



A novel conversion prediction method of MCI to AD based on longitudinal dynamic morphological features using ADNI structural MRIs

Man Guo¹ · Yongchao Li¹ · Weihao Zheng² · Keman Huang³ · Li Zhou⁴ · Xiping Hu^{1,5} · Zhijun Yao¹ · Bin Hu¹

Received: 18 December 2019 / Revised: 3 May 2020 / Accepted: 5 May 2020 / Published online: 4 June 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Mild cognitive impairment (MCI) is a pre-existing state of Alzheimer's disease (AD). An accurate prediction on the conversion from MCI to AD is of vital clinical significance for potential prevention and treatment of AD. Longitudinal studies received widespread attention for investigating the disease progression, though most studies did not sufficiently utilize the evolution information. In this paper, we proposed a cerebral similarity network with more progression information to predict the conversion from MCI to AD efficiently. First, we defined the new dynamic morphological feature to mine longitudinal information sufficiently. Second, based on the multiple dynamic morphological features the cerebral similarity network was constructed by sparse regression algorithm with optimized parameters to obtain better prediction performance. Then, leave-one-out cross-validation and support vector machine (SVM) were employed for the training and evaluation of the classifiers. The proposed methodology obtained a high accuracy of 92.31% (Sensitivity = 100%, Specificity = 82.86%) in a three-year ahead prediction of MCI to AD conversion. Experiment results suggest the effectiveness of the dynamic morphological feature, serving as a more sensitive biomarker in the prediction of MCI conversion.

Keywords Mild cognitive impairment · Dynamic morphological features · Elastic network · Magnetic resonance imaging

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00415-020-09890-5>) contains supplementary material, which is available to authorized users.

✉ Xiping Hu
huxp@lzu.edu.cn; xp.hu@siat.ac.cn

✉ Zhijun Yao
yaozj@lzu.edu.cn

✉ Bin Hu
bh@lzu.edu.cn

¹ College of Information Science and Engineering, Lanzhou University, Lanzhou, China

² College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

³ MIT Sloan School of Management, Cambridge, US

⁴ National University of Defense Technology, Changsha, China

⁵ Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Introduction

Mild cognitive impairment (MCI) is an intermediate state between normal aging and dementia [1]. Studies have shown that MCI subjects tend to progress to probable Alzheimer's disease (AD) at a rate of 10–15% each year [2]. MCI can be divided into two subtypes, converting MCI and non-converting MCI. The MCI converter (MCI-C) indicates the group of patients who is likely to progress to AD in a short period of time, but the MCI non converter (MCI-NC) remains stable for a certain period of time, with smaller risk of conversion to AD than the former [3]. Identifying these two different kinds of MCIs can predict the conversion from MCI to AD as early as possible, which is helpful for the prevention and treatment of AD.

Neuroimaging has been proven to be a useful tool to understand the pathology of AD and MCI as well as many medical assistant diagnosis systems based on which have been applied in clinical practice [4]. In previous AD mechanism studies, structural magnetic resonance imaging (s-MRI) was the most widely used imaging tool in AD detection and prediction, with wide practicality, high

diagnostic accuracy, and moderate cost [5]. However, s-MRI can only reflect the current state of the structure and further mining of s-MRI information will enhance its ability of clinical application. Many cross-sectional studies obtained classification accuracy of more than 95% which has achieved practical results in the diagnosis and identification of AD [6, 7]. Unfortunately, this method is not so effective in MCI conversion prediction. Because MCI is a state of conversion process, disease progression over time will be more indicative than static assessment using a snapshot [8, 9]. Longitudinal information has been added to the scope of research, while the existing longitudinal studies mostly selected data based on the longitudinal criteria and used only baseline data for further analysis [10, 11]. However, the progression information contained in follow-up time points, which is of great importance for studying progressive diseases [8], is not utilized effectively. Therefore, taking the evolution information into consideration is critical to improve the performance for MCI conversion prediction.

In addition to the original features acquired by MRI, brain structural network measures which are also referred to as anatomical connection patterns between different brain regions [12], providing new insights into brain network organization, topology, and complex dynamics, as well as a further understanding of the pathogenesis of the disease [13–15]. Previous studies show the significant differences in the measures of structural network between MCI subjects and Normal Controls (NC) [16, 17]. In the brain network based on cortical thickness structure, MCI group shows the decreased nodal centrality in the left lingual gyrus, middle temporal gyrus (MTG) and the increased nodal centrality in the precuneus cortex compared with NC, in which the nodal centrality is estimated by the betweenness, measuring the importance of a node in the network [16]. Similarly, the network properties such as small world attributes, local efficiency and degree have also greatly improved the predictive performance of MRI images [17]. In the studies of brain structure, the methods summarized the anatomical development of the entire brain into several scalar measures, such as cortical thickness, cortical volume and hippocampus, etc. [18, 19]. Researchers studied the morphologic changes of AD and MCI by voxel-based analysis [20, 21], or through analyzing region of interests (ROI) which is more targeted [15, 18, 19]. However, previous studies usually explored only a single measure (volume and cortical thickness) or several ROIs, ignoring the multiple morphological changes in the whole brain with the progress of AD [22, 23]. For these ROIs, research on their similar variation from a network perspective is focused rarely, though existing studies showed that many brain areas have the similar abnormal morphological patterns as the progression of AD [15, 24]. Recently, sparse learning techniques have attracted

increasing attention due to their excellent performances in a series of neuroimaging applications on different modalities [25, 26]. A voxel-based sparse classifier on basis of a l_1 -norm regularized linear regression model is employed to classify AD and MCI which achieves significant results [27]. Elastic networks are used to identify neuroimaging and proteomic biomarkers of AD and MCI [28].

This paper proposes a novel method to address the above problems, namely: limited evolution information, lacking similar variation research, and single measure studies by constructing longitudinal dynamic multi-morphological network (LDMN) using s-MRI data. Our hypothesis is that the dynamic morphological feature is a more sensitive biomarker and the sparse regression cerebral network constructed by the multiple dynamic morphological features can achieve a better performance in MCI conversion prediction. We highlight the contribution of this paper as follows: (1) Based on the longitudinal data to further mine the image information of s-MRI, we defined the dynamic morphological feature benefiting from the progression information. (2) We selected the subjects with 4 years follow-up and chose four time points without conversion, extracting five cerebral cortex measures closely related to AD progression and calculating their dynamic features. (3) We adopted LASSO and elastic network to perform sparse regression on the multiple dynamic morphological features to construct the cerebral similarity network at different phases, employing the two-level tuning method to optimize parameters of sparse regression to make more excellent prediction results. Finally, we classified MCI-C and MCI-NC based on the commonly used network attributes, utilizing the method combining leave-one-out cross-validation and support vector machine (SVM) to train and evaluate the classifiers.

The remainder of this paper mainly includes the following parts. Section "Material" illustrates the selection of the subjects and the pre-processing of the images. Section "Methods" describes the definition and calculation of the dynamic morphological feature, the methods of network construction and classification. Section "Results" reports the experimental results. Section "Discussions" discusses the achieved results and the clinical significance. Finally, Sect. "Conclusion" gives the conclusion.

Material

Image data

The subjects were obtained from the publicly open Alzheimer's disease neuroimaging initiative (ADNI) database (<https://adni.loni.usc.edu/>) [29]. The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the

US Food and Drug Administration, private pharmaceutical companies, and several non-profit organizations [25].

MCI is determined based on routine cognitive behavioral tests such as Mini-Mental-State -Examination (MMSE) scores, between 24 and 30, and Clinical Dementia Rating (CDR) score of 0.5. Using month18 as a reference point, subjects converted to AD before month 18 (including month 18) were excluded. The MCI-C included subjects who converted to AD between month 18 and month 48. Similarly, subjects were marked as MCI-NC once no conversion occurred in the period. We selected subjects with s-MRI data at four longitudinal time points without conversion, baseline (BL), 6th month (M6), 12th month (M12) and 18th month (M18). Seventy-eight participants were chosen, including 35 (75.3 ± 6.7) MCI-NC and 43 (73.6 ± 7.6) MCI-C. Their MMSE scores were from 24 to 30 and the score of CDR was 0.5. Details of the characteristics of the subjects selected are presented in Table 1. To evaluate the specific effects that age and gender might have on the classification results, we performed the Two-sample T test on age and the Chi-square test on gender, and found no statistical differences ($p = 0.31 > 0.05$; $p = 0.61 > 0.05$) [30].

Pre-processing

The images of the subjects were acquired by the T1 scanner from the ADNI library. We chose the Neuroimaging Informatics Technology Initiative (NIfTI) format, which had undergone spatial distortion correction processing due to gradient nonlinearity and B1 field non-uniformity.

