

RESEARCH ARTICLE

WILEY

Longitudinal structural MRI analysis and classification in Alzheimer's disease and mild cognitive impairment

Yingteng Zhang¹  | Shenquan Liu² | Xiaoli Yu¹

¹Department of Cardiovascular Surgery, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

²School of Mathematics, South China University of Technology, Guangzhou, China

Correspondence

Yingteng Zhang, Department of Cardiovascular Surgery, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou 510260, China.
Email: xiaoteng28@163.com

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 11572127, 11702064, 11872183

Abstract

Early diagnosis or detection of Alzheimer's disease (AD) and mild cognitive impairment (MCI) is crucial so as to intervene in advance and to better understand the neurodegenerative process. Gray matter volume (GMV) plays an important role in demonstrating unique anatomical characteristics of the brain regions and further differentiates AD, MCI and normal control (NC). In this study, 317 subjects (100 NC, 58 stable MCI (sMCI), 53 converted MCI (cMCI) and 106 AD) are selected from the Alzheimer's Disease Neuroimaging Initiative database. First, the differences of GMV patterns among the between-group comparisons at different time points and the development of longitudinal pattern within the same group are compared. Next, the longitudinal feature combination strategy is applied to construct the classification model by using a support vector machine (SVM) combined with the nested leave-one-out cross-validation (LOOCV) method. The brain structure experiences a gradual change in the process of developing from NC to AD. In addition, the baseline GMV combined with the longitudinal measurements for 2 years of follow-up data yielded optimal classification results. Specifically, the AD-NC comparison achieves the best classification performance with 98.06% accuracy, 97.17% sensitivity, 99.00% specificity, 99.04% positive predictive value (PPV) and 97.06% negative predictive value (NPV). The comparison of the two subtypes of MCI (ie, sMCI and cMCI) also achieves high accuracy. Other between-group comparisons also receive high classification performance. According to statistics, caudate nucleus, hippocampus, temporal pole and lenticular putamen are the most important contribution areas to the between-group comparisons. Our research has the potential to improve the clinical diagnosis of subtypes of MCI and predict the risk of its conversion to AD.

KEYWORDS

Alzheimer's disease, gray matter volume, longitudinal analysis, longitudinal classification, mild cognitive impairment

1 | INTRODUCTION

Mild cognitive impairment (MCI) is usually regarded as a precursor to Alzheimer's disease (AD). Previous studies

indicate that the pathological time axis of AD may occur several years to decades before the clinical diagnosis. It initially happens without any symptoms and then progresses into the MCI stage.^{1,2} Therefore, the early

precaution and detection of AD, as well as its tracking and prediction, are particularly important. It could help prevent the occurrence of AD or slow down the speed of neurodegeneration.

In the past decades, there had been many studies to find biomarkers that could predict the conversion of MCI to AD. Some possible biomarkers include genetic, cerebrospinal fluid proteins, cognitive measurements, glucose metabolism and structural/functional brain abnormalities. In a review of a large number of studies, Landau et al³ used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to compare biomarkers that could predict conversion efficiency. They found 73% to 88% specificity but about 40% less significant sensitivity in classification. Another notable study on predictive transformation was that of Killiany et al,⁴ who used a structural Magnetic Resonance Imaging (MRI) to predict the transformation of MCI and found that the entorhinal cortex, superior temporal sulcus and anterior cingulate gyrus were the most significant classification features with 75% accuracy, but the specificity was only 48%. Davatzikos et al⁵ tried to combine the information of MRI and Cerebro-Spinal Fluid (CSF) biomarkers to predict AD and obtained the highest accuracy of 61.7%. Therefore, more explorative researches are needed to improve the accuracy of predicting the conversion of MCI to AD. In addition, there are more studies based on structural MRI to distinguish patients with MCI or AD from normal controls (NC), to distinguish stable MCI (sMCI) from converted MCI (cMCI) and to track and predict the transition from NC to MCI and from MCI to AD.

Recently, several longitudinal neuroimaging studies have collected much longitudinal data to better understand the progress of neuropsychiatric and neurodegenerative diseases.⁶⁻⁸ Therefore, the longitudinal parameter change in MRI may be a crucial factor in the prediction of future conversion from MCI to AD.^{6,8,9} Until recently, only a few researchers started to use longitudinal data for cMCI/sMCI classification.^{6,8,10} Li et al⁸ investigated the longitudinal cortical thickness changes of 75 MCI subjects to distinguish cMCI from sMCI. Moreover, Zhang et al⁶ proposed an AD prediction method with ROI-based features from longitudinal data. The experiments were performed on 88 MCI subjects, and the results show that the performance of their method with longitudinal data was better than that with baseline data. Despite these efforts, extracting discriminative features from longitudinal data for the early diagnosis and prediction of AD progression is still challenging and requires more research. Specifically, longitudinal images obtained from MCI patients are researched to acquire valuable information on the longitudinal changes that can be used to classify MCI subjects as a cMCI or sMCI subtype. From a clinical

perspective, an observed trend can show the tendency of an MCI subject to become an AD patient or to remain stable. If such trends are dynamically monitored with longitudinal data, AD-related changes can be determined, and an AD prediction model can be constructed with the longitudinal data. In clinical settings, when a new MRI scan of an MCI subject is provided, the future development of the MCI subject (progressing to AD or remaining stable) can be predicted with previous MRI scans and the constructed prediction model. Therefore, more valuable information can be extracted from the longitudinal data to help reinforce prediction accuracy.

Until now, however, most classification studies are based on cross-sectional data. Considering that the process of AD has the natural property of neurodegeneration, we deduce that combining the baseline and longitudinal changes of brain structure over time may improve the accuracy of prediction and transformation and provide a good measurement value for AD. In addition, tracking individual changes over time can naturally reduce some of the underlying confounding factors, such as gender, age, education and diet. Inhomogeneity in populations may be a cause of low accuracy in existing studies.

In this study, we investigate the change of gray matter volume (GMV) among the four groups (NC, sMCI, cMCI and AD) over time and use the machine learning method to explore whether the GMV can be used as an effective feature to distinguish NC, sMCI and cMCI from AD. First, the differences of GMV patterns between groups at different time points and the development of longitudinal pattern within the same group is compared. Then, the longitudinal feature combination strategy is used to construct the classification model by using a support vector machine (SVM) combined with the nested leave-one-out cross-validation (LOOCV) method. Finally, we relate the discriminating features to the pathological features of AD.

2 | MATERIALS AND METHODS

2.1 | Participants

In this study, 317 subjects are collected from the ADNI database, including 100 NCs, 58 sMCI patients, 53 cMCI patients and 106 AD patients. In each group, MRI images are collected at baseline period (bl), 12 months after baseline period (12 m) and 24 months after baseline period (24 m). All of them have mini-mental state examination (MMSE) and clinical dementia rate (CDR) scores. During the tracking period from baseline to 24 months, the MCI that converted to AD is defined as cMCI. If it remained

in the MCI state, it is defined as sMCI. The scanning parameters of whole-brain MRI are as follows: TR, 9 ms; TE, 3.9 ms; field strength, 1.5 T; slice thickness, 1.2 mm; and spatial resolution, 1.25×1.25 mm. Each parameter fluctuates slightly across different scans. We carried out a chi-square test on gender in the four groups and found that there are significant differences in several between-group comparisons, including NC-sMCI comparison ($P = .022$), sMCI-AD comparison ($P = .008$) and cMCI-AD comparison ($P = .042$). There is no significant difference in age between the groups. The scores of MMSE and CDR in cognitive ability assessment are significantly different by analysis of variance (ANOVA) and post-hoc test ($P < .05$). The demographic data and cognitive assessment scores for each group are shown in Table 1.

