Heliyon 7 (2021) e07287

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identifying and characterizing different stages toward Alzheimer's disease using ordered core features and machine learning

Jinhua Sheng [a,b,*], Bocheng Wang [a,b,c], Qiao Zhang [d], Rougang Zhou [b,e,f], Luyun Wang [a,b], Yu Xin [a,b]

[a] College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, Zhejiang, 310018, China
[b] Key Laboratory of Intelligent Image Analysis for Sensory and Cognitive Health, Ministry of Industry and Information Technology of China, Hangzhou, Zhejiang, 310018, China
[c] Communication University of Zhejiang, Hangzhou, Zhejiang, 310018, China
[d] Beijing Hospital, Beijing, 100730, China
[e] College of Mechanical Engineering, Hangzhou Dianzi University, Hangzhou, Zhejiang, 310018, China
[f] Mstar Technologies Inc., Hangzhou, Zhejiang, 310018, China

ARTICLE INFO

Keywords:
Alzheimer's disease (AD)
Mild cognitive impairment (MCI)
Machine learning
The area under ROC curve (AUC)
Human connectome project (HCP)
Connectivity network analysis

ABSTRACT

Based on the joint HCPMMP parcellation method we developed before, which divides the cortical brain into 360 regions, the concept of ordered core features (OCF) is first proposed to reveal the functional brain connectivity relationship among different cohorts of Alzheimer's disease (AD), late mild cognitive impairment (LMCI), early mild cognitive impairment (EMCI) and healthy controls (HC). A set of core network features that change significantly under the specifically progressive relationship were extracted and used as supervised machine learning classifiers. The network nodes in this set mainly locate in the frontal lobe and insular, forming a narrow band, which are responsible for cognitive impairment as suggested by previous finding. By using these features, the accuracy ranged from 86.0% to 95.5% in binary classification between any pair of cohorts, higher than 70.1%–91.0% when using all network features. In multi-group classification, the average accuracy was 75% or 78% for HC, EMCI, LMCI or EMCI, LMCI, AD against baseline of 33%, and 53.3% for HC, EMCI, LMCI and AD against baseline of 25%. In addition, the recognition rate was lower when combining EMCI and LMCI patients into one group of mild cognitive impairment (MCI) for classification, suggesting that there exists a big difference between early and late MCI patients. This finding supports the EMCI/LMCI inclusion criteria introduced by ADNI based on neuropsychological assessments.

1. Introduction

Worldwide, around 50 million people have Alzheimer's disease (AD) or a related dementia, and there are nearly 10 million new cases every year. These diseases are expected to triple by 2050. AD is a chronic neurodegenerative dysfunction that may contribute to nearly 60%–70% of dementia [1], but only 25% people with AD have been diagnosed. The progression from healthy state to AD spans over many years. These changes can be measured using medical imaging [2] and other techniques. The clinical and pathological characteristics of AD have attracted much attention in recent years. Researchers are working on ways to combine methods from various fields for prodromal diagnosis, prevention and treatment. Methods include neuropsychological assessment, biochemistry pathology, genetic analysis, advanced medical imaging, etc. Arevalo-Rodriguez et al. used the mini–mental state examination (MMSE) to estimate the severity and progression of cognitive impairment, and tried to detect AD in people with MCI [3]. Nineteen areas were found in genes that appear to affect this disease by genome-wide association studies [4]. Another research proved that tau protein abnormalities initiate the disease cascade [5]. To establish indicators of AD during the preclinical stage for early diagnosis and intervention is crucial. The main focus of current studies is to predict the conversion of MCI to AD. Despite that many advances have been made in developing biomarkers for AD using neuroimaging approaches [6], researchers suggest that no single biomarker can accurately diagnose AD by itself alone. Therefore, people started to look at the combination of biomarkers. For instance,

* Corresponding author.
E-mail address: jinhua_sheng@yahoo.com (J. Sheng).

https://doi.org/10.1016/j.heliyon.2021.e07287
Received 25 September 2020; Received in revised form 18 February 2021; Accepted 7 June 2021

biomarkers assessed through imaging and cerebrospinal fluid yields better diagnostic accuracy [7].