All the structural images were pre-processed with FreeSurfer v5.3.0 (<http://surfer.nmr.mgh.harvard.edu>) [31] running under Matlab2012 on the CentOS 7.5.1804 operating system. The main steps of the pre-processing are recapitulated in the following contents. Firstly, the structural images were performed non-uniformity artifacts correction. Secondly, coordinate transformation was implemented [11, 32]. Next, the corrected images were segmented into gray matter, white matter, cerebrospinal fluid and other background categories. Then, the reconstruction of grey/white matter

Table 1 Demographic and clinical data for all participants

Means \pm SD, n	MCI-C	MCI-NC	<i>p</i> value
Sample size	43	35	–
Male/female	35/8	30/5	0.611*
Age (years)	73.6 ± 7.6	75.3 ± 6.7	0.314 [#]
MMSE score	27.0 ± 1.4	28.1 ± 1.4	0.01
CDR score	0.5	0.5	–

SD standard deviation, MCI-C Mild cognitive impairment converter, MCI-NC Mild cognitive impairment non-converter. *Chi-square test. [#]Two-sample T test. MMSE Mini-mental state examination, CDR clinical dementia rating

boundaries was followed [32]. After completing the cortical models, surface expansion, registration to the average template with group subjects' information were executed [31, 33]. Finally, cortical features were extracted, including: cortical thickness (CT), surface area (SA), volume (VOL), sulcal depth (SD), and gyrus height (GH) [34]. A full width at half maximum (FWHM) of 30 mm Gaussian kernel was used for smoothing the images [35]. The segmentation was performed automatically whose errors were visually checked in FreeView and manually corrected [32]. The technical details of pre-processing by FreeSurfer have been published online (<https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferMethodsCitation>).

The CT reflects the closest distance between the white matter and the grey matter surface in each vertex [36]. The SA is the average of the surrounding triangles at each vertex. The volume is a product of CT and SA [37]. The distance below (above) the average surface at each vertex reflects the depth (height) of sulci (gyrus) [38].

Methods

The overall structure of the proposed prediction framework is depicted in Fig. 1. The method mainly consists of following steps: the pre-processing of images, the dynamic morphological feature calculation, multi-morphological similarity network construction, and network attributes classification.

The dynamic morphological feature

The developments of AD and MCI are often accompanied by the atrophy of brain regions and the decreased complexity of whole cortex with a shallower depth of the sulcus [39]. Surface area is also one of the important indicators to quantify changes in the brain structure [40]. Studies have shown that the use of multiple cortical measures can improve AD classification accuracy [32, 41]. Therefore, in addition to the two commonly used indicators, CT and VOL, in MCI conversion prediction, SD, GH and SA were also added. We partitioned the cerebral cortex data pre-processed by FreeSurfer according to the multi-modal parcellation (MMP) atlas [42], excluding the subcortical structures.

$$D_j = \sum_{n=1}^{N_k} \left(\frac{Baseline_{kn} - M_i^{kn}}{Baseline_{kn}} \right) / N_k \quad (1)$$

The static feature was the mean of the morphological measure in each brain region at a single time point, which was commonly used in cognitive science research based on the cerebral cortex [15]. To make full use of the progress information contained in the longitudinal time points, we

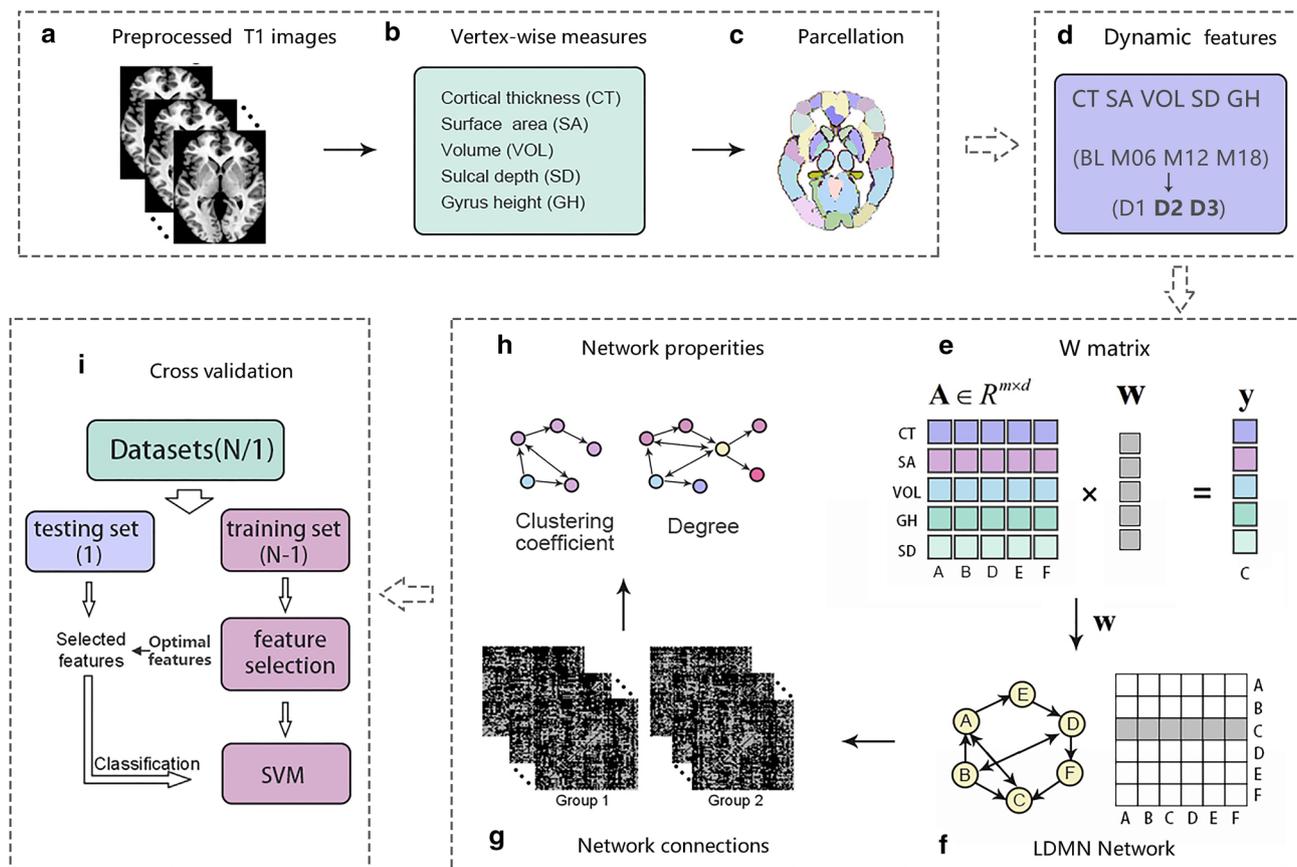


Fig. 1 Illustration of the proposed method. **a, b, c** Represents the procedure of pre-processing. **d** Represents the calculation and extraction of the dynamic morphological features of each measure. **e, f, g, h** Represent the construction of multi-morphological similarity network and the calculation of network attributes. **i** Represents the leave-

one-out cross validation. BL, baseline. M06, 6th month. M12 12th month. M18, 18th month. D1, dynamic features in the first phase. D2, dynamic features in the second phase. D3, dynamic features in the third phase

proposed the dynamic morphological feature relative to the static feature. For the definition of the dynamic feature, as shown in (1), where $Baleline_{kn}$ was the morphological measure value at the n_{th} vertex in the k_{th} brain region at baseline time point and M_i^{kn} denoted the value at the n_{th} vertex in the k_{th} brain region of the measure in the i_{th} month. The number of vertex in region k was defined as N_k . D_j stood for the dynamic morphological feature in phase j on the brain region level. Dynamic morphological features of five measures (CT, SA, VOL, SD and GH) were calculated according to Eq. (1). Because we selected four follow-up time points of subjects, four static features and three dynamic features were obtained on each cortical measure. To select a more superior phase (D1, D2 or D3) for the network construction and compare the ability of the static and dynamic features, the effectiveness of the static and dynamic morphological features to identify MCIs were evaluated by classification performances.

Network construction

In previous studies, a multifeature-based network (MFN) with LASSO algorithm was proposed and achieved praiseworthy results in AD and NC classification [32]. Although LASSO regression has been widely applied in many studies, it also has limitations. LASSO is not robust enough as it selects only one random feature and neglects the others, especially when the variables are high correlation [43]. For instance, if there is a group of brain areas with high pairwise correlation, LASSO algorithm tends to select only one region from the group, regardless of which one is selected. Obviously, LASSO can-not solve the grouping effect problem, and may miss some relevant brain regions during the process of selecting related areas within the designated regions [28]. The elastic network is an extension of LASSO regression that is robust to extreme correlations among the predictors, making the result more credible [43]. Similar

to LASSO, the elastic network can also solve the problem of sparse representation. Therefore, we adopted LASSO and elastic-network regressions for individual network construction.