2.2 | Data processing

A GMV map in the Montreal Neurological Institute (MNI) space is generated for each individual using the SPM12 toolbox (<http://www.fil.ion.ucl.ac.uk/spm/>). This processing procedure includes the following steps: (a) correcting for bias-field inhomogeneity; (b) segmenting the brain into gray matter (GM), white matter and cerebrospinal fluid density maps using the “new-segment” approach; (c) applying diffeomorphic anatomical registrations through exponentiated lie algebra (DARTEL) to generate a custom template¹¹ via default parameter settings; (d) registering each subject’s GM density (GMD) image of the native space to the customized template¹²; (e) registering the resultant image to the MNI space and standardizing the GMD map; (f) applying the modulation by multiplying the resulting GMD map with the nonlinear components of Jacobian determinant, which

result in the GMV maps representing the local native space GM volume after correcting for individual differences in whole brain size; and (g) smoothing GMV maps using a 8-mm full-width at half maximum (FWHM) Gaussian kernel. Finally, we average all subjects’ GMD maps to create a GM mask and apply a threshold of 0.2 to this average map. The GMV features are then restricted to this GM mask. Among the existing atlases, the Automated Anatomical Labeling (AAL-90) atlas is still the most popular atlas in the brain studies,¹³ and it has been widely used in discriminative studies for different disorders.^{14–16} Therefore, we divide the whole brain GMV into 90 regions of interest (ROIs) according to the AAL-90 atlas. By using the volume modulation map generated by the DARTEL method, the cross-sectional between-group comparisons and longitudinal intragroup comparisons are statistically analyzed using the two-sample T-test and paired T test, respectively. At the same time, the total intracranial volumes (TIV) are regressed to remove the effect of their differences on the results. Finally, we use FWE correction ($P < .05$) for the cross-sectional between-group comparisons and use uncorrected $P < .05$ for longitudinal intragroup comparisons to control the effect of false positive results.

2.3 | Machine learning method and analysis

The goal of machine learning is to construct a classifier that can accurately predict the category of a new subject. If it successfully predicts the category, the classifier captures the difference between the two groups. The analysis framework of machine learning in neuroimaging data research mainly includes the following five parts: feature

TABLE 1 Participant demographics and clinical information

Characteristic	Sample size	Gender (male/female)	Age (years)	MMSE	CDR
NC	100	54/46	75.53 ± 5.34	29.25 ± 0.91	0.03 ± 0.11
NC_12 m				29.21 ± 1.17	0.05 ± 0.26
NC_24 m				29.30 ± 0.89	0.06 ± 0.27
sMCI	58	42/16	74.48 ± 7.63	27.69 ± 1.73	1.26 ± 0.58
sMCI_12 m				27.97 ± 1.97	1.49 ± 1.05
sMCI_24 m				27.55 ± 2.51	1.72 ± 1.40
cMCI	53	36/17	75.30 ± 6.55	26.43 ± 1.77	1.81 ± 1.07
cMCI_12 m				24.85 ± 2.81	3.05 ± 1.48
cMCI_24 m				22.72 ± 3.55	4.65 ± 1.56
AD	106	54/52	75.33 ± 7.35	23.29 ± 1.97	4.31 ± 1.54
AD_12 m				21.12 ± 4.62	5.67 ± 2.70
AD_24 m				18.83 ± 5.79	7.53 ± 3.55

Note: The data are represented as mean ± SD (SD).

extraction, feature selection, model training and testing, evaluation of model prediction ability and location of features that contribute to prediction.

In this paper, we use a SVM with a linear kernel (LSVM) combined with the nested leave-one-out cross-validation (LOOCV) method and use GMV as a feature to classify the four groups. Then, we use accuracy and another index to quantify the performance of the classifier.

As GMV is the unique feature in this study, we adopt the strategy of longitudinal feature combination, which is to combine the features of each time point.¹⁷ There is still a lack of research on the use of this strategy. We think that it may have a better ability to distinguish between groups because it can dig into the existing differences between groups.

2.3.1 | Feature selection

The feature selection method is an important step of machine learning. It can select features with significant differences and filter out redundant features. It has been extensively used to improve the performance of classifiers.^{18,19} This study applies a nested LOOCV using

the outer loop to estimate classification accuracy and the inner loop to determine the optimal feature selection. Feature selection using inner LOOCVs avoids overfitting. The process of this approach is shown as Figure 1: the selection of $N-1$ subjects as the training set for each outer LOOCV fold, and the remaining subject is used as the test sample, where N is the number of all subjects. Inner LOOCVs are further applied in each of the outer LOOCV folds. A two-sample t -test is conducted on each of the features in the training set for each of the inner LOOCVs ($N-2$ subjects), which yields a p value for each feature. A P threshold from 0 to 1 with a .01 interval is applied, and the features below the p threshold are retained and above the p threshold are excluded. The p threshold is applied for each inner LOOCV, resulting in 99 inner LOOCVs in total, and 99 classification accuracies are obtained for each inner LOOCV. We take the accuracy of each p value and the average accuracy corresponding to the two p values adjacent to it as the final classification accuracy, which is similar to the smoothing operation in image processing to control some outliers or noise effects. The optimal p threshold is used to train the final classifier on the $N-1$ training samples of the outer LOOCV. The obtained classifier is used to predict the category of the