Recently, with the advancement of magnetic resonance imaging (MRI), high temporal and spatial resolution facilitates the exploration of the most sophisticated organ of human being: the brain. Researchers believe that structural or functional alterations in the brain bring irreversible cognitive disorders [8]. Various neuroimaging technologies like functional MRI (fMRI) are becoming widely adopted. A key development of fMRI is functional connectivity analysis, which explores the brain from a network perspective. Functional connectivity has become an important tool in studying AD. Researchers compared the connectivity patterns of the default mode network (DMN) by using resting state fMRI with 12 amnestic mild cognitive impairments, 13 AD patients and 13 healthy controls, and found that AD was associated with opposing connectivity effects in the DMN (decreased) and frontal networks (enhanced) [9]. Wang et al. found increased functional connectivity between the left hippocampus and the right lateral prefrontal cortex and diminished rightward asymmetry of hippocampal connectivity in AD [10]. Graph theoretical approaches were combined with automated anatomical labeling (AAL) atlas on fMRI data to study functional brain network alteration in patients with AD [11]. However, there is no consensus on which part of brain plays a key role in the progressive degeneration of Alzheimer's disease. It is more likely to be attributed to the variety of brain parcellation such as Brodmann, Talairach, MNI structural atlas, or probabilistic cerebellar atlas [12] etc. Since 2016, a multi-modal parcellation (HCP MMP) was delineated by Glasser et al. based on human brain cortical architecture, function, connectivity, and topography characteristics [13]. 180 areas per hemisphere were proposed and a machine-learning classifier was designed to recognize the multi-modal 'fingerprint' of each cortical area. It would be of great help to study the brain in health and even disease by using the fine-grained parcellation.

Graph theory-based analysis of complex brain networks applied to AD/MCI classification using AI technologies have attracted wide attention since last decade. In 2010, Rubinov et al. introduced an effective approach for the study of structural and functional connectivity in the human brain. It originates from the topological analysis method for social networks. A complex network in brain is defined by a collection of nodes (regions of interest) and links (correlation coefficient or any other information between nodes) between pairs of nodes. The definition of the nodes is through parcellation of the brain. Four types of networks can be generated based on the weight and directionality within edges. The functional integration and segregation can be characterized using the Brain Connectivity Toolbox (BCT) [14]. Meanwhile, with the development of artificial intelligence, the network metrics can be trained in learning models and have achieved high classification accuracy for AD. Frank de Vos *et al.* computed eight functional connectivity measures as predictors in an elastic net logistic regression, and obtained the area under the receiver operating characteristic curve (AUC) of 0.85 in a classification of 77 AD patients and 173 controls [15]. Andrés Ortiz et al. trained an unsupervised deep learning architecture for the early diagnosis of the AD and provided an accuracy value up to 0.90 [16]. They reported that 44 features were found to achieve 88.4% accuracy based on a parcellation of 264 putative areas as well as the AAL template. Our previous study improved the performance of binary recognition between any pair of cohorts for healthy controls (HC), early MCI (EMCI), late MCI (LMCI) and AD, by computing 3,240 network measures as features in machine learning [17]. Although recent studies have achieved pretty good capability in binary-group recognition, further research is needed from two aspects. One is the interpretation for the selected features in machine learning or deep neural network (DNN); the other is the capacity in multi-group classification. Facing hundreds of thousands of candidate features, it is not easy to explain why those selected areas are suitable to distinguish patients from healthy controls despite the excellent behavior of classification. Especially for DNN, inexplicable number of medial nodes and layers make it even hard to reproduce results. In fact, this might be the reason for the failure of multi-group classification.

In feature selection, the algorithm always randomly searches for an appropriate dimension by which classification error rate declines. It is inexplicable why these selected network metrics are suitable for classification. Therefore, analysis of functional connectivity characteristics before numerical feature selection is necessary, which helps to interpret the underlying mechanism of these features in separating patients and healthy controls.

In this study, we aimed at optimizing feature selection to improve classification accuracy of various stages in AD progression. We started with the non-HCP preprocessing method to prepare and parcellate structural and functional brain MRI data into 360 areas as the network nodes. An ordered core feature set was found following analysis of variability to reduce the dimension of training data. The criterion of ordered core feature is that significant alterations with specifically progressive relationship (AD > LMCI > EMCI > HC or in the reverse order) must exist in certain regions rather than the whole brain in EMCI, LMCI, AD compared with HCs. Finally, machine learning was used as the validation method to test the effectiveness of selected features.

2. Material and methods

2.1. Preparation and parcellation

The overall analysis pipeline proposed in this study is depicted in Figure 1. Subjects acquired from ADNI2 were firstly preprocessed by preparation and parcellation, in which the raw DICOM data were converted into CIFTI space and registered into the HCP MMP atlas with 360 areas. A 360 × 360 dense matrix was generated by computing the correlation coefficients between any two brain areas. Dynamic proportion of the strong weights (dPSW) value varying with individual was used to remove spurious connections and produce both weighted and binary adjacent networks. The optimal dPSW is computed for each subject in groups.