As shown in (2), the single penalty term $l1$ was used in LASSO regression. The elastic network had the mixed penalty term of $l1$ norm (LASSO) and $l2$ norm (ridge regression), which could be expressed as (3) regularized objective function optimization problem:

$$\min \|x_m - A_m \alpha_m\| + \lambda \|\alpha_m\|_1 \quad (2)$$

$$\min \|x_m - A_m \alpha_m\| + \lambda_1 \|\alpha_m\|_1 + \lambda_2 \|\alpha_m\|_2^2 \quad (3)$$

We represented single participant as $X = [x_1, x_2, \dots, x_m]^T \in R^{m \times d}$ with m brain regions. Since the five kinds of morphological measures and the MMP atlas were used [42], m was 358 (excluding subcortical tissues) and each x_m stood for the m_{th} region containing five morphological measures ($d=5$). Every x_m was a target vector and could be expressed by a linear combination of other $m-1$ prediction vectors [32]. Before the construction of the cerebral similarity network, we normalized the five kinds of dynamic morphological features. The commonly used Min–Max normalization method was employed whose principle is as (4) shown [44]. X_{\max} represented the maximum value and X_{\min} meant the minimum value of each row in the original matrix. X was defined as each element of the row in the matrix that needed to be normalized. Y_{\max} , Y_{\min} denoted the maximum and minimum values of the range that needed to be standardized, respectively.

$$Y = [(Y_{\max} - Y_{\min}) \times (X - X_{\min})] / [X_{\max} - X_{\min}] + Y_{\min} \quad (4)$$

$$y = Aw \quad (5)$$

The regression model was set to $y = Aw$, where y represented the target vector, A ($A = [x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m]$) was a matrix contained all regional vectors except x_i , and w denoted the regression coefficient. We obtained the sparse solution by solving the corresponding optimization problem shown in (6).

$$\min_w \|y - Aw\| + \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2 \quad (6)$$

To optimize the network ability and promote the accuracy of the classification, the two-level tuning method was proposed for the majorization of regression parameters. We first varied a range for the parameters, and based on the first classification results, delineated an elaborate range for fine-grained optimization. Referred to the previous studies for model parameters setting in LASSO and elastic network [5, 32], we constructed the longitudinal dynamic

multi-morphological network with parameters in this range ($\lambda_1 \in \{\alpha \times [10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]\}$, $\alpha \in \{2, 3, 4, 6, 8, 9\}$, $\lambda_2 \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$) of elastic network first. For contrasting the performance of LASSO and elastic network, the λ of LASSO took the same range with λ_1 of elastic network ($\lambda \in \{\alpha \times [10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]\}$, $\alpha \in \{2, 3, 4, 6, 8, 9\}$). Then, we made a fine-grained division in the smaller range according to the distribution of the parameter optimization results in the first step, and this will be discussed in Sect. "Results" in more detailed. The SLEP package was used to solve the optimization problem [45]. We set the parameter of the SLEP package *opts.rFlag* = 1 to make the maximal value of λ_1 and λ_2 , above which shall obtain the zero solution. Figure 1e and f show the network construction process, since each row of the sparse solution came from different regression processes, and the solution finally obtained was a 358×358 asymmetric matrix.

Previous research showed that network attributes present obvious advantages in classification [16, 17]. We calculated the common used network attributes, including clustering coefficient (CC), network degree (Deg), global efficiency (GE) and edge density (ED), of LDMN with the Brain Connectivity Toolbox (BCT) for classification [46]. The CC indicates the degree of aggregation of nodes in a graph. The Deg is the total number of edges connected to a node [47]. The GE measures the efficiency of distant information transmission in the network [48]. The ED is the fraction of present edges to possible edges [46, 49].

Classification and evaluation

SVM is a supervised multivariate classification method that identifies the optimal hyper plane of the maximum margin [50]. SVM usually need to map samples to a high-dimensional feature space, so that the samples are linearly separable, and the kernel function determines the mapping form [51]. Common ones include the linear kernel, the radial basis function (RBF) kernel, the polynomial kernel, etc. [52]. Among them, the linear kernel is more robust to high-dimensional features, and the RBF kernel is more suitable for low-dimensional features [51, 52]. In our study, according to the dimension differences in different types of features and data characteristics [52, 53], support vector machines based linear kernel and RBF kernel were used to classify the MCI-C and MCI-NC on the basis of their multiple dynamic morphological features, where we employed F-score for feature selection. Considering the small number of participants, leave-one-out cross validation (LOOCV) should be adopted to obtain a credible estimate for a classification algorithm [54–56]. In each LOOCV trial for n samples, $n-1$ samples were selected as training set elements and the leave-one sample was used for the testing step. In addition, a tenfold cross validation was repeated ten times to evaluate the robustness

of LOOCV, where the subjects were divided into 10 folds randomly and each fold was used for testing and the rest for training [57, 58]. A permutation test was conducted to estimate the statistical significance of the observed classification accuracy. Concretely, both the entire LOOCV and tenfold cross-validation procedures were repeated 5000 times using randomly shuffled labels [47, 59, 60].

F-score selected significant features by measuring the recognition ability of features in the classification of two-type pattern recognition problems. The training sample $x^k \in R^n, k = 1, 2 \dots$, was divided into two categories: positive and negative, where the number of positive samples was n_+ and the negative sample was n_- . The score of the i_{th} feature could be expressed as:

$$F(i) = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1} \sum_{k=1}^{n_+} (x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1} \sum_{k=1}^{n_-} (x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \tag{7}$$

\bar{x}_i was the average eigenvalue of the i_{th} feature over the entire training set. $\bar{x}_i^{(+)}$ meant the average eigenvalue of the i_{th} feature in the positive class while $\bar{x}_i^{(-)}$ represented the average eigenvalue in the negative class. $x_{k,i}^{(+)}$ was defined as the eigenvalue of the i_{th} feature on the k_{th} positive class sample. On the contrary, the negative class sample was described by $x_{k,i}^{(-)}$.

$$\{w, b\} = \arg \min \|w\|_2^2 + c \sum_{i=1}^N \max(0, 1 - y_i(\langle w, \theta(f_i) \rangle + b)) \tag{8}$$

$$K(f_j, f_i) = f_j^T f_i \tag{9}$$

$$K(f_j, f_i) = \exp\left(-g \|f_j - f_i\|^2\right), g > 0 \tag{10}$$

Given a training set $\{f_i, y_i\}$, where f_i was the eigenvector and y_i was the label of the subject i (MCI-C or MCI-NC). f_i was mapped to high-dimensional feature spaces via kernel functions as $\theta(f_i)$. The optimal SVM model $\{w, b\}$ was calculated by minimizing the cost function [8]. $\{w, b\}$ were the learned parameters of the model. Equations (9) and (10) are the digital representations of linear and RBF kernel functions, respectively. Here we implemented the classification via the LIBSVM toolkit [61].

Results

Classification performances

On each cortical measure, we calculated the static features at four time points and the dynamic features of its three stages, comparing the classification performances of the dynamic

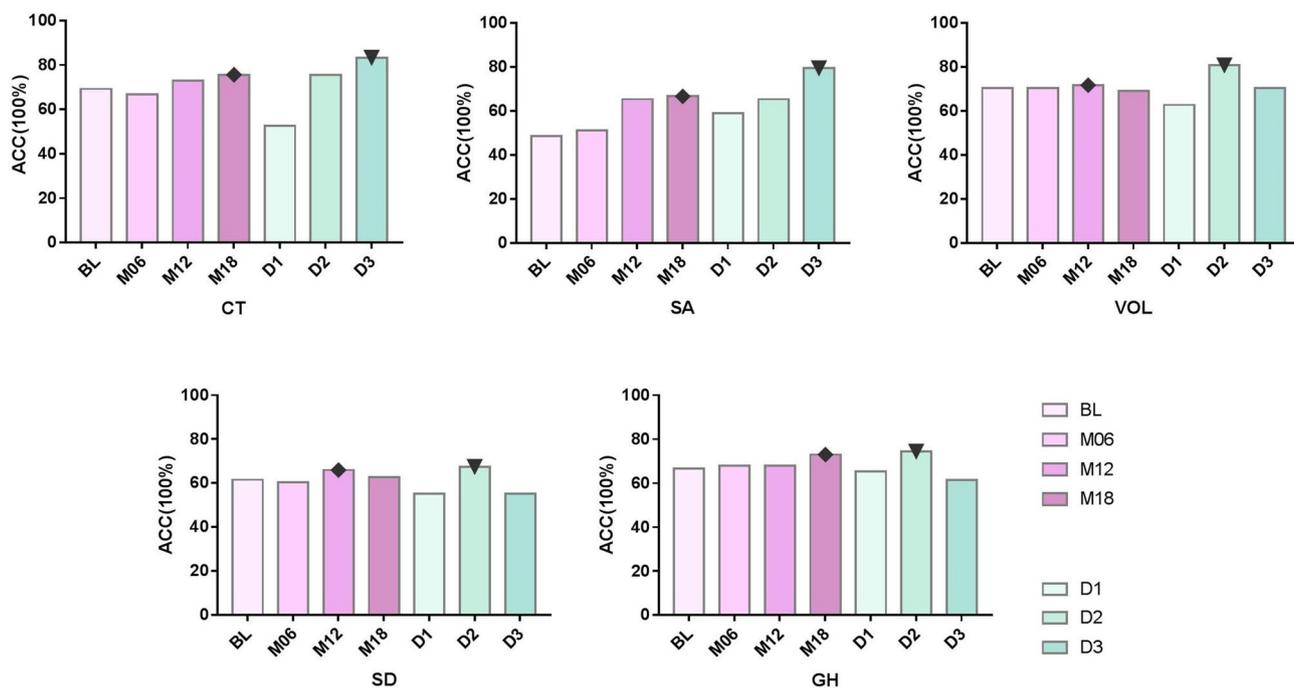


Fig. 2 The classification results of the five cortical measures in the comparison. The static and dynamic features are represented by different colors. The triangle represents the best result in dynamic fea-

tures and the rhombus represents the best result in static features. *CT* Cortical thickness. *SA* Surface area. *VOL* Volume. *SD* Sulcal depth. *GH* Gyrus height

and static features based on the measure values directly. The results demonstrated that the dynamic morphological feature was superior to the static feature in each cortical measure on the optimal results, as shown in Fig. 2. More detailed, SD, GH and VOL obtained the best classification accuracy in D2 phase, while CT and SA performed best in D3 phase. The global (among CT, SA, VOL, SD and GH) best accuracy 83.33% was achieved in CT as the most common cortical indicator in MCI and AD classification research. Therefore, the dynamic morphological features of phases D2 and D3 were selected for the next cerebral similarity network construction. In our classification experiments based on LDMN, we employed accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC) obtained from

the receiver operating characteristic (ROC) to evaluate the performance of the classification.