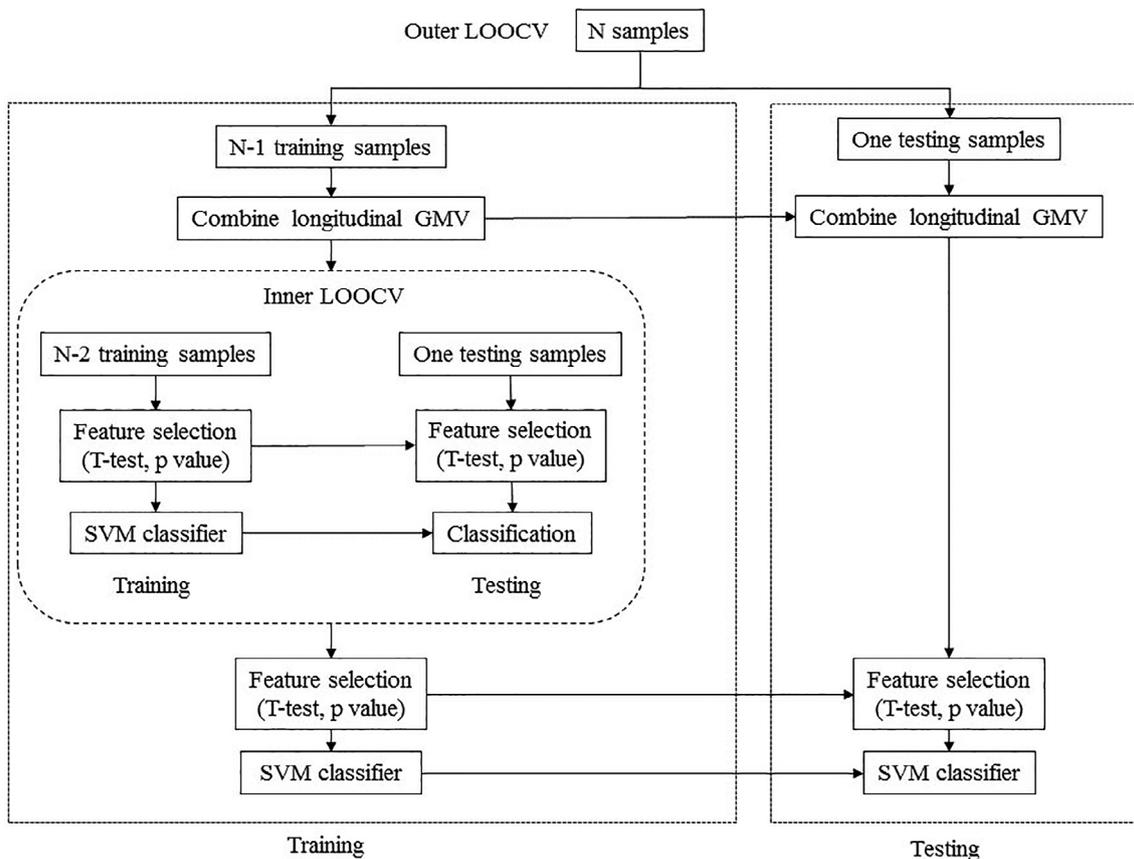


FIGURE 1 Nested LOOCV classification process framework using the GMV feature

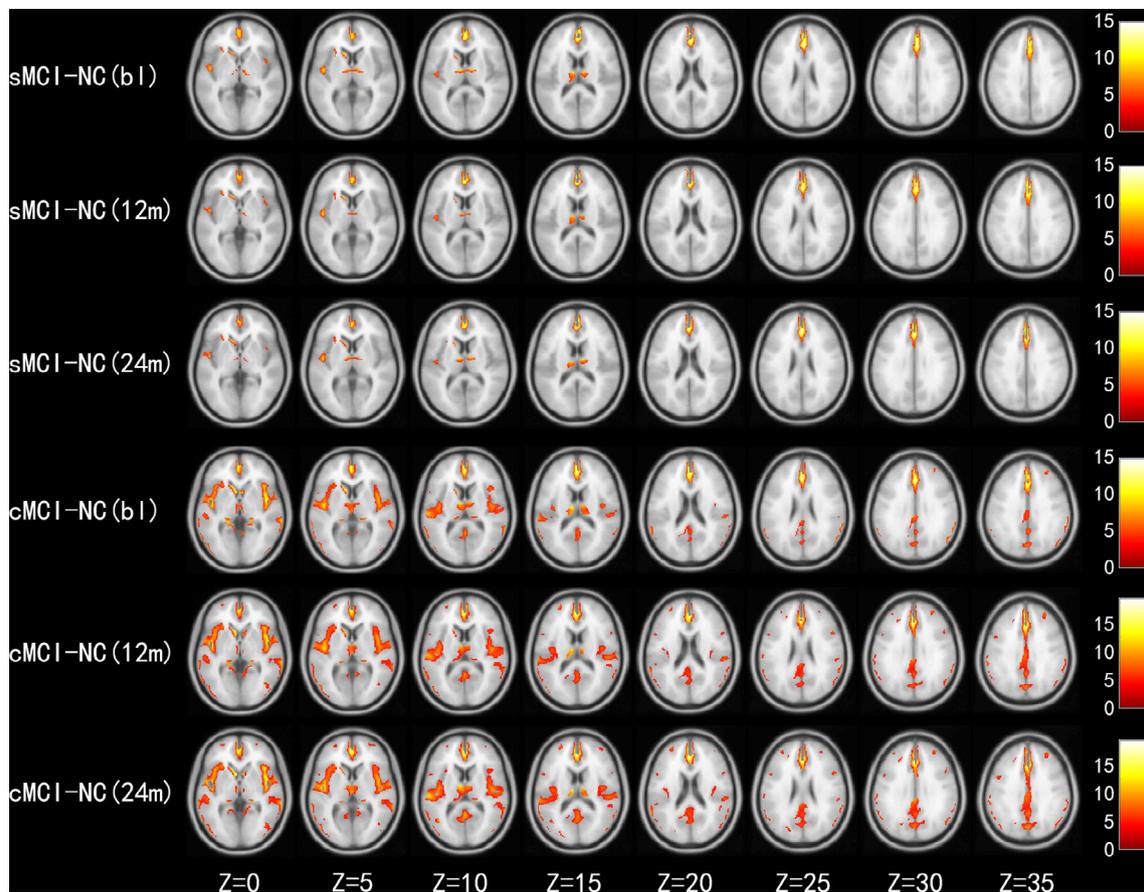


FIGURE 2 Significantly different regions in sMCI-NC and cMCI-NC comparisons at different scanning time ($p\text{-FWE} < 0.05$). Time point: baseline, b1; 12 months, 12 m; 24 months, 24 m. Red and yellow represent the decreased gray matter volume of brain regions [Color figure can be viewed at wileyonlinelibrary.com]

remaining test sample. The process is repeated N times so that each sample is tested once.

2.3.2 | SVM implementation

The SVM is one of the most widely used supervised machine learning methods, which takes into account the experience error and the complexity of the model at the same time to train a classifier with higher prediction capability.^{20,21} We apply the LIBSVM toolbox for MATLAB to implement the SVM classification (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>).²² It has a parameter C to control the tradeoff between empirical errors and model extensibility, which is set to the default value of 1. Specifically, each subject can be regarded as a point in multidimensional space, and each dimension is associated with a feature. The SVM algorithm uses some training data as input to fit a hyperplane in this multidimensional space, which best separates the input data into two categories matching with the known class labels. The hyperplane is represented as a decision function $y = f(x)$, in which y is the classification score and x is the multidimensional

feature vector. Once the decision function is learned from the training data, it will be implemented to predict the class of new testing samples.

2.3.3 | Evaluation of classification performance

Accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) are computed to quantify the classification performance. Specifically, accuracy is the proportion of subjects who are correctly classified into the disease group or control group. Sensitivity and specificity are the proportion of patients and controls classified correctly, respectively. PPV and NPV are the proportion of correct patient predictions and control predictions, respectively. Furthermore, we use receiver operating characteristic (ROC) analysis to evaluate the performance of the classifiers. The area under the ROC curve (AUC) represents the classification power of a classifier, and a larger AUC indicates a better classification power.²³ The ROC curve is generated using sequential thresholding at the classification score of each subject.

