early detection of Alzheimer's disease stage. *Scientific Reports*, *11*(1), 1–13.

Wang, T., Qiu, R. G., & Yu, M. (2018). Predictive modeling of the progression of Alzheimer's disease with recurrent neural networks. *Scientific Reports*, *8*(1), 9161.

Zhang, J., Wu, J., Li, Q., Caselli, R. J., Thompson, P. M., Ye, J., et al. (2021). Multi-resemblance multi-target low-rank coding for prediction of cognitive decline with longitudinal brain images. *IEEE Transactions on Medical Imaging*, *40*(8), 2030–2041.

Zhang, C., Zhao, S., & He, Y. (2021). An integrated method of the future capacity and RUL prediction for lithium-ion battery pack. *IEEE Transactions on Vehicular Technology*, *71*(3), 2601–2613.

Zhao, S., Zhang, C., & Wang, Y. (2022). Lithium-ion battery capacity and remaining useful life prediction using board learning system and long short-term memory neural network. *Journal of Energy Storage*, *52*, Article 104901.

Zhou, T., Thung, K.-H., Zhu, X., & Shen, D. (2019). Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human Brain Mapping*, *40*(3), 1001–1016.

Zhu, W., Sun, L., Huang, J., Han, L., & Zhang, D. (2021). Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI. *IEEE Transactions on Medical Imaging*, *40*(9), 2354–2366.

Zu, C., Wang, Y., Zhou, L., Wang, L., & Zhang, D. (2018). Multi-modality feature selection with adaptive similarity learning for classification of Alzheimer's disease. In *2018 IEEE 15th international symposium on biomedical imaging* (pp. 1542–1545). IEEE.