In each of the selected phases (D2 and D3), we constructed the LDMN with LASSO (L-LDMN) and the LDMN based on elastic network (E-LDMN), optimizing the parameters of sparse regression for all networks. Next, the network properties (CC, Deg, GE, ED) were computed as features to train classifiers and then the concatenation of them was also used for classification.

The best classification accuracy 92.31% was obtained in the E-LDMN of D2 stage in results of LOOCV. Except the global efficient of L-LDMN in D3 and the edge density in each stage, all classification results passed the permutation test ($p < 0.05$). As depicted in Fig. 3 and Table 2, the classification performances in D2 phase were always better

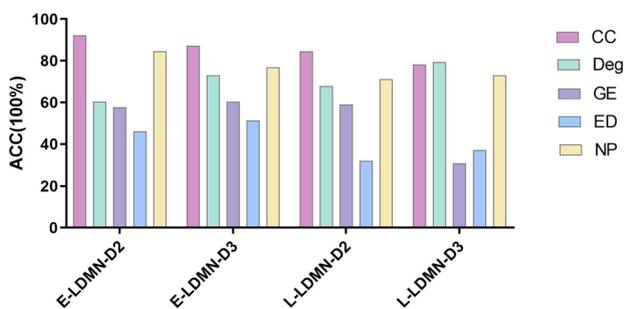


Fig. 3 The accuracy statistics of the four classification results based on network properties, including the clustering coefficient (CC), the degree of network (Deg), the global efficiency (GE), the edge density (ED) and the combination of the network properties (NP)

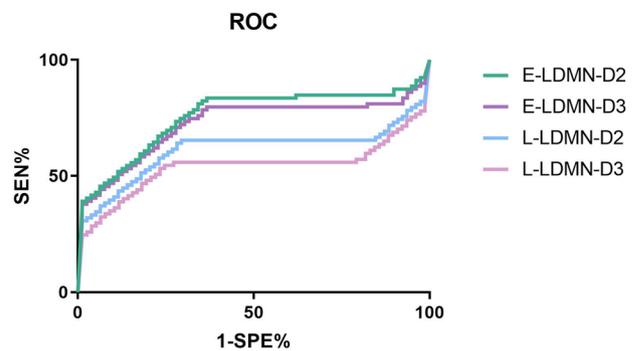


Fig. 4 ROC curves of four classification results based on the clustering coefficients

Table 2 Classification performances of LDMN based on LASSO and elastic network in stage D2 and D3

Kernel	ACC	L- LDMN-D2			$\lambda 1$	p	ACC	E- LDMN -D2			$\lambda 1$	$\lambda 2$	p	
		SEN	SPE	AUC				SEN	SPE	AUC				
CC	Linear	84.65%	100%	65.71%	0.6571	0.009	0.0002	92.31%	100%	82.86%	0.8465	0.009	0.001	0.0002
Deg	Linear	60.26%	67.44%	51.43%	0.5355	0.009	0.0218	67.95%	93.02%	37.14%	0.6841	0.009	0.001	0.0006
GE	RBF	58.97%	72.09%	42.86%	0.5894	0.009	0.0456	57.69%	90.70%	17.14%	0.4807	0.009	0.001	0.0426
ED	RBF	32.05%	51.16%	08.57%	0.2993	0.009	0.9990	46.15%	83.72%	00.00%	0.4478	0.009	0.001	0.7401
NP	Linear	71.79%	74.42%	68.57%	0.7229	0.009	0.0004	84.62%	90.70%	77.14%	0.9037	0.009	0.001	0.0002
Kernel	ACC	L- LDMN-D3			$\lambda 1$	p	ACC	E- LDMN -D3			$\lambda 1$	$\lambda 2$	p	
Kernel	ACC	SEN	SPE	AUC				SEN	SPE	AUC				
CC	Linear	78.25%	100%	51.43%	0.5694	0.0008	0.0002	87.18%	100%	71.43%	0.8027	0.05	1	0.0002
Deg	Linear	79.48%	88.37%	68.57%	0.8306	0.0008	0.0002	73.08%	86.05%	57.14%	0.7548	0.05	1	0.0002
GE	RBF	30.77%	51.16%	05.71%	0.1528	0.0008	0.9998	60.26%	97.67%	14.29%	0.2492	0.05	1	0.0122
ED	RBF	37.18%	67.44%	00.00%	0.0000	0.0008	0.9808	51.28%	93.02%	00.00%	0.1548	0.05	1	0.3831
NP	Linear	73.08%	83.72%	60.00%	0.7542	0.0008	0.0002	76.92%	83.72%	68.57%	0.7508	0.05	1	0.0002

ACC Accuracy, SEN Sensitivity, SPE Specificity, AUC Area under the curve, CC Clustering coefficient, Deg Degree, GE Global efficiency, ED Edge density, NP Combination of Network properties, RBF Radial basis function

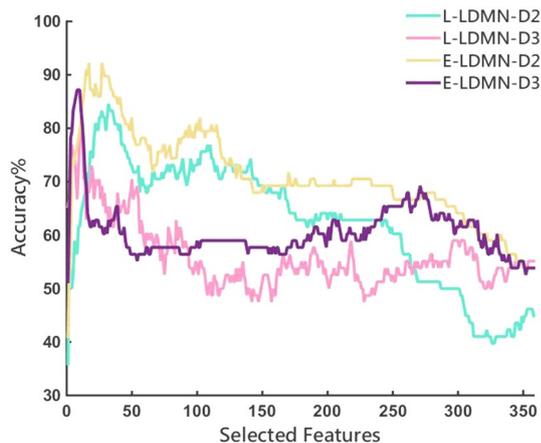


Fig. 5 The classification accuracies with varying number of selected features

than D3 phase both in L-LDMN and E-LDMN which is also displayed in the ROC curves (Fig. 4). In addition, the method using elastic network obtained a better performance than LASSO. Figure 5 shows that only less than 10% of the features were selected when reaching the best accuracies in results of L-LDMN and E-LDMN. The tenfold cross validation was performed on the optimal feature of each LDMN in LOOCV results to estimate the robustness of results. As shown in Table 3, although the accuracy is slightly lower than that of LOOCV, the best result still expressed in D2 of E-LDMN, and all results passed the permutation test ($p < 0.05$).

Effect of regression parameters

Existing research showed that parameter λ of regression algorithm has great impact on the network topology [62]. The size of λ is positively related to the network sparsity and negatively related to the noise content, affecting the credibility of the network and the performance of the classification ultimately.

The method of two-level tuning was used for parameters optimization. Firstly, the previous sparse regression studies were referred to make parameters setting. Zheng et al. varied the parameter λ range ($\lambda \in \{10^{-4}, a \times 10^{-3}, b \times 10^{-2}, c \times 10^{-1}\}$, where $a, b \in \{1, 2, \dots, 9\}, c \in \{1, 2, \dots, 5\}$) for LASSO in MFN,

while the best result was attained when $\lambda = 0.06$ and $\lambda = 0.003$ [32]. In our study, the parameter λ_1 was varied in the range ($\lambda_1 \in \{\alpha \times [10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]\}$, $\alpha \in \{2, 3, 4, 6, 8, 9\}$), where α was set to multiples of 2 and 3 ($\alpha < 10$). The range of λ_2 was set according to the parameter standard of Tong et al. [5] ($\lambda_2 \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$). Since the first parameter has a greater influence on the result in elastic network, the secondary tuning was mainly used for the first parameter λ_1 . The best classification results were obtained in the range from 10^{-4} to 10^{-2} of λ_1 according to the first classification results. Classification accuracy can be improved with the increased number of λ values [63]. So the range (from 10^{-4} to 10^{-2}) of λ_1 was varied further. The range of α would include all positive integers less than 10. Finally, E-LDMN-D3 achieved better classification performance at $\lambda_1 = 0.05$ within the more specific range, but the second tuning did not bring obvious improvement for E-LDMN-D2. The optimizations of parameter λ_1 in all networks are shown in Fig. 6 and the first parameter tuning of E-LDMN-D2 is depicted in Fig. 7.