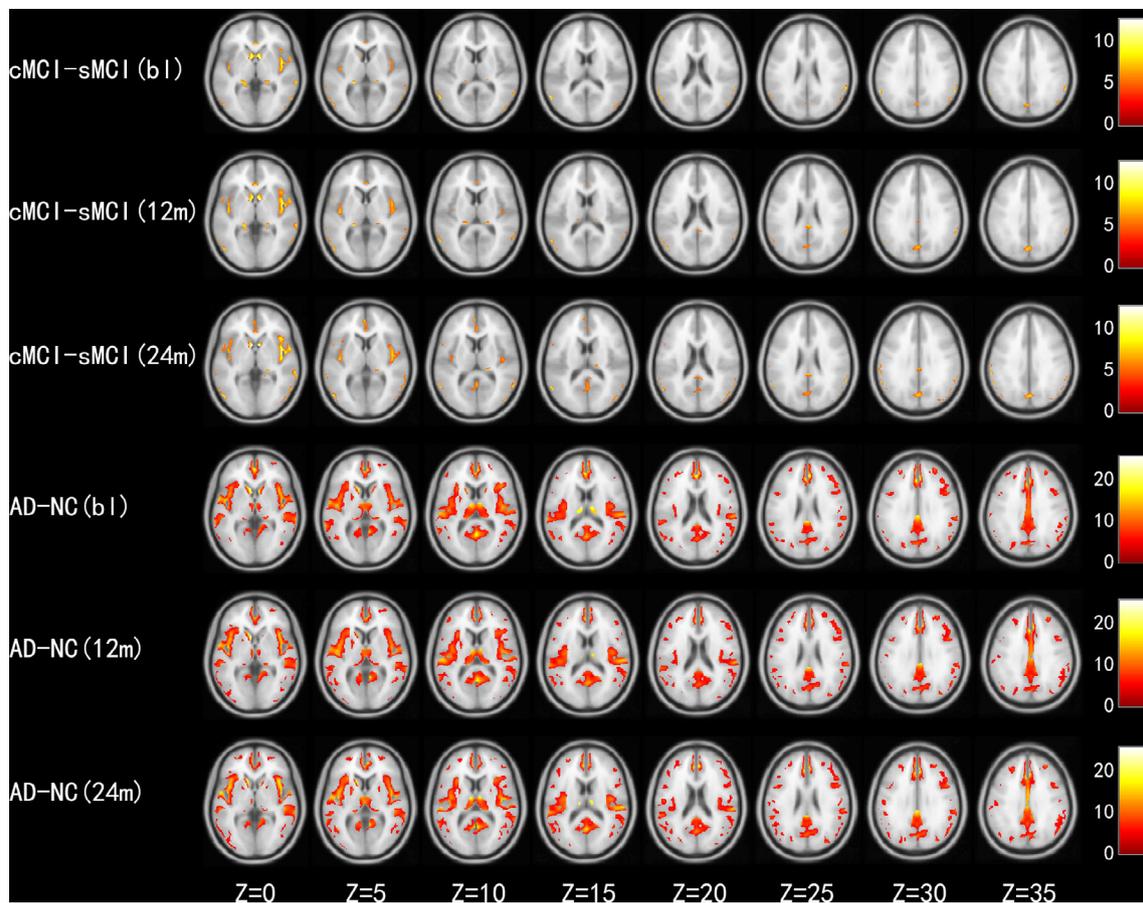


FIGURE 3 Significantly different regions in cMCI-sMCI and AD-NC comparisons at different scanning time ($p\text{-FWE} < 0.05$). Time point: baseline, bl; 12 months, 12 m; 24 months, 24 m. Red and yellow represent the decreased gray matter volume of brain regions [Color figure can be viewed at wileyonlinelibrary.com]

2.3.4 | Discriminative features

Feature selection in each fold of the outer LOOCV is implemented using a slightly different sample subset, which leads to a different set of selected features across folds. The recurrent features that are selected on all folds of the outer LOOCV are defined as the discriminative features as described previously. The discriminative weight for each feature is defined as the average of their absolute weights across all folds. A higher absolute value of the discriminative weight indicates a greater contribution of the corresponding feature to the classification.

3 | RESULTS

3.1 | Voxel-based morphometry analysis

The different patterns of the whole brain are found in each time point in Figures 2, 3 and 4. According to the pattern of changes in the whole brain, AD-NC

comparison is the most significant difference among all between-group comparisons, and the decrease of GMV is almost all over the brain. cMCI-NC comparison and AD-sMCI comparison take the second place. cMCI will gradually develop into AD with progression of disease, yet sMCI remains stable. According to the comparison of the degree of whole-brain GM atrophy, we find that cMCI-NC is more serious than AD-sMCI. cMCI and sMCI are two subtypes of MCI, and the lesion degree between them expands gradually with the development of time. Finally, compared with sMCI-NC and AD-cMCI, sMCI stays in the stable stage for a long time, while cMCI will eventually transform into AD, which is more similar to AD. Therefore, the difference in AD-cMCI comparison is smaller than that of sMCI-NC comparison.

The GMV change patterns of the whole brain in different time periods (ie, 12 m-bl and 24 m-bl), as shown in Figure 5. It is obvious that the development pattern of the NC group is similar to that of the sMCI group. A similar development pattern was observed for the cMCI group and the AD group.

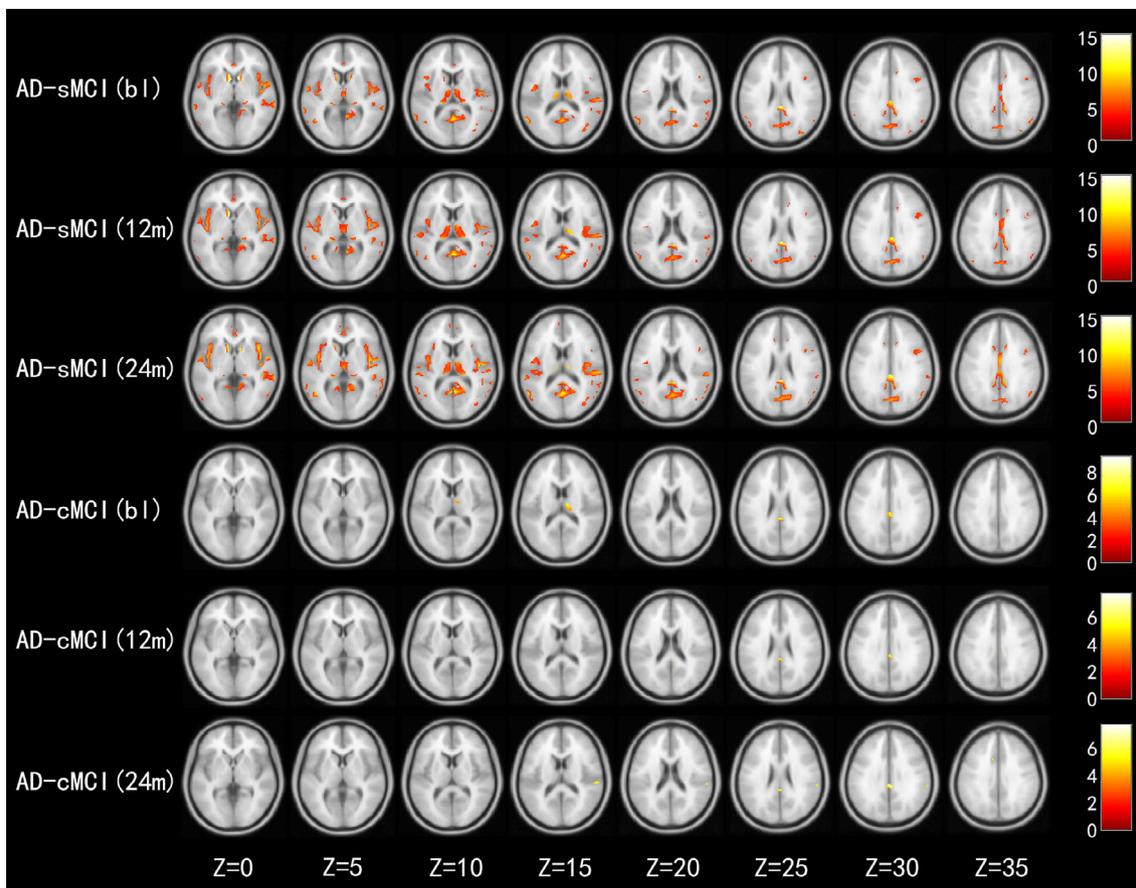


FIGURE 4 Significantly different regions in AD-cMCI and AD-sMCI comparisons at different scanning time (p -FWE < 0.05). Time point: baseline, bl; 12 months, 12 m; 24 months, 24 m. Red and yellow represent the decreased gray matter volume of brain regions [Color figure can be viewed at wileyonlinelibrary.com]