T1-weighted MPRAGE and resting-state fMRI data were processed firstly by J-HCPMMP [17] method. In this paper, a HCP-based multi--modal parcellation of human cerebral cortex (i.e. HCP MMP) was used to parcellate human brain into 180 areas per hemisphere. Although HCP MMP explicitly requires MRI data to follow HCP sampling protocols, J-HCPMMP helps to convert non-HCP style ADNI2 data into HCP CIFTI (Connectivity Informatics Technology Initiative) space using only T1W and fMRI data, without T2W data or field map information. Several brain data processing toolkits were used in J-HCPMMP including FreeSurfer, fMRIPrep, CIFTIFY and HCP minimal preprocessing pipeline. It is basically a fully automatic process with default parameters. First, some necessary steps like motion correction, normalization, segmentation and smoothing etc. are carried out in FreeSurfer for the structural MRI data by the help of recon-all command. Second, fMRIPrep pipeline is used for the registration of functional MRI data. The execution process is also automatic with default set except for the surface preprocessing option. The cerebral cortex is registered into the CIFTI gray-ordinates space which consists of 32,492 vertices per hemisphere for cortical surface and 26,298 NIFTI voxels for subcortical volume. 360 HCP MMP areas in cortical surface are delineated based on these 32,492 × 2 = 64,984 vertices. Thus, we set the surface preprocessing parameter with –cifti-output flag to preprocess BOLD as CIFTI dense timeseries. Finally, we use CIFTIFY tools to convert the FreeSurfer output directory into the CIFTI space and project one subject's NIFTI functional scan to the CIFTI dense series file (*. dtseries). Two commands are executed successively: ciftify_recon_all and ciftify_subject_fmri with default parameters. All the fMRI data within these 360 areas were then used for constructing the brain connectivity network.

2.2. Connectivity network construction

To study the functional association between brain regions, Pearson correlation coefficients were calculated using fMRI signals in the 360

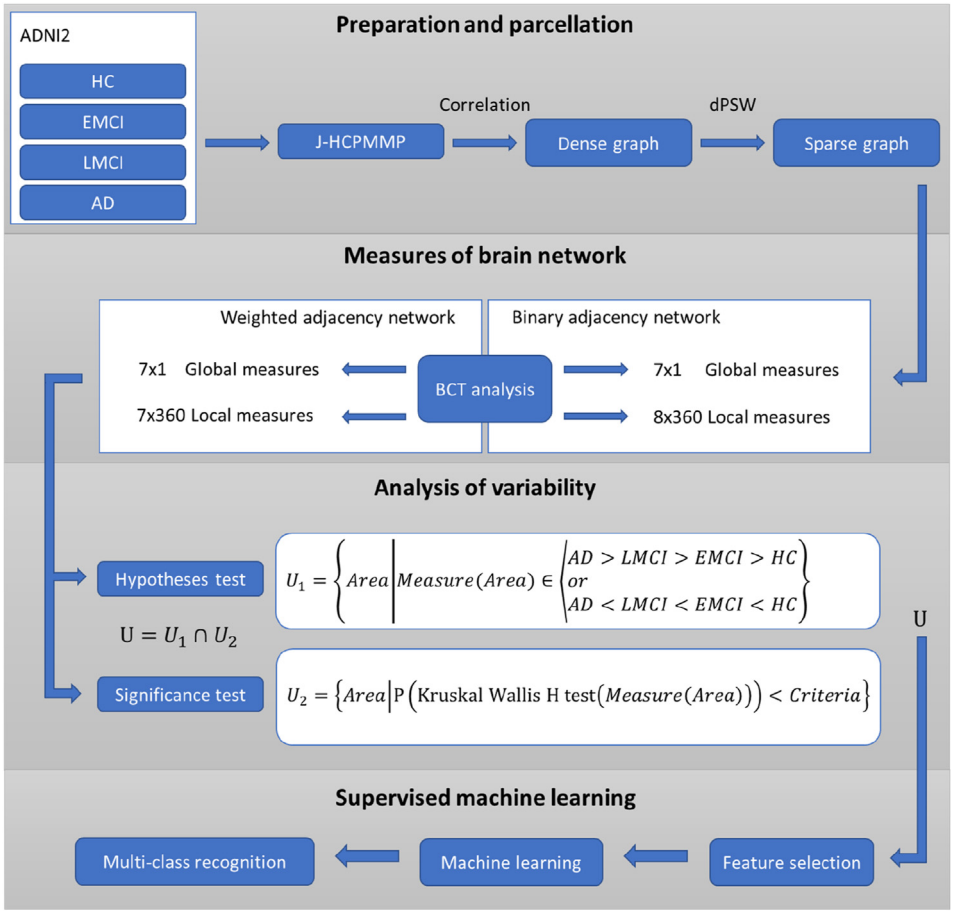


Figure 1. Analysis pipeline proposed in this study.