Brain regions and brain structure analysis

We performed statistical analysis on the differential brain regions of morphological measures in the D2 phase, after the comparison between the static and dynamic features. Since CT and SA got the best performances in the D3 phase, we also took the D3 phases of these two measures into account. In total 57 main brain regions with overlapped measures were implemented further analysis. As shown in Fig. 8,

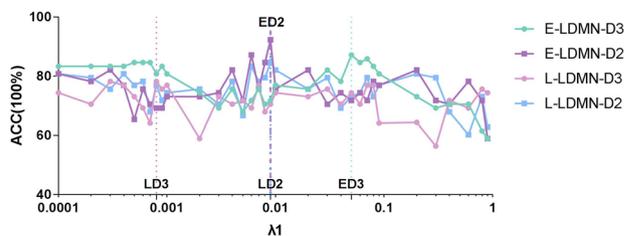


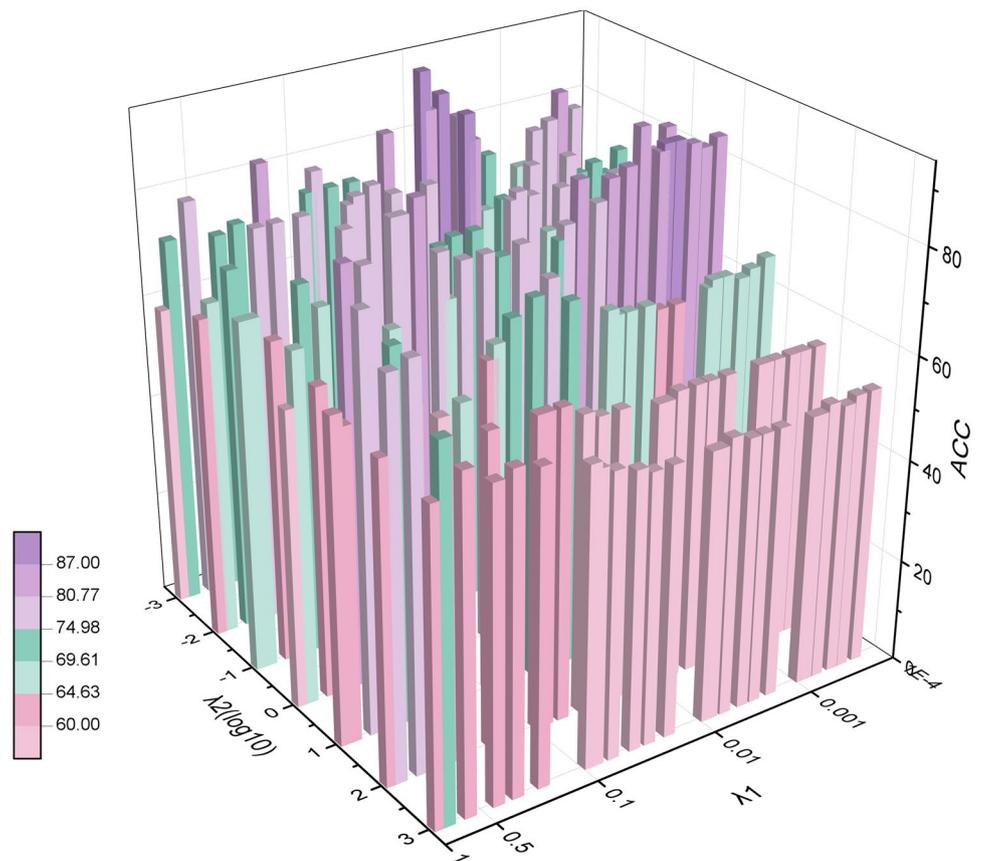
Fig. 6 Regularization parameter λ_1 (λ) with α including all positive integers less than 10. The value of λ_2 are set to 0.001 and 1 for E-LDMN-D2 and E-LDMN-D3, respectively. The dotted lines are located at the extreme points of the network classification performances

Table 3 Classification results of tenfold cross validation

	ACC	SEN	SPE	AUC	λ_1	λ_2	p
L-LDMN-D2	78.26%	87.25%	61.67%	83.12%	0.0090	–	0.0002
L-LDMN-D3	75.44%	89.20%	58.33%	71.30%	0.0008	–	0.0002
E-LDMN-D2	84.62%	93.00%	73.33%	92.46%	0.0090	0.001	0.0002
E-LDMN-D3	81.38%	90.25%	69.58%	80.47%	0.0500	1	0.0002

ACC Accuracy, SEN Sensitivity, SPE Specificity, AUC Area under the curve

Fig. 7 The first elastic-network tuning in E-LDMN-D2. ($\lambda_1 \in \{a \times [10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]\}$, $a \in \{2, 3, 4, 6, 8, 9\}$, $\lambda_2 \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$)



different colors are used to represent brain regions that have different number of morphological measure overlaps.

The main morphological differential brain regions were located in cingulate gyrus [64, 65], insular lobe [65], parahippocampal gyrus [65], olfactory cortex [66], partial front lobe (inferior frontal sulcus, frontal pole, operculum frontale) [67], partial parietal lobe (parietal operculum, superior parietal cortex, the parietal lobe in groove) [67], partial occipital lobe (transverse occipital sulcus, anterior occipital lobe) [64], and transverse temporal gyri [65, 68]. The detailed information for each main morphological differential brain region is provided in the Supplement Table 1.

We used the LDMN under the best classification performance for network structure analysis to characterize clinical significance. Since the brain structure of MCI-NC has similarity to MCI-C, and most of the data did not satisfy the normal distribution. The chi-square test was carried out on the network connections between group MCI-C and MCI-NC with $p < 0.0001$ (uncorrected). The cerebral regions connected by these difference links mainly included anterior hippocampal [65], cingulate gyrus [64, 65], insular lobe [65], frontal pole and operculum [67], partial occipital lobe (parietal-occipital, occipital belly) [64], superior temporal gyrus and sulcus [65] and tectum [67]. As the abnormal links were mostly long connections across the hemisphere,

which may be related to information transmission, we also analyzed the global efficiency of the two groups by t test and the result showed that significant changes occurred ($p = 0.012 < 0.05$).

The difference of the connection averages between group MCI-C and MCI-NC was computed by $(avg_c - avg_{nc})$ to further reveal the structural changes. The negative difference indicated that the relationship of two brain regions connected by this link was decreased in MCI-C. Conversely, it indicated the relationship was enhanced when the difference was positive. Figure 9 illustrates that decreased links were shown in most of differential connections. The specific p values are presented in the Supplement Table 2. The same analysis was also performed on the global efficiency averages and the declined global efficiency was found in MCI-C compared with MCI-NC.

Comparison with other methods

We compared the present results with the state-of-the-art results, which also used MRI data from the ADNI database to predict MCI conversion.

As Table 3 demonstrated, the proposed method achieved competitive performance both in the metrics of ACC and SEN. Among all results in Table 4, the best classification

Fig. 8 The main morphological differential brain regions selected by dynamic features of multiple measures. The different colors represent the number of overlaps with different measures. The green area is the region selected by two cortical measures (Two-overlap). Similarly, purple stands for Three-overlap, and pink represents Four-overlap

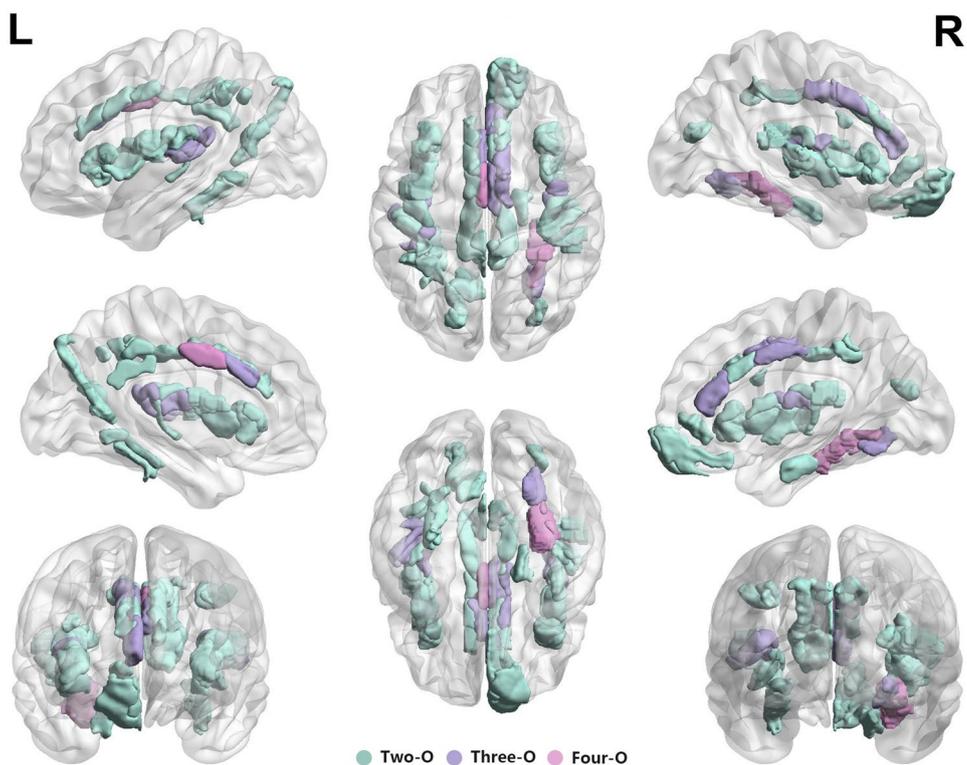
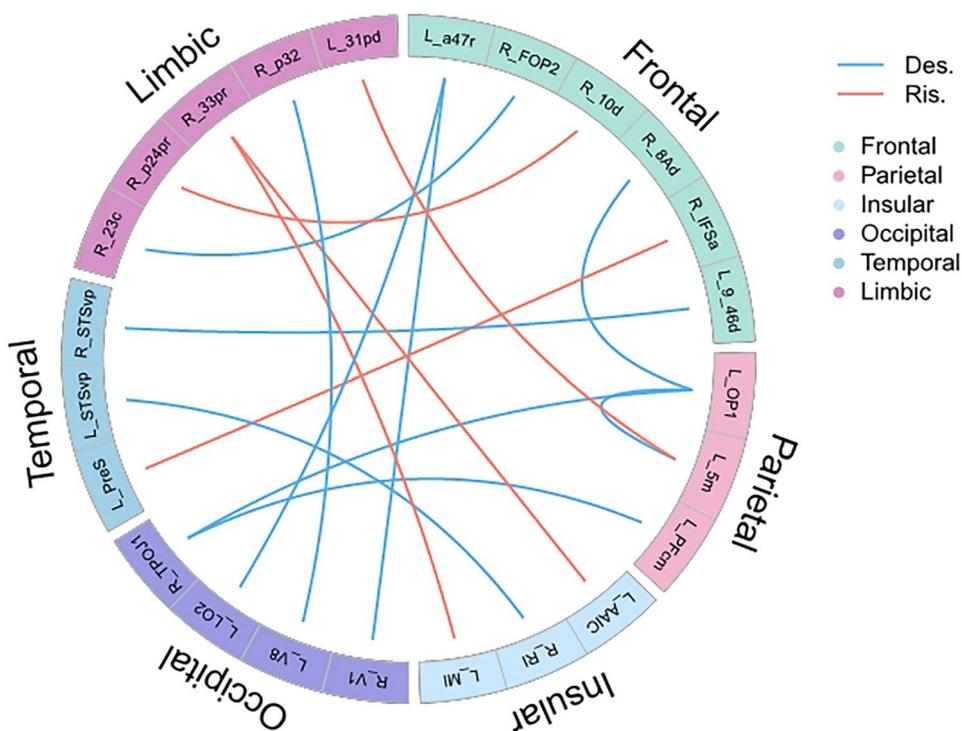