3.2 | Classification performance

The SVM classification of the four groups is performed using the GMV features. The classification performance is reported using accuracy, sensitivity, specificity, PPV, NPV and AUC. The criteria for determining the best feature dimension is that the SVM should achieve the best comprehensive performance that would consist of a balance between accuracy, sensitivity, specificity, PPV, NPV and AUC. The primary measure considered is the accuracy, followed by the sensitivity, specificity, PPV, NPV and AUC. Each subject has 90 baseline GMV features and 90 longitudinal GMV features at each time point (eg, 12 and 24 months).

We first calculate the classification performance without feature selection (Table 2). By comparing Tables 2 and 3 (with feature selection), it is obvious that feature selection can improve the classification performance to some extent. As shown in Table 3, when using the combined features from all time points, the accuracies, sensitivities, specificities, PPVs, NPVs and AUCs of all between-group comparisons achieve great improvement. No matter in the baseline

period or after the longitudinal combination, the classification performance of AD-NC comparison is higher than that of other between-group comparisons. After the longitudinal feature combination strategy, the classification performance of cMCI-sMCI comparison improves most obviously, and the accuracy rate increases from 74.77% to 91.89%. In addition, the classification performance of AD-cMCI comparison is significantly improved, along with an AUC increase from 72.41% to 91.05%. The sensitivity of sMCI-NC and cMCI-sMCI comparisons increases from 63.79% to 89.08% and from 69.81% to 90.57%, respectively. The specificity and NPV of AD-cMCI comparison increases from 39.62% to 70.44% and from 53.85% to 80.00%, respectively. With the exception of AD-cMCI comparison, the accuracies among other between-group comparisons are over 90% after longitudinal feature combination. Specifically, using the selected baseline features, the SVM produces poor accuracies ranging from 68.55% to 94.66%, low sensitivities ranging from 63.79% to 95.28%, low specificities ranging from 39.62% to 94.00%, low PPVs ranging from 73.33% to 94.39%, low NPVs ranging from 53.85% to 94.95% and low AUCs ranging from 72.41% to 97.22% for all the comparisons. However, the

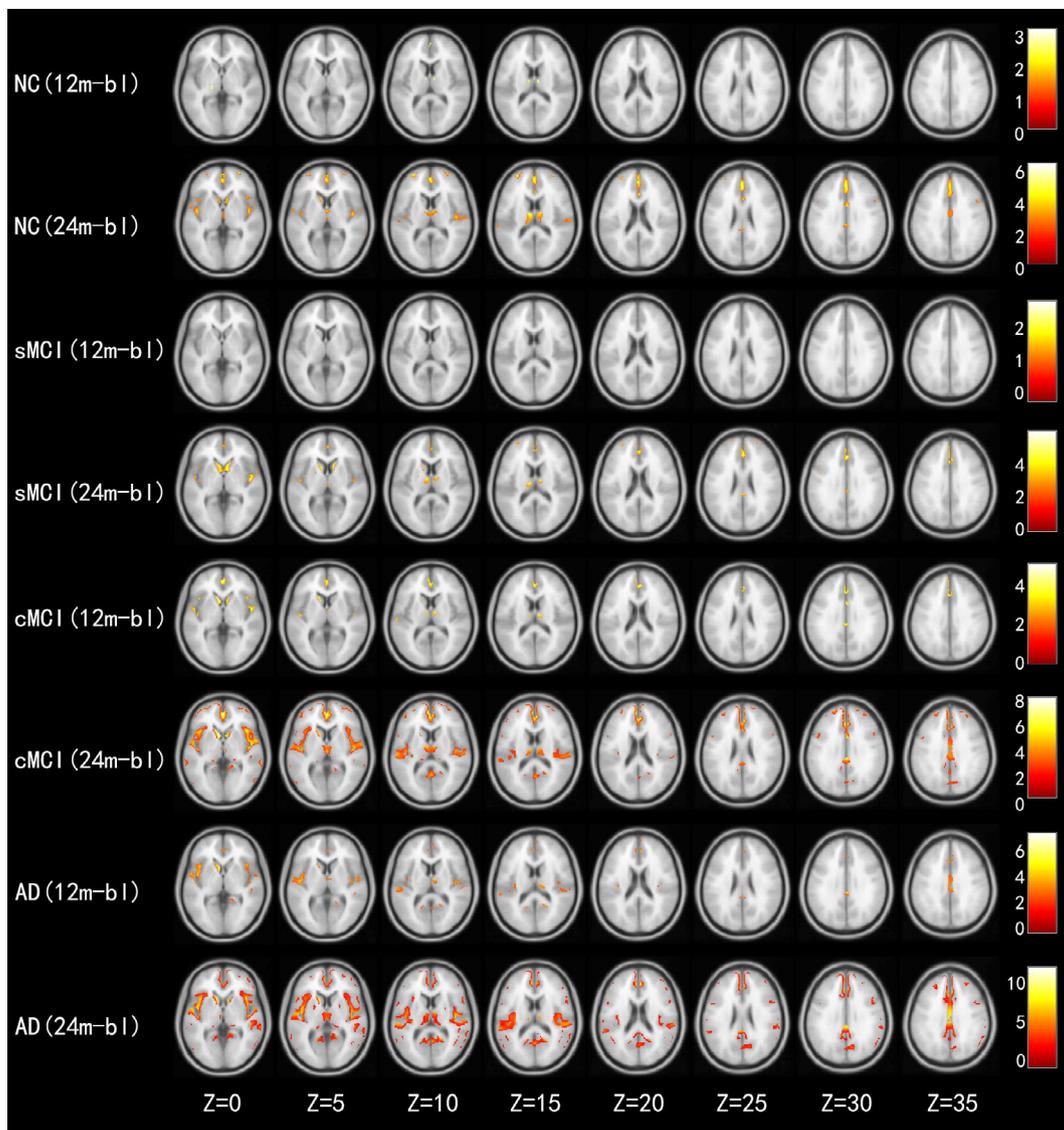


FIGURE 5 The GMV changes of intragroup during longitudinal comparison (uncorrected $P < .01$). Red and yellow represent the decreased gray matter volume of brain regions [Color figure can be viewed at wileyonlinelibrary.com]

classifier performs better using the selected baseline and longitudinal features in 12 months with accuracies of 78.62% to 97.57%, sensitivities of 76.72% to 97.17%, specificities of 60.38% to 98.00%, PPVs of 81.58% to 98.10%, NPVs of 71.11% to 97.49% and AUCs of 84.76% to 99.10%, and it achieved the highest performance with accuracies of 84.28% to 98.06%, sensitivities of 89.08% to 97.17%, specificities of 70.44% to 99.00%, PPVs of 86.05% to 99.04%, NPVs of 80.00% to 97.36% and AUCs of 91.05% to 99.49%.