parcellated areas. Brain connectivity network was constructed by defining each area as a graph node and correlation coefficient as the weight of the graph edge between nodes. Thus, for each subject, functional connectivity of human brain was represented by a 360 × 360 correlation adjacency matrix, and all diagonal weights (representing self-connections) were set to zero. Initially, the matrix would be dense because of noisy correlations and spurious connections due to the scanning environment. An appropriate thresholding is necessary prior to further network anal, 19, ysis. Various methods for de-noising were described in literatures [18, 19, 20], in this study, the well-known Brain Connectivity Toolbox (https://sites.google.com/site/bctnet/) was adopted to generate sparse network by preserving a dynamic proportion of the strongest weights (dynamic PSW, dPSW). It was defined as the number of retained strong weights divided by the total number of weights, given by Eq. (1). In ref [21], the optimal threshold value was determined by graph theory-based measures global cost efficiency (GCE) and global efficiency (E), described in Eq. (2) and Eq. (3):

$$\mathrm{dPSW} = \frac{\text{the number of retained strong weights}}{\text{the total number of weights}} \tag{1}$$

$$E = \frac{1}{n}\sum_{i\in N} E_i = \frac{1}{n}\sum_{i\in N} \frac{\sum_{j\in N, j\neq i} d_{ij}}{n-1} \tag{2}$$

$$\max_{\mathrm{dPSW}}(GCE) = \mathrm{E} - \mathrm{dPSW} \tag{3}$$

where E_i is the efficiency of node i, d_{ij} is the shortest path length between node i and j, n is the total number of nodes in the correlation network, and N is the vector of all nodes. A range of candidate PSWs from 0.01 to 1 with a step 0.05 was tested for maximizing the GCE value. For each subject, the optimal PSW was different.

We calculated dPSW varying from person-to-person instead of a group average value. The optimal dPSWs maintain at a stable level with the normal aging (HC group), while for other groups the linear fitting curves indicate a decline of PSW, meaning that fewer functional connectivity is effective enough to keep brain working well with the increasing of patients' age. Larger dPSWs are gotten which compensate the process of binarization that discards correlation coefficients from the weighted network. Weighting information in brain network is often used to quantitatively analyze the connectivity among brain regions, while the overall communication architecture can be characterized by a binary signal between regions: topological connected or not. So, the binary brain connectivity network was also constructed in this study. After the optimal PSW was determined, the remaining strongest connections in the network would be set to 1 and others would be 0.

2.3. Measures of brain networks

Previous studies reported that graph-based measures of brain network could be effective to reveal characteristics of topological or functional connectivity in human brain. We computed various global and local measures for both weighted and binary networks to quantitatively analyze the significant differences among subjects, and further used them as machine learning features to classify patients and HCs.

Global measures, including global efficiency (E), maximized modularity, assortativity coefficient, optimal number of modules, small worldness index (SWI), characteristic path length (CPL) and mean clustering coefficient (MCC), were thought to be related to the overall per-

formance of the human brain. They have only one value (1 × 1) for each subject. All these measures could be computed directly through BCT software except for SWI, which was given by Eq. (4):

$$\mathrm{SWI} = \frac{MCC/MCC_{rand}}{CPL/CPL_{rand}} \tag{4}$$

where the subscript *rand* represented the average value came from the randomized networks.

In this study, we set the number of randomized networks to be 100. Different local measures in weighted and binary networks were calculated, which characterized regional behavior in the brain. Each local measure can be represented by a vector of 360 values (1 × 360). Seven local measures including strength (S), clustering coefficient (CC), local efficiency (LE), betweenness centrality (BC), eigenvector centrality (EC), page rank centrality (PC) and degree (D) were computed in weighted network. Strength is defined as the sum of neighboring edge weights of a node, and degree is the number of edges connected to a node. In addition to these local measures calculated in weighted network, other measures like k-coreness centrality (KC) and flow coefficient (FC) were computed for binary network.

2.4. Analysis of variability

For each subject, a vector of 5,414 measures (including 7 × 1 global and 7 × 360 local measures in weighted network, 7 × 1 global and 8 × 360 local measures in binary network) was set up. All the measures were standardized to [-1, 1] prior to subsequent analysis. To examine the inter-group variability among EMCI, LMCI, AD and HC, we averaged all the global and local measures within each group. Thus, a 4 × 5,414 matrix representing the average of the four groups was formed. Ordered core features (OCF) was proposed here to distinguish different stages toward Alzheimer's disease. On the basis of previous studies [22, 23, 24] that regional abnormalities in brain causes cognitive impairment, we tested the hypothesis that significant alterations with specifically progressive relationship (AD > LMCI > EMCI > HC or in the reverse order) must exist in certain regions rather than the whole brain. Due to the unknown of data distribution, non-parametric null hypothesis Kruskal-Wallis H test followed by calculating a Bonferroni-corrected P-value for the pairwise comparison was carried out to determine whether these results were statistically significance among the four groups. IBM SPSS software was used for statistics analysis in this study with the varying confidence levels of 90% and 95%. Eqs. (5), (6), and (7) depicted the analysis above.