Fig. 9 The significantly different brain network connections in MCI-C compared with MCI-NC ($p < 0.0001$). The orange and blue lines indicate the significantly increased and decreased interregional relationships between the corresponding regions, respectively. The six brain lobes are represented by different colors



performance was obtained by Seyed et al. [11]. The high accuracy was mainly due to the combination of multimodal data in which the functional-MRI (f-MRI) complemented the finer dynamic information lacking in s-MRI [11].

Single modal (s-MRI) was used in the proposed method which achieved better accuracy than the single modal result (s-MRI) in study of Seyed et al. (92.31% > 89.40%), and combining multiple modes may further improve our

Table 4 Comparison with recent state-of-the-art results based on ADNI database

Author	Data	Subjects	Manner	ACC	SEN	SPE	AUC	Year	Method
Moradi [79]	MRI/CT	164C/100NC	Cross-sectional	82.00%	87.00%	84.00%	0.9020	2015	TSVM
Wei [15]	MRI	76C/83NC	Longitudinal	76.39%	65.57%	84.34%	0.8130	2016	SVM
Liu [69]	MRI	117C/117NC	Cross-sectional	79.25%	87.02%	75.54%	0.8344	2016	SVM
Minhas [10]	MRI/NM	16C/13NC	Longitudinal	89.66%	87.50%	92.31%	–	2017	NPC
Tong [5]	MRI/CT	171C/129NC	Cross-sectional	84.10%	88.70%	76.50%	0.9170	2017	SVM
Sun [8]	MRI	67C/43NC	Longitudinal	92.00%	95.00%	90.00%	0.9400	2017	SVM
Hojjati [11]	MRI/fMRI	18C/62NC	Cross-sectional	97.00%	95.00%	100%	0.9800	2018	SVM
Minhas [80]	MRI/NM	54C/65NC	Longitudinal	84.29%	70.36%	92.31%	0.8893	2018	SVM
Proposed	MRI	43C/35NC	Longitudinal	92.31%	100%	82.86%	0.8465	2019	SVM

CT Cognitive test, NM Neuropsychological measures, ACC Accuracy, SEN Sensitivity, SPE Specificity, AUC Area under the curve, TSVM Transductive support vector machine, NPC No parametric classification

classification effect. Besides, it is worth noting that the proposed method obtained more excellent sensitivity than [11]. The high sensitivity is considered to be advantageous for confident MCI diagnosis, which can provide potentially clinical aid in early intervention [69]. Our proposed method employed the dynamic morphological feature that utilized the longitudinal evolution information more sufficiently and provided a good basis for the superior classification results. Similarly, Sidra et al. [10] also obtained nearly 90% classification accuracy based on progression information by synthesizing information from multiple time points. Therefore, we believe that the full use of evolution information is conducive to improving classification accuracy. Sun et al. [8] gained a commendable accuracy using longitudinal information combined with the anatomy of the whole brain. Also from the longitudinal standpoint, our proposed method constructed the cerebral network combining multiple morphological measures from the perspective of the whole brain. These indicate that the prediction of AD benefits from the whole brain perspective.

Discussions

Analysis of classification results

The comparison result of the static and dynamic features proves that more progress information is helpful for the diagnosis of MCIs. From the network classification results, the elastic-network based method achieved better performance than the LASSO based approach which is consistent with previous study [62]. The classification results on basis of CC performed better than other network attributes and the ED has limited diagnostic value of MCIs. The fusion of the network properties did not improve the classification accuracy, which may be associated with the plane-based classification mechanism of SVM [70].

Wei et al. found that more stable and higher classification accuracy can be obtained from the short-term prediction (12 month) compared with the long-term prediction (18 month) [15]. From Yuan's research, we acknowledge that MCI-NC and MCI-C gray matter atrophy occur differently [71]. The atrophy in MCI-NC is later than MCI-C, and the significant changes of MCI-C have been clearly reflected in M12 [71]. In our study, the best prediction performance was obtained in the second stage (from baseline to month 12) with the superior accuracy 92.31%, which shows that the morphological difference between MCI-NC and MCI-C in the second phase is more significant than the third phase. Therefore, we deem that the second phase would be a better clinical diagnosis period. Longitudinally, the significance difference of brain atrophy between MCI-C and MCI-NC presented a non-linear trend, which may be due to some compensation mechanism of the brain. In addition, we found that with disease progression, the alterations of cortical measures differed. The significant difference between MCI-C and MCI-NC of CT and AREA achieved the best recognition effect in the third stage, while SD, GH and VOL were in the second stage, which may be related to the diversity of AD progress.

The results of randomization tests showed that in both LOOCV and tenfold cross validation, the main results are significantly higher than the random performance ($p < 0.05$) with statistical significance [59, 60]. While compared with LOOCV, a slight decrease was happened in tenfold cross-validation results. We have already used a filtering feature selection algorithm F-score to prevent over-fitting, and only a small part of features (far less than the number of subjects) were employed when the highest classification accuracies were achieved in LOOCV results [47]. Notably, the trends that D2 stage is better than D3 phase and the elastic network-based method is superior to the lasso based approach in classifying MCIs are expressed both in LOOCV and tenfold cross-validation results, proving the robustness of the

dynamic features. In addition, the main results of LOOCV and tenfold cross validation all passed the permutation test. Therefore, we deem that the small decrease in the accuracies of tenfold cross validation compared to LOOCV is mainly due to insufficient model training as the result of limited samples, rather than over-fitting of LOOCV. The results of LOOCV are reliable.

Analysis of characteristic brain regions and LDMN structure

In the main 57 morphological differential brain regions, the left and right brains were basically symmetrical [24]. While the number of ROIs in the right brain was slightly more than the left, which consists with the right brain being more sensitive to human memory [72]. The fact that the morphological characteristic brain regions mostly overlapped with the ROIs in previous AD and MCI research proves the dynamic morphological feature has certain robustness.

The difference of LDMN structure between MCI-NC and MCI-C indicates that the changes in the cortical structure of the brain region significantly affect the relationship among cerebral areas. Consistent with previous studies, the structural links showing significant difference between MCI-C and MCI-NC were mainly long-range connections across the cerebral hemisphere [16]. The abnormal long-range connections might mean a change in the information transfer path of the patient group across global brain areas. Further analysis of network attributes suggested that the global efficiency of MCI-C declined than that of MCI-NC. The MCI-C showed the decreased connections between mostly cerebral regions compared to MCI-NC [32], which might indicate descending efficiency of collaborative work between these brain regions. The decreased links also demonstrate that the similarity of brain regions would subside during the conversion from MCI to AD, as the connections of LDMN reflect a similar relationship between cerebral regions.

The affected brain regions found in our study in terms of dynamic features, were mainly located in limbic, insula, frontal and occipital lobes. Structures of parietal and temporal lobes were also affected. According to the histopathological staging, MCI belongs to the limbic stages of AD [73, 74]. The key characteristics of the stages are the severe involvement of the transentorhinal and entorhinal regions and higher order association areas of the neocortex, as well as interrupted connections between components of the limbic loop at multiple sites [73]. The limbic loop plays an important role in maintaining memory function and emotional balance, and these special sites are most prone to AD-related pathology [74, 75]. The abnormal limbic loop components between MCI-C and MCI-NC were detected in our study, which may hamper the exchange of data between the higher order components of the limbic system and the

prefrontal, occipital cortex [74]. The change of temporal lobe is related to the accumulation of neurofibrillary tangles and neuropil threads in the hippocampus [76, 77]. Parietal lobe has sensory centers and many other important areas, whose abnormality may result face agnosia and is an important clinical manifestation of AD [74, 78]. These affected regions all have been extensively reported in previous AD and MCI studies, which demonstrates that these characteristic areas are more significant in the development of AD and more attention should be paid to them clinically.