3.3 | Distinguishing GM feature

Features with different weight values will be obtained in the process of classification and will reflect the contribution to the classification. As Table 4 shows, we apply the

combination of three scanning time points (bl + 12 m + 24 m) to obtain the weight values. According to statistics, caudate nucleus, hippocampus, temporal pole and lenticular putamen are the most important contribution areas to the between-group classification.

4 | DISCUSSION

As is known to all, voxel-based morphometry (VBM) analysis has been widely used in cross-sectional and longitudinal studies to detect the difference between the aged and AD patients and predicts the transformation of MCI to AD.²⁴ We observe the GMV change pattern of four groups in the whole brain from a cross-sectional and longitudinal viewpoint, especially the changes in sMCI

TABLE 2 Classification performance without feature selection

Between-group comparison	Feature number	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AUC (%)
sMCI-NC(bl)	90	74.05	51.72	87.00	69.77	75.65	79.45
sMCI-NC(bl + 12 m)	180	80.70	67.24	88.50	77.23	82.33	89.95
sMCI-NC(bl + 12 m + 24 m)	270	85.44	71.26	93.67	86.71	84.89	93.12
cMCI-NC(bl)	90	84.97	75.47	90.00	80.00	87.38	92.28
cMCI-NC(bl + 12 m)	180	91.83	89.62	93.00	87.16	94.42	97.59
cMCI-NC(bl + 12 m + 24 m)	270	93.03	88.05	95.67	91.50	93.79	98.34
sMCI-cMCI(bl)	90	74.77	75.47	74.14	72.73	76.79	80.45
sMCI-cMCI(bl + 12 m)	180	89.19	88.68	89.66	88.68	89.66	94.20
sMCI-cMCI(bl + 12 m + 24 m)	270	93.39	94.97	91.95	91.52	95.24	96.48
AD-NC(bl)	90	92.23	92.45	92.00	92.45	92.00	96.27
AD-NC(bl + 12 m)	180	95.63	94.34	97.00	97.09	94.17	98.07
AD-NC(bl + 12 m + 24 m)	270	96.60	95.60	97.67	97.75	95.44	98.92
AD-sMCI(bl)	90	86.59	89.62	81.03	89.62	81.03	92.65
AD-sMCI(bl + 12 m)	180	93.60	93.87	93.10	96.14	89.26	97.83
AD-sMCI(bl + 12 m + 24 m)	270	94.92	96.23	92.53	95.92	93.06	98.75
AD-cMCI(bl)	90	66.67	80.19	39.62	72.65	50.00	71.61
AD-cMCI(bl + 12 m)	180	76.10	85.85	56.60	79.82	66.67	83.58
AD-cMCI(bl + 12 m + 24 m)	270	79.87	88.36	62.89	82.65	72.99	85.51

Abbreviations: ACC, accuracy; SEN, sensitivity; SPE, specificity.

TABLE 3 Classification performance after feature selection

Between-group comparison	Optimal feature dimension	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AUC (%)
sMCI-NC(bl)	65	79.75	63.79	89.00	77.08	80.91	85.93
sMCI-NC(bl + 12 m)	69	87.34	76.72	93.50	87.25	87.38	96.09
sMCI-NC(bl + 12 m + 24 m)	84	92.83	89.08	95.00	91.18	93.75	98.49
cMCI-NC(bl)	61	83.66	73.58	89.00	78.00	86.41	93.89
cMCI-NC(bl + 12 m)	84	96.41	95.28	97.00	94.39	97.49	98.83
cMCI-NC(bl + 12 m + 24 m)	87	97.17	94.97	98.33	96.79	97.36	99.39
sMCI-cMCI(bl)	56	74.77	69.81	79.31	75.51	74.19	82.30
sMCI-cMCI(bl + 12 m)	77	91.44	91.51	91.38	90.65	92.17	93.94
sMCI-cMCI(bl + 12 m + 24 m)	79	91.89	90.57	93.10	92.31	91.53	96.84
AD-NC(bl)	86	94.66	95.28	94.00	94.39	94.95	97.22
AD-NC(bl + 12 m)	87	97.57	97.17	98.00	98.10	97.03	99.10
AD-NC(bl + 12 m + 24 m)	87	98.06	97.17	99.00	99.04	97.06	99.49
AD-sMCI(bl)	70	87.20	89.62	82.76	90.48	81.36	92.52
AD-sMCI(bl + 12 m)	83	93.90	94.34	93.10	96.15	90.00	98.07
AD-sMCI(bl + 12 m + 24 m)	85	96.14	96.23	95.98	97.76	93.30	98.67
AD-cMCI(bl)	37	68.55	83.02	39.62	73.33	53.85	72.41
AD-cMCI(bl + 12 m)	54	78.62	87.74	60.38	81.58	71.11	84.76
AD-cMCI(bl + 12 m + 24 m)	74	84.28	91.19	70.44	86.05	80.00	91.05

Abbreviations: ACC, accuracy; SEN, sensitivity; SPE, specificity.

TABLE 4 The discriminative features of SVM classifier

sMCI-NC	Weight	cMCI-NC	Weight
Caudate nucleus (L)	0.408	Hippocampus (R)	0.339
Temporal pole: superior temporal gyrus (L)	0.294	Caudate nucleus (L)	0.337
Hippocampus (R)	0.251	Insula (R)	0.277
Superior frontal gyrus, medial (L)	0.228	Lenticular nucleus, putamen (L)	0.274
Heschl gyrus (L)	0.209	Temporal pole: superior temporal gyrus (L)	0.260
Calcarine fissure (R)	0.208	Hippocampus (L)	0.260
Lingual gyrus (R)	0.180	Temporal pole: middle temporal gyrus (L)	0.227
Lenticular nucleus, putamen (R)	0.174	Calcarine fissure (R)	0.214
Lenticular nucleus, putamen (L)	0.162	Lenticular nucleus, putamen (R)	0.184
Lenticular nucleus, putamen (R)	0.161	Inferior parietal gyrus (R)	0.183
cMCI-sMCI	Weight	AD-NC	Weight
Temporal pole: superior temporal gyrus (R)	0.260	Caudate nucleus (L)	0.403
Hippocampus (L)	0.227	Lenticular nucleus, putamen (L)	0.371
Temporal pole: middle temporal gyrus (R)	0.223	Hippocampus (R)	0.359
Olfactory cortex (L)	0.213	Temporal pole: superior temporal gyrus (L)	0.316
Posterior cingulate gyrus (R)	0.206	Hippocampus (L)	0.274
Parahippocampal gyrus (L)	0.204	Temporal pole: superior temporal gyrus (R)	0.174
Insula (R)	0.200	Calcarine fissure (R)	0.169
Superior parietal gyrus (R)	0.191	Temporal pole: middle temporal gyrus (L)	0.150
Precentral gyrus (R)	0.184	Heschl gyrus (L)	0.139
Inferior parietal gyrus (R)	0.182	Insula (R)	0.138
AD-sMCI	Weight	AD-cMCI	Weight
Temporal pole: superior temporal gyrus (R)	0.306	Parahippocampal gyrus (L)	0.319
Posterior cingulate gyrus (R)	0.300	Median cingulate gyrus (R)	0.266
Hippocampus (L)	0.243	Caudate nucleus (R)	0.246
Middle frontal gyrus (R)	0.222	Temporal pole: superior temporal gyrus (L)	0.239
Gyrus rectus (R)	0.188	Temporal pole: middle temporal gyrus (R)	0.225
Middle frontal gyrus (R)	0.182	Inferior parietal gyrus (L)	0.215
Calcarine fissure (L)	0.178	Inferior occipital gyrus (R)	0.214
Heschl gyrus (R)	0.175	Lenticular nucleus, putamen (R)	0.209
Inferior frontal gyrus, triangular part (R)	0.148	Posterior cingulate gyrus (R)	0.201
Inferior temporal gyrus (R)	0.143	Precuneus (R)	0.195