$$U_1 = \left\{ Area \middle| Measure(Area) \in \left\langle \begin{matrix} AD > LMCI > EMCI > HC \\ or \\ AD < LMCI < EMCI < HC \end{matrix} \right\rangle \right\} \tag{5}$$

$$U_2 = \{Area | P(Kruskal\ Wallis\ H\ test(Measure(Area))) < Criteria\} \tag{6}$$

$$U = U_1 \cap U_2 \tag{7}$$

where *Area* is each of the 360 regions from HCP MMP. U_1 is the set of specific areas containing progressive relationship among groups, U_2 is the set of areas showing two levels of significant measure results, in which 0.05 and 0.1 are considered as the criteria. U is the intersection of U_1 and U_2.

To compare the variability of different HCP MMP areas measured by metrics from BCT, coefficient of variation (CV), which was also known as the relative standard deviation, was computed within each group. CV was defined as the ratio of standard deviation to mean value.

2.5. Supervised machine learning

In order to verify the effectiveness of set U drawn above, network-based brain characteristic measures in each area of that were preserved as the keep-in features in machine learning to distinguish subjects in different groups. In the other areas, network measures were passed through feature selection progress. Many algorithms [25, 26, 27] could be used for reducing dimensions of training data before classification. In this study, we carried out two steps to select the optimal features: filter and wrapper feature selection. In filter-based selection, ReliefF algorithm provided in MATLAB was applied to rank and select top scoring features. It was classifier independent, which considered only interactions among individual values rather than the final effectiveness of machine learning. Strongly correlated features would be considered redundant and assigned lower scores. Then top 1/5 of the ranked features with the highest discrimination ability were treated as the candidate features for the following wrapper-based selection. As a supervised strategy, wrapper algorithm is classification dependent. The selected features vary with machine learning methods. There mainly exist two methods in wrapper selection: forward and backward sequential feature selection (FSFS and BSFS). In FSFS, an empty feature set is initially created. Different features are tested successively in training model. Features that help improving classification accuracy would be added to the set. By contrast, in BSFS, a set of the whole candidate features selected by Filter algorithm is initially set up. Features that have no effect on recognition for groups would be eliminated. We adopted both methods. Measures were tested randomly by FSFS or BSFS strategy with the fixed core features, thus the chosen feature vector varied each time. We evaluated the average performance in recognition with these selected features.

Various classifiers, including decision tree, K-nearest neighbor (KNN), support vector machine (SVM) and ensemble method, were examined in this study. In the choice of parameters in SVM, first, we used the Radial Basis Function (RBF) as kernel in libSVM library (-t = 2) for the non-linear classification. The type of SVM was set to be C-SVC (-s = 0 in libSVM). Then, the optimal combination of regularization parameter C and Gaussian width could be searched by that gave the highest cross validation accuracy. We used the easy/grid.py tool in libSVM to achieve this. Balanced data were used in training, testing and validation to avoid biased results. MATLAB Classification Toolbox was used to achieve multi-class classification. Previously binary classification for EMCI, LMCI, AD and HC were implemented, while it was complicated to be directly applied in multi-class situation.

There are two ways to enhance the performance of a classifier: one-vs-one [28] and one-vs-all [29]. In one-vs-all method, four binary classification models were trained first, namely EMCI vs. others, LMCI vs. others, AD vs. others and HC vs. others. Each subject was scored by these four models to determine the probability it belonged to the class, and it would be judged based on the highest score among the four results. In one-vs-one, there are $\frac{K \times (K-1)}{2}$ binary classifiers for any two classes, in this study K = 4, resulting in 6 classifiers: HC vs. EMCI, HC vs. LMCI, HC vs. AD, EMCI vs. LMCI, EMCI vs. AD, and LMCI vs. AD. Each classifier had a decision about the subject. Finally, subject was classified with the most votes.

To evaluate the performance of classifiers, four indicators were computed in this study including true positive (TP), true negative (TN), false negative (FN) and false positive (FP). N-fold cross validation was employed for robust classification. The number N in cross validation is a crucial tuning parameter that impacts model performance. In common practice, it should be a large number. Considering the number of AD/MCI patients available, we limited N to be 5 in this study. A larger one will lead to few samples in each "fold", which likely causes underfitting. Thus, the whole training data were equally divided into five subsets. We chose four subsets to train classifier and the remaining subset as testing data. The process was repeated for five times, with different testing set at each time. The finally accuracy was the average value of these five recognitions results.