Possible reasons for the excellent performance of LDMN

The satisfied performances in prediction from MCI to AD indicate the practicability of our method. Compared with other longitudinal research, our study used the new longitudinal biomarker “the dynamic morphological feature” which considers the pathological information evolving over time. Existing longitudinal studies mostly used the longitudinal data as a criterion to select subjects, finally the baseline data was selected for classification and data analysis [8, 10, 11], which still belong to a kind of “cross-sectional” study. The dynamic morphological feature used in our method reflects the course of disease progression and complements the evolution information in time dimension of s-MRI, providing richer information for disease conversion prediction. The feature enhances the difference between MCIs (MCI-C and MCI-NC) in s-MRI data, and is more sensitive to the recognition of two kinds of MCI. By adopting the multiple dynamic morphological features, the classification and prediction abilities of LDMN were further improved.

The LDMN combined multiple cortical measures (CT, SA, VOL, SD and GH) to build brain networks. These measures are thought to be closely related to the development of AD and MCI [39]. The fusion of multi-measure features ensured the LDMN included more detailed and comprehensive information [32, 41]. Meanwhile, the use of MMP atlas provided a boost for the excellent classification results due to finer brain areas parcellation [11]. In addition, the application of sparse multiple regression model can more fully reflect the nature of brain network connections [32, 62]. Different from the commonly used pair-wise correlation network, which only considers the relationship between the two simple brain regions [62], the multivariate regression model in our study takes all cerebral regions into consideration, according to that AD and MCI are accompanied by multiple brain regions atrophy usually [22, 23].

Previous study showed that sparse regression can be employed to construct networks and achieve commendable classification performance [25]. We utilized the

elastic-network method to integrate the multiple dynamic morphological features. Elastic networks, as a robust and more applicable sparse regression method, also obtained better classification ability than LASSO regression in our study. Compared with LASSO, elastic network has one more parameter (l_2 -norm) which can greatly affect the calculation result and work well in solving the grouping effect [43]. Therefore, with l_1 being for automatic variable selection and l_2 encouraging grouped selection [62], the integration of l_1 and l_2 significantly improved the construction of LDMN. A larger range of parameters was set for optimization and a more reasonable two-level tuning method was took for tuning, which also ensured the superior classification results.

Limitations and future directions

Due to the limitation of subjects' number, the conclusions need to be further validated on a larger data set in future work. The hippocampus and amygdaloid nucleus are important clinical features for AD patients and MCI subjects, while our study was based only on the cerebral cortex. Subsequent work will attempt to integrate the cortical and subcortical features for prediction. In addition, the neuropsychological measure is also a commonly used biomarker in neurological diseases studies. The combination of different biomarkers is also our follow-up focus.

Conclusion

The dynamic morphological feature is proved to be a more sensitive biomarker for predicting the conversion from MCI to AD in our study. Based on the biomarker, the cerebral similarity network was constructed with multiple morphological features using sparse regression algorithm. The dynamic multi-morphological network integrates the multiple cortical measures, incorporating the longitudinal evolution information of the MCI conversion process. The proposed method achieves an excellent performance in MCI conversion prediction and provides possible assistance for clinical diagnosis. In addition, the analysis of LDMN connections offers an interesting perspective for revealing the complex lesions of AD.

Acknowledgements This work was supported by the National Basic Research Program of China (973 Program) (No.2014CB744600), the National Natural Science Foundation of China (Grant No.61632014, No.61210010), the Program of Beijing Municipal Science & Technology Commission (No.Z171100000117005), the Program of International S & T Cooperation of MOST (No.2013DFA11140), Fundamental Research Funds for the Central Universities (lzujbky-2018-it64), and the Postdoctoral Funding of Zhejiang Province, China (514000-X81901).

Compliance with ethical standards

Conflicts of interest None.

Ethical standards Data involved in the study came from the publicly open Alzheimer's disease neuroimaging initiative (ADNI) database. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the Helsinki Declaration of 1975.

References

- Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *J Intern Med* 256:183–194
- Hänninen T, Hallikainen M, Tuomainen S, Vanhanen M, Soininen H (2002) Prevalence of mild cognitive impairment: a population-based study in elderly subjects. *Acta Neurol Scand* 106:148–154
- Anter AM, Wei Y, Su J, Yuan Y, Lei B, Duan G, Mai W, Nong X, Yu B, Li C (2019) A robust swarm intelligence-based feature selection model for neuro-fuzzy recognition of mild cognitive impairment from resting-state fMRI. *Informat Sci*
- Li Z, Suk H-I, Shen D, Li L (2016) Sparse multi-response tensor regression for alzheimer's disease study with multivariate clinical assessments. *IEEE Trans Med Imaging* 35:1927–1936
- Tong T, Gao Q, Guerrero R, Ledig C, Chen L, Rueckert D, AsDN I (2017) A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. *IEEE Trans Biomed Eng* 64:155–165
- Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR Jr, Ashburner J, Frackowiak RS (2008) Automatic classification of MR scans in Alzheimer's disease. *Brain* 131:681–689
- Jie B, Zhang D, Cheng B, Shen D, AsDN I (2015) Manifold regularized multitask feature learning for multimodality disease classification. *Hum Brain Mapp* 36:489–507
- Sun Z, van de Giessen M, Lelieveldt BP, Staring M (2017) Detection of conversion from mild cognitive impairment to alzheimer's disease using longitudinal brain MRI. *Front Neuroinform* 11:16
- Zhang J, Liu M, An L, Gao Y, Shen D (2017) Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images. *IEEE J Biomed Health Informat* 21:1607–1616
- Minhas S, Khanum A, Riaz F, Alvi A, Khan SA (2017) A non-parametric approach for mild cognitive impairment to ad conversion prediction: Results on longitudinal data. *IEEE J Biomed Health Informat* 21:1403–1410
- Hojjati SH, Ebrahimzadeh A, Khazae A, Babajani-Feremi A, AsDN I (2018) Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI. *Comput Biol Med* 102:30–39
- Griffa A, Baumann PS, Thiran J-P, Hagmann P (2013) Structural connectomics in brain diseases. *NeuroImage* 80:515–526
- Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10:186
- Zalesky A, Fornito A, Bullmore ET (2010) Network-based statistical: identifying differences in brain networks. *NeuroImage* 53:1197–1207
- Wei R, Li C, Fogelson N, Li L (2016) Prediction of conversion from mild cognitive impairment to Alzheimer's Disease using MRI and structural network features. *Front Aging Neurosci* 8:76