Abbreviations: L, left hemisphere; R, right hemisphere.

and cMCI. The GMV atrophy of cMCI from the baseline to 12 months appears to be significantly higher than that of sMCI. From the longitudinal pattern of AD-cMCI comparison, it is possible to speculate that the brain atrophy pattern of cMCI is similar to that of AD because the cMCI patients will develop into AD within 2 years. Similarly, for the longitudinal pattern of sMCI-NC comparison, the sMCI patients remains stable for 2 years, and it is possible to speculate that the brain atrophy is slow, which is similar to that of the NC group. The GM atrophy rate of AD-NC comparison is significantly faster than

that of other groups, which also reflects that MCI is a transitional state between NC and AD. In addition, from the viewpoint of differences in the longitudinal intra-group comparisons, the NC group is roughly the same as the sMCI group, and the cMCI group is roughly the same as that of the AD group, which reflected the development trend of two different subtypes of MCI.

In the classification of the four groups, we use the SVM with linear kernel and combine the nested LOOCV method. This method also has a feature selection function. Table 3 shows that the nested LOOCV method can

select the optimal feature dimension for different between-group comparisons and receives the best classification performance in this dimension. The accuracy rates obtained by the classification method and the feature combination strategy are all higher than those of previous studies.²⁵⁻²⁸ The accuracies, sensitivities, specificities, PPVs, NPVs and AUCs have been greatly improved after the combination of longitudinal time points (ie, bl + 12 m and bl + 12 m + 24 m). The reason for high classification performance of the classifier model mainly depends on the following two aspects¹: the nested LOOCV method used in this study has the function of feature selection and can select the best feature dimension²; compared with single feature classification, the strategy of longitudinal feature combination greatly improves the classification performance of the classifier model. These findings suggest that longitudinal change is a crucial factor for the prediction of future conversion of MCI to AD. The longitudinal changes in the GMV could provide effective information for discrimination and improved the performance of the classifier.

In order to link the discriminative features extracted from the classifier with the pathological mechanism of AD, we list the features/brain regions that have important contributions on between-group classification (Table 4). In earlier anatomic studies of some nonhuman primates, the temporal pole was considered to be an important area of interaction that mediated the marginal cortex and interconnected with visual, auditory, olfactory and multichannel-associated cortex.^{29,30} In addition, it was considered to be important for visual recognition, nonverbal memory, spontaneous activity, sociality and other complex behaviors in human and primate behavior studies.³¹⁻³³ Several early studies of AD also confirmed that the temporal pole was one of the most severely damaged cortical areas.^{34,35} The temporal pole extracted from the classification model in this study has a high weight value in each between-group comparison, indicating that the temporal pole is an important area for identifying AD, which indirectly provides clues to its pathological mechanism. De et al³⁶ found that the volume of the bilateral lenticular putamen in AD significantly decreased compared with that in MCI, and the reduced lenticular putamen was positively correlated with cognitive function. However, few studies have been conducted on the role of the lenticular putamen in cognition. As a part of the striatum, it had been found that the lenticular putamen was activated in the probability learning task^{37,38} and working memory task.³⁹ The reason for the cognitive effects of the lenticular putamen in AD is because of the damage of the primary cognitive function in lenticular putamen or the incomplete neurochemical function or the disconnected cortical thalamic projection, which

remains unknown. So, it will be a hot spot in the future. In addition to the lenticular putamen, which has a significant contribution to the classification, this study also found that the caudate nucleus plays an important role in the classification of AD. A number of studies^{40,41} found that the volume of the caudate nucleus in AD patients is significantly lower than that in NCs. The caudate nucleus plays a vital role in acquiring explicit memory and in sport learning,^{42,43} and its damage will result in the implementation of control and cognition.⁴⁴ Therefore, the combination of the caudate nucleus and other regions may be an effective target in exploring the biomarkers of AD. In addition, a large number of research⁴⁵⁻⁴⁹ point out the hippocampal area was seriously injured in AD patients and caused episodic memory disorder, which also played a key role in guiding the clinical symptoms of patients.

5 | CONCLUSION AND PERSPECTIVE

This study mainly analyzes AD and MCI in two aspects. First, the GMV is used as a feature to explore the changes in the brain structure in AD and MCI. In this study, the differences among the NC, sMCI, cMCI and AD groups are compared, and the longitudinal time course is used to prove that the brain structure gradually changes in the development of NC to AD. Therefore, it is proved that MCI is the transition stage from NC to AD from the perspective of brain structural change. Second, linear SVM with nested LOOCV are used to classify the four groups of population. The longitudinal combination strategy is used to improve the classification performance and excavate the features that have a distinguished ability, which provides a new idea for the diagnosis of AD in the future.

In the follow-up study, we hope to focus on the following aspects: (a) cooperation with the hospital to obtain different subtypes of AD, mining internal patterns of differences, to establish a more specific diagnostic model; (b) in order to seek a more sensitive and effective biological feature for the transformation of MCI, more imaging modes and biomarkers should be added to compare; (c) seek a better classification algorithm to further improve the diagnostic performance; and (d) a more comprehensive analysis should be carried out on the level of brain network to study the process of the change of disease topological pattern.

ACKNOWLEDGMENTS

This work has been supported by the National Natural Science Foundation of China (Grant Nos. 11572127, 11872183 and 11702064).