Four kinds of classification experiments were designed for EMCI, LMCI, AD and HC: binary recognition, three-class recognition, four-class recognition, and a special grouping in which two strategies of sampling

were implemented: one randomly selecting 30 patients in the mix of EMCI and LMCI, and the other randomly selecting 15 patients in group of EMCI and LMCI respectively for the recognition of HC, MCI and AD with balanced training data.

3. Results

Structural and functional MRI data from ADNI database were analyzed in this study with 30 AD patients, 29 LMCI patients, 40 EMCI patients and 33 HCs. To ensure the uniformity of data acquisition protocols and formats, all images were downloaded from ADNI2 (Philips Medical Systems, MRI: matrix = 256 × 256, slice thickness = 1.2mm, TE/TR = 3.2/6.8 ms; fMRI: matrix = 64 × 64, slice thickness = 3.3mm, slices = 6,720, TE/TR = 30/3000 ms). There is no significant difference in the average age between the four cohorts. The average age was 75.6 ± 5.8 for HC subjects (14 male, 19 female), 71.8 ± 6.5 for EMCI subjects (14 male, 26 female); 72.9 ± 7.8 for LMCI subjects (16 male, 13 female), and 73.1 ± 6.8 for AD patients (12 male, 18 female).

Figure 2 shows the seven global network measures in weighted and binary network. For the purpose of comparison, each measure is normalized to [0, 1]. Groups are distinguished in colors. The local quantities of the network were calculated for nodal degree, nodal strength, betweenness centrality, local efficient, clustering coefficient, eigenvector centrality and page rank centrality. The areas showing significant alterations with specifically progressive relationship (AD > LMCI > EMCI > HC or in the reverse order) in weighted and binary network are selected in Figure 3, respectively, and all of them make up the set U_1 introduced in Sec. Analysis of variability. The areas presented the characteristics in weighted (Figure 3L: (A) Betweenness Centrality, (B) Strength, (C) Clustering Coefficient, (D) Local Efficiency, (E) Eigenvector Centrality, (F) Pagerank Centrality, (G) Degree) and binary (Figure 3R: (A) Strength, (B) Clustering Coefficient, (C) Local Efficiency, (D) Betweenness Centrality, (E) Eigenvector Centrality, (F) Pagerank Centrality, (G) K-coreness Centrality, (H) Flow Coefficient) network are selected, respectively, and all of them make up the set U_1 introduced in Sec. Analysis of variability.

Figure 4 shows the areas with at least four network measures satisfying the OCF hypothesis. Among them, R-44 area has 8 measures with AD > LMCI > EMCI > HC and R-PBelt has 7 measures with AD<LMCI<EMCI<HC. Hence, all the areas in U_1 must be validated by significance test. Significance test results are calculated for weighted and binary network. Only tuples (Area, BCT) with significance ($P < 0.05$) are shown (in the following analyses, we also retained the results with $P < 0.1$, which are not presented here). Colors represent different levels of significance. Areas in these significant points make up the set U_2.

The final set $U = U_1 \cap U_2$ is shown in Table 1. For the relationship HC < EMCI < LMCI < AD, Table 1 lists the area with P-value <0.05 (underline) and 0.1 respectively. The highest CV values were found in set U, including the clustering coefficient in R-AAIC area and the eigenvector centrality in R-7AM area. By using the areas with network measure from Table 1 as the core features that were fed into filter and wrapper feature selection process, HC, EMCI, LMCI, and AD were distinguished by machine learning for binary classification and multi-class recognition. We found that SVM always achieved the highest accuracy against decision tree, KNN, and ensemble method we tested in this study. Five-fold cross validation was carried out. During each of the five repetitions, only 4/5 samples were used in training and 1/5 samples were under tested. Thus, the number of samples actually involved in modeling was less than the original weighted data set. The receiver operating characteristic curves (ROC) of both binary recognition and multi-class recognition are shown in Figure 5. From Figure 5L (A)-(F) are binary classification results for HC vs. EMCI, HC vs. LMCI, HC vs. AD, EMCI vs. LMCI, EMCI vs. AD and LMCI vs. AD respectively. Figure 5R (A-B) are three-class recognitions, one for HC vs. EMCI vs. LMCI and another for EMCI vs. LMCI vs. AD. Figure 5R (C) is the full class recognition in which subjects of HC, EMCI, LMCI and AD are mixed together. Figure 5R (D) shows the results of all groups classification too, while EMCI and LMCI are labeled as one class: MCI. Consequently, it is still three-class recognition. The TPR/FPR are the averaged classification results from five-fold cross validation.

The accuracy was computed by averaging the TP rate (vertical axis value) of the optimal cut-point pointed by arrow. In binary classification, the average accuracies are as following: (HC vs. EMCI) = 90.5%, (HC vs. LMCI) = 92%, (HC vs. AD) = 95.5%, (EMCI vs. LMCI) = 86.0%, (LMCI vs. AD) = 87.0%, and (EMCI vs. AD) = 88.5%. All the areas under ROC are far larger than 0.50, which represents the superior behavior of the classifier. In three-class classification, the average accuracies are 77.7% for HC vs. EMCI vs. LMCI and 75.0% for EMCI vs. LMCI vs. AD. The average accuracy of four-class recognition is 53.3% when HC, EMCI, LMCI and AD are mixed together. The average accuracy is 58.7% when EMCI and LMCI are labeled as one class: MCI is essentially a three-class recognition problem.

Figure 6 illustrates the distribution of the areas in set U, in which network measures showed significant alterations with progressive relationship. The top row shows the left hemisphere, and the bottom row shows for the right. A total of 30 network measures exhibit the progressive relationship in 21 HCP MMP areas in set U. We combine all the chosen 21 areas together and draw them in each hemisphere, as shown in Figure 7, a distinct banding area appears (only left hemisphere showed, contoured with yellow lines). Most regions locate in the frontal lobe (6v, 6d, 55b, IFJp and area 44) and part of insular (FOP5, AAIC, AVI, MI).

4. Discussion

We proposed a hypothesis that progressive relationship AD > LMCI > EMCI > HC or reverse order exists for some network measures in certain functional regions rather than the whole brain. In contrast to previous

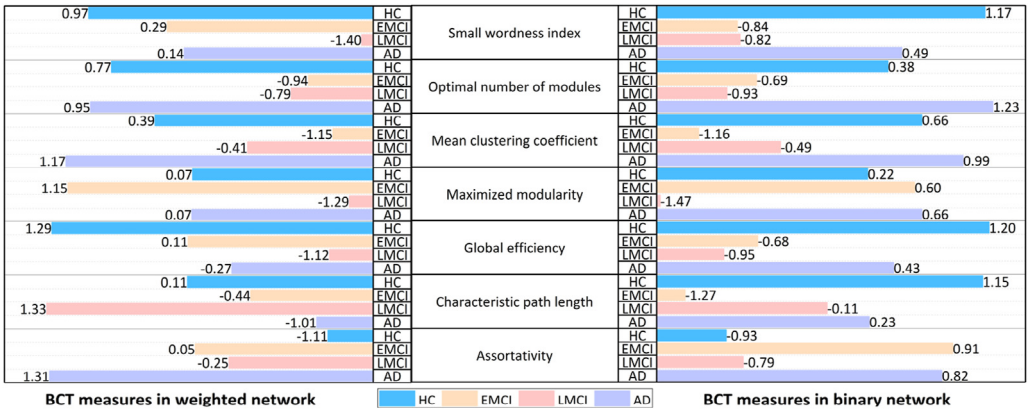


Figure 2. Normalized global network measures.

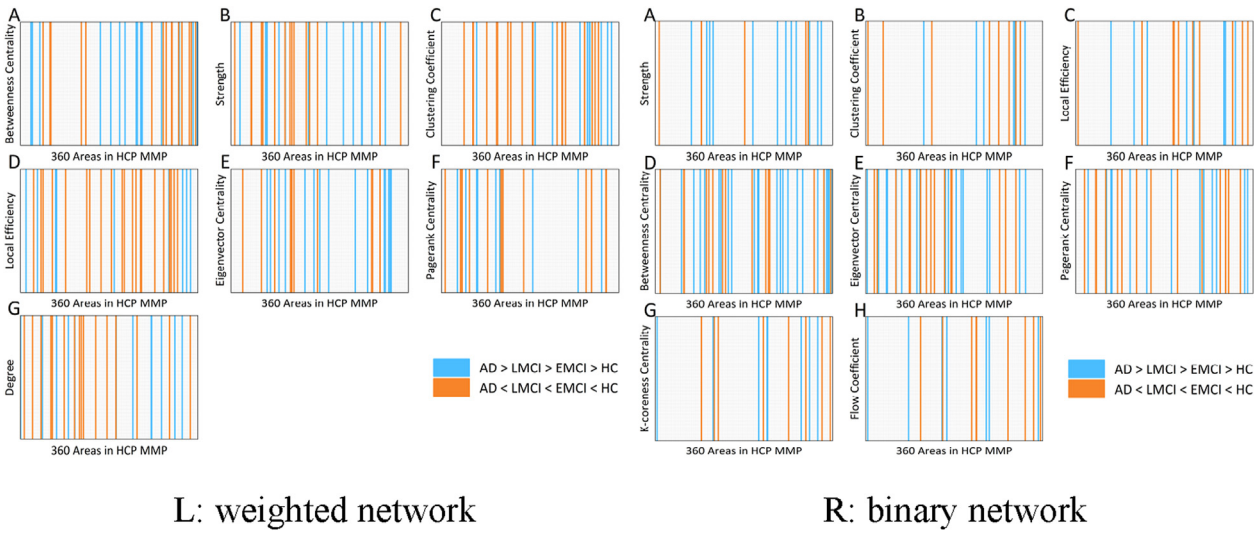


Figure 3. L: Distribution of areas satisfying OCF hypothesis in weighted network as horizontal axis. (A) Betweenness Centrality, (B) Strength, (C) Clustering Coefficient, (D) Local Efficiency, (E) Eigenvector Centrality, (F) Pagerank Centrality, (G) Degree; R: Distribution of areas those satisfied OCF hypothesis in binary network. (A) Strength, (B) Clustering Coefficient, (C) Local Efficiency, (D) Betweenness Centrality, (E) Eigenvector Centrality, (F) Pagerank Centrality, (G) K-coreness Centrality, (H) Flow Coefficient. Vertical axis indicated BCT local measure and horizontal axis was the HCP MMP areas.

studies that all of the functional connectivity based network measures were used directly in training model as candidate features, we exploited the variability of fine-grained human brain parcellation among subjects in HC, EMCI, LMCI and AD. Network measures in these special areas with significant alterations were treated as the keep-in features and classified with training model to validate their effectiveness in distinguishing different levels of cognitive impairment or even Alzheimer's disease. Combining with statistical method and machine learning, a total of 30 network measures distributing in 21 HCP MMP areas showed significant alterations and the accuracy of binary classification exceeds 90%. Most of the areas locate in the frontal lobe and insular of the brain.

It is impossible to discern any distinguishable patterns from 5,414 global and local measures directly, and it is also a very challenging and exhausting work to do feature selection process among such high data dimensions in machine learning. Under the hypothesis proposed in this study, we reduced the number of dimensions to 359 of which 180 network measures present the progressive relationship of AD $>$ LMCI $>$ EMCI $>$ HC and 179 in the reverse order. Tested by different levels of statistical analysis (P-value $<$ 0.1 and 0.05), 30 core network measures are finally chosen which shows significance alterations. These 30 measures mainly locate in the areas of Premotor Cortex (Inferior Premotor Cortex L-6v, Superior Premotor Cortex R-6d and R-55b that sits between them), Inferior Frontal Cortex (L-p47r, L-IFJp and R-44), Insular and Frontal Opercular Cortex (Frontal Opercular areas L-FOP5 and R-FOP5, Middle Insular area L-MI, Anterior Ventral Insular area L-AVI and Anterior Agranular Insular Complex R-AAIC), Early Auditory Cortex (R-PBelt), Superior Parietal Cortex (L-7Pm and R-7Am), Anterior Cingulate and Medial Prefrontal Cortex (L-33pr), Medial Temporal Cortex (perihippocampal areas 1, L-PHA1), Orbital and Polar Frontal Cortex (L-10pp), Ventral Stream Visual Cortex (L-VVC), Dorsal Stream Visual Cortex (L-V3B), and its superior neighbor Inferior Parietal Cortex (L-IP0 and L-IP1), as shown in Figure 6. Symmetry in hemispheres can be found that both left and right FOP5 areas are significantly altered with the status HC $<$ EMCI $<$ LMCI $<$ AD, and adjacent area showing the same pattern can be found in the contralateral hemisphere with most regions, like L-6v and R-6d, L-7Pm and R-7Am, L-p47r and R-44. It is in consensus with previous report that symmetrically distributed pathological changes are accompanied with brain degeneration [30].

Most brain regions shown in Figure 7 locate in the frontal lobe and part of insular. The frontal lobe contains most of the dopamine neurons in the cerebral cortex, the dopaminergic pathways are associated with attention, short-term memory tasks, planning and motivation. There is considerable evidence to suggest that the frontal lobes undergo structural, functional, and pathological changes that ultimately have negative effects on cognitive functioning [31]. Sacuiu et al. found that chronic depressive symptomatology in mild cognitive impairment was associated with frontal atrophy rate that hastens conversion to Alzheimer dementia [32]. Jeremy Koppel demonstrated that psychosis in Alzheimer's disease was associated with frontal metabolic impairment and accelerated decline in working memory [33]. Olazarán found that depressive symptoms were related to atrophy in the left precentral gyrus (Brodmann area 6, BA6) [34]. Brodmann area 6 participates in the capacity of