16. Yao Z, Zhang Y, Lin L, Zhou Y, Xu C, Jiang T, AsDN I (2010) Abnormal cortical networks in mild cognitive impairment and Alzheimer's disease. *PLoS Comp Biol* 6:e1001006
17. Zhou Y, Lui YW (2013) Small-world properties in mild cognitive impairment and early Alzheimer's disease: a cortical thickness MRI study. *ISRN geriatrics*
18. Chupin M, Gérardin E, Cuingnet R, Boutet C, Lemieux L, Lehericy S, Benali H, Garnero L, Colliot O (2009) Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus* 19:579–587
19. Eskildsen SF, Coupé P, García-Lorenzo D, Fonov V, Pruessner JC, Collins DL, AsDN I (2013) Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *NeuroImage* 65:511–521
20. Ota K, Oishi N, Ito K, Fukuyama H, Group S-JS (2014) A comparison of three brain atlases for MCI prediction. *J Neurosci Methods* 221:139–150
21. Ferreira LK, Diniz BS, Forlenza OV, Busatto GF, Zanetti MV (2011) Neurostructural predictors of Alzheimer's disease: a meta-analysis of VBM studies. *Neurobiol Aging* 32:1733–1741
22. Du A-T, Schuff N, Kramer JH, Rosen HJ, Gorno-Tempini ML, Rankin K, Miller BL, Weiner MW (2007) Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. *Brain* 130:1159–1166
23. Dai Z, Yan C, Wang Z, Wang J, Cao M, Xia M, Song H, Shu N, Han Y, Li K (2012) Disrupted cortical hubs in functional brain networks in early-stage Alzheimer's disease. *Alzheimer's Dementia* 8:P536–P537
24. Fan Y, Batmanghelich N, Clark CM, Davatzikos C, AsDN I (2008) Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *NeuroImage* 39:1731–1743
25. Jie B, Liu M, Liu J, Zhang D, Shen D (2017) Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer's disease. *IEEE Trans Biomed Eng* 64:238–249
26. Zhu X, Suk H-I, Wang L, Lee S-W, Shen D, AsDN I (2017) A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. *Med Image Anal* 38:205–214
27. Liu M, Zhang D, Shen D, AsDN I (2012) Ensemble sparse classification of Alzheimer's disease. *NeuroImage* 60:1106–1116
28. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J Roy Stat Soc Ser B (Stat Method)* 67:301–320
29. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack CR, Jagust W, Liu E (2013) The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's Dementia* 9:e111–e194
30. Wang L, Zang Y, He Y, Liang M, Zhang X, Tian L, Wu T, Jiang T, Li K (2006) Changes in hippocampal connectivity in the early stages of Alzheimer's disease: evidence from resting state fMRI. *NeuroImage* 31:496–504
31. Fischl B (2012) FreeSurfer *NeuroImage* 62:774–781
32. Zheng W, Yao Z, Xie Y, Fan J, Hu B (2018) Identification of Alzheimer's disease and mild cognitive impairment using networks constructed based on multiple morphological brain features. *Biolog Psychiatry* 3:887–897
33. Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *NeuroImage* 9:195–207
34. Destrieux C, Fischl B, Dale A, Hagren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* 53:1–15
35. Hogstrom LJ, Westlye LT, Walhovd KB, Fjell AM (2013) The structure of the cerebral cortex across adult life: age-related patterns of surface area, thickness, and gyrification. *Cereb Cortex* 23:2521–2530
36. Fischl B, Dale AM (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci* 97:11050–11055
37. Rimol LM, Nesvåg R, Hagler DJ Jr, Bergmann Ø, Fennema-Notestine C, Hartberg CB, Haukvik UK, Lange E, Pung CJ, Server A (2012) Cortical volume, surface area, and thickness in schizophrenia and bipolar disorder. *Biol Psychiatry* 71:552–560
38. Schaer M, Cuadra MB, Tamarit L, Lazeyras F, Eliez S, Thiran J-P (2008) A surface-based approach to quantify local cortical gyrification. *IEEE Trans Med Imaging* 27:161–170
39. Matthews SC, Strigo IA, Simmons AN, Yang TT, Paulus MP (2008) Decreased functional coupling of the amygdala and supragenual cingulate is related to increased depression in unmedicated individuals with current major depressive disorder. *J Affect Disord* 111:13–20
40. Hogstrom LJ, Westlye LT, Walhovd KB, Fjell AM (2012) The structure of the cerebral cortex across adult life: age-related patterns of surface area, thickness, and gyrification. *Cereb Cortex* 23:2521–2530
41. de Vos F, Schouten TM, Hafkemeijer A, Dopfer EG, van Swieten JC, de Rooij M, van der Grond J, Rombouts SA (2016) Combining multiple anatomical MRI measures improves Alzheimer's disease classification. *Hum Brain Mapp* 37:1920–1929
42. Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M (2016) A multi-modal parcellation of human cerebral cortex. *Nature* 536:171
43. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33:1
44. Jain A, Nandakumar K, Ross A (2005) Score normalization in multimodal biometric systems. *Pattern Recogn* 38:2270–2285
45. Liu J, Ji S, Ye J (2009) SLEP: Sparse learning with efficient projections. *Arizona State University* 6:7
46. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* 52:1059–1069
47. Zheng W, Eilamstock T, Wu T, Spagna A, Chen C, Hu B, Fan J (2019) Multi-feature based network revealing the structural abnormalities in autism spectrum disorder. *IEEE Trans Affect Computing*
48. Latora V, Marchiori M (2001) Efficient behavior of small-world networks. *Phys Rev Lett* 87:198701
49. Wang J, Wang L, Zang Y, Yang H, Tang H, Gong Q, Chen Z, Zhu C, He Y (2009) Parcellation-dependent small-world brain functional networks: A resting-state fMRI study. *Hum Brain Mapp* 30:1511–1523
50. Vapnik VN (1999) An overview of statistical learning theory. *IEEE Trans Neural Networks* 10:988–999
51. Hsu C-W, Chang C-C, Lin C-J (2003) A practical guide to support vector classification. In: Taipei
52. Novakovic J, Veljovic A (2011) C-support vector classification: Selection of kernel and parameters in medical diagnosis. In: 2011 IEEE 9th International Symposium on Intelligent Systems and Informatics. IEEE, pp 465–470
53. Milton A, Roy SS, Selvi ST (2013) SVM scheme for speech emotion recognition using MFCC feature. *Int J Computer Applic* 69
54. Wong T-T (2015) Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recogn* 48:2839–2846
55. Steyerberg EW, Harrell FE (2016) Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 69:245–247
56. Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage* 45:S199–S209

57. Masso M, Rao N, Pyarasani P (2018) Modeling transcriptional activation changes to Gal4 variants via structure-based computational mutagenesis. *PeerJ* 6:e4844
58. Liu X, Huang H (2020) Alterations of functional connectivities associated with autism spectrum disorder symptom severity: a multi-site study using multivariate pattern analysis. *Sci Rep* 10:1–13
59. Huang P, Cui L-B, Li X, Lu Z-L, Zhu X, Xi Y, Wang H, Li B, Hou F, Miao D (2018) Identifying first-episode drug naïve patients with schizophrenia with or without auditory verbal hallucinations using whole-brain functional connectivity: A pattern analysis study. *NeuroImage: Clinical* 19:351–359
60. Geng X, Xu J, Liu B, Shi Y (2018) Multivariate classification of major depressive disorder using the effective connectivity and functional connectivity. *Front Neurosci* 12:38
61. Chang C-C, Lin C-J (2011) LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol (TIST)* 2:27
62. Guo H, Li Y, Xu Y, Jin Y, Xiang J, Chen J (2018) Resting-State Brain Functional Hyper-Network Construction Based on Elastic Net and Group Lasso Methods. *Front Neuroinform* 12
63. Jie B, Wee C-Y, Shen D, Zhang D (2016) Hyper-connectivity of functional networks for brain disease diagnosis. *Med Image Anal* 32:84–100
64. Chételat G, Desgranges B, Landeau B, Mézenge F, Poline J, de La Sayette V, Viader F, Eustache F, Baron J-C (2007) Direct voxel-based comparison between grey matter hypometabolism and atrophy in Alzheimer's disease. *Brain* 131:60–71
65. Misra C, Fan Y, Davatzikos C (2009) Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. *NeuroImage* 44:1415–1422
66. Vasavada MM, Wang J, Eslinger PJ, Gill DJ, Sun X, Karunanayaka P, Yang QX (2015) Olfactory cortex degeneration in Alzheimer's disease and mild cognitive impairment. *J Alzheimer's Dis* 45:947–958
67. Di Paola M, Macaluso E, Carlesimo G, Tomaiuolo F, Worsley K, Fadda L, Caltagirone C (2007) Episodic memory impairment in patients with Alzheimer's disease is correlated with entorhinal cortex atrophy. *J Neurol* 254:774–781
68. Liu Y, Paajanen T, Zhang Y, Westman E, Wahlund L-O, Simmons A, Tunnard C, Sobow T, Mecocci P, Tsolaki M (2011) Combination analysis of neuropsychological tests and structural MRI measures in differentiating AD, MCI and control groups—the AddNeuroMed study. *Neurobiol Aging* 32:1198–1206
69. Liu M, Zhang D, Shen D (2016) Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment. *IEEE Trans Med Imaging* 35:1463–1474
70. Xizhe Z, Jie Z, Chen W, Hegao C (2004) Study on a SVM-based data fusion method. In: *IEEE Conference on Robotics, Automation and Mechatronics*. IEEE, pp 413–415
71. Yuan G, Zhuo Z, Li H (2013) Longitudinal progression of grey matter atrophy morphological characteristics in MCI patients. In: *IEEE International Conference on Medical Imaging Physics and Engineering*. IEEE, pp 66–71
72. Sperry RW (1961) Cerebral organization and behavior. *Science* 133:1749–1757
73. Spulber G, Niskanen E, MacDonald S, Kivipelto M, Ferreira Padilla D, Julkunen V, Hallikainen M, Vanninen R, Wahlund L-O, Soininen H (2012) Evolution of global and local grey matter atrophy on serial MRI scans during the progression from MCI to AD. *Curr Alzheimer Res* 9:516–524
74. Braak E, Griffing K, Arai K, Bohl J, Bratzke H, Braak H (1999) Neuropathology of Alzheimer's disease: what is new since A. Alzheimer? *Eur Arch Psychiatry Clin Neurosci* 249:S14–S22
75. Li X, Wang H, Tian Y, Zhou S, Li X, Wang K, Yu Y (2016) Impaired white matter connections of the limbic system networks associated with impaired emotional memory in Alzheimer's disease. *Front Aging Neurosci* 8:250
76. Nagy Z, Hindley N, Braak H, Braak E, Yilmazer-Hanke D, Schultz C, Barnetson L, King E-F, Jobst K, Smith A (1999) The progression of Alzheimer's disease from limbic regions to the neocortex: clinical, radiological and pathological relationships. *Dement Geriatr Cogn Disord* 10:115–120
77. Chételat G, Landeau B, Eustache F, Mézenge F, Viader F, de La Sayette V, Desgranges B, Baron J-C (2005) Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study. *NeuroImage* 27:934–946
78. Balestrini S, Francione S, Mai R, Castana L, Casaceli G, Marino D, Provinciali L, Cardinale F, Tassi L (2015) Multimodal responses induced by cortical stimulation of the parietal lobe: a stereo-electroencephalography study. *Brain* 138:2596–2607
79. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, AsDN I (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage* 104:398–412
80. Minhas S, Khanum A, Riaz F, Khan SA, Alvi A (2018) Predicting progression from mild cognitive impairment to Alzheimer's disease using autoregressive modelling of longitudinal and multimodal biomarkers. *IEEE journal of biomedical and health informatics* 22:818–825