ORCID

Yingteng Zhang  <https://orcid.org/0000-0002-7456-5579>

REFERENCES

- Jack CR Jr, Wiste HJ, Vemuri P, et al. Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain*. 2010;133(11):3336-3348.
- Hinrichs C, Singh V, Xu G, Johnson SC, Alzheimer's Disease Neuroimaging Initiative. Predictive markers for AD in a multimodality framework: an analysis of MCI progression in the ADNI population. *Neuroimage*. 2011;55(2):574-589.
- Landau SM, Harvey D, Madison CM, et al. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*. 2010;75(3):230-238.
- Killiany RJ, Gomez-Isla T, Moss M, et al. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. *Ann Neurol*. 2010;47(4):430-439.
- Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging*. 2011;32(12):2322-2327.
- Zhang D, Shen D. Alzheimer's disease neuroimaging I. predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS One*. 2012;7(3):e33182.
- Hyun JW, Li Y, Huang C, Styner M, Lin W, Zhu H. STGP: Spatio-temporal Gaussian process models for longitudinal neuroimaging data. *Neuroimage*. 2016;134:550-562.
- Li Y, Wang Y, Wu G, et al. Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. *Neurobiol Aging*. 2012;33(2):427.e15-e30.
- Risacher SL, Shen L, West JD, et al. Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. *Neurobiol Aging*. 2010;31(8):1401-1418.
- Mcevoy LK, Holland D, Hagler DJ, et al. Mild cognitive impairment: baseline and longitudinal structural MR imaging measures improve predictive prognosis. *Radiology*. 2011;259(3):834-843.
- Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage*. 2007;38(1):95-113.
- Ashburner J, Friston KJ. Unified segmentation. *Neuroimage*. 2005;26(3):839-851.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002;15(1):273-289.
- Zhengjia D, Chaogan Y, Zhiqun W, et al. Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). *Neuroimage*. 2012;59(3):2187-2195.
- Chong-Yaw W, Pew-Thian Y, Daoqiang Z, et al. Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage*. 2012;59(3):2045-2056.
- Zeng LL, Shen H, Liu L, Wang LB, Li BJ, Fang P, Zhou Z, Li Y, Hu D Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. *Brain* 2012; 135(Pt 5):1498-507.
- Guo S, Lai C, Wu C, Cen G. Alzheimer's disease neuroimaging I. conversion discriminative analysis on mild cognitive impairment using multiple cortical features from MR images. *Front Aging Neurosci*. 2017;9:146.
- Hahn T, Kircher T, Straube B, et al. Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. *JAMA Psychiat*. 2015;72(1):68-74.
- Robert W, Richard W, Orr CA, et al. Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature*. 2014;512(7513):185-189.
- Cherkassky V. The nature of statistical learning theory. *IEEE Trans Neural Netw*. 1997;8(6):1564.
- Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Machine Learning*. 2002;46:34.
- Chang CC, Lin CJ. LIBSVM: a library for support vector machines. 2011;2:1-27.
- Fawcett T. An introduction to ROC analysis. *Pattern Recog Lett*. 2006;27(8):861-874.
- Yang H, Liu W, Xia H, Zhou Z, Tong L, editors. Longitudinal change of the grey matter of mild cognitive impairment patients over 3 years by using voxel-based morphometry. International Conference on Biomedical Engineering and Informatics; 2013.
- Zhang J, Liu M, Le A, Gao Y, Shen D. Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images. *IEEE J Biomed Health Inform*. 2017;21(6):1607-1616.
- Sun Z, van de Giessen M, Lelieveldt BP, Staring M. Detection of conversion from mild cognitive impairment to Alzheimer's disease using longitudinal brain MRI. *Front Neuroinform*. 2017; 11:16.
- Platero C, Lin L, Tobar MC. Longitudinal neuroimaging hippocampal markers for diagnosing Alzheimer's disease. *Neuroinformatics*. 2019;17(1):43-61.
- Huang M, Yang W, Feng Q, Chen W. Alzheimer's disease neuroimaging I. longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease. *Sci Rep*. 2017;7:39880.
- Morán MA, Mufson EJ, Mesulam MM. Neural inputs into the temporopolar cortex of the rhesus monkey. *J Comp Neurol*. 2010;256(1):88-103.
- Markowitsch HJ, Emmans D, Irle E, Streicher M, Preilowski B. Cortical and subcortical afferent connections of the primate's temporal pole: a study of rhesus monkeys, squirrel monkeys, and marmosets. *J Comp Neurol*. 2010;242 (3):425-458.
- Horel JA, Voytko ML, Salsbury KG. Visual learning suppressed by cooling the temporal pole. *Behav Neurosci*. 1984;98(2):310-324.
- Rausch R, Serafetinides EA, Crandall PH. Olfactory memory in patients with anterior temporal lobectomy. *Cortex*. 1977;13(4):445-452.
- Franzen EA, Myers RE. Neural control of social behavior: prefrontal and anterior temporal cortex. *Neuropsychologia*. 1973;11 (2):141-157.
- Arnold SE, Hyman BT, Flory J, Damasio AR, Van Hoesen GW. The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. *Cereb Cortex*. 2012;1(1):103.

35. Arnold SE, Hyman BT, Van Hoesen GW. Neuropathologic changes of the temporal pole in Alzheimer's disease and Pick's disease. *Arch Neurol*. 1994;51(2):145-150.
36. de Jong LW, van der Hiele K, Veer IM, Houwing JJ, Westendorp RG, Bollen EL, et al. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain* 2008;131(Pt 12):3277-85.
37. Graybiel AM. The basal ganglia: learning new tricks and loving it. *Curr Opin Neurobiol*. 2005;15(6):638-644.
38. Christian B, Benno K, Michael S, Irene D. Focal basal ganglia lesions are associated with impairments in reward-based reversal learning. *Brain*. 2008;131(Pt 3):829.
39. Erika D, Anna Stigsdotter N, Anne L, Lars BC, Lars N. Transfer of learning after updating training mediated by the striatum. *Science*. 2008;320(5882):1510-1512.
40. Good CD, Scahill RI, Fox NC, et al. Automatic differentiation of anatomical patterns in the human brain: validation with studies of degenerative dementias. *Neuroimage*. 2002;17(1):29-46.
41. Rombouts SA, Barkhof F, Witter MP, Scheltens P. Unbiased whole-brain analysis of gray matter loss in Alzheimer's disease. *Neurosci Lett*. 2000;285(3):231-233.
42. Nakamura T, Ghilardi MF, Mentis M, et al. Functional networks in motor sequence learning: abnormal topographies in Parkinson's disease. *Hum Brain Mapp*. 2001;12(1):42-60.
43. Knowlton BJ, Mangels JA, Squire LR. A neostriatal habit learning system in humans. *Science*. 1996;273(5280):1399-1402.
44. Rubin DC. Frontal-striatal circuits in cognitive aging: evidence for caudate involvement. *Aging Neuropsychol Cognition*. 1999;6(4):241-259.
45. Tsao SC, Gajawelli N, Zhou JY, et al. Feature selective temporal prediction of Alzheimer's disease progression using hippocampus surface morphometry. *Brain and Behavior*. 2017;7(7):e00733.
46. Stepan-Buksakowska I, Szabo N, Horinek D, et al. Cortical and subcortical atrophy in Alzheimer disease parallel atrophy of thalamus and hippocampus. *Alzheimer Dis Associat Disorders*. 2014;28(1):65-72.
47. Jack CR Jr, Petersen RC, Xu YC, et al. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*. 1999;52(7):1397-1403.
48. Morris JC, Csernansky J, Price JL. MRI measures of entorhinal cortex versus hippocampus in preclinical AD. *Neurology*. 2002;59(9):1474-1475.
49. Jack CR, Xu YC, Petersen RC. Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD - reply. *Neurology*. 2001;56(6):820-821.

How to cite this article: Zhang Y, Liu S, Yu X. Longitudinal structural MRI analysis and classification in Alzheimer's disease and mild cognitive impairment. *Int J Imaging Syst Technol*. 2020;30:421-433. <https://doi.org/10.1002/ima.22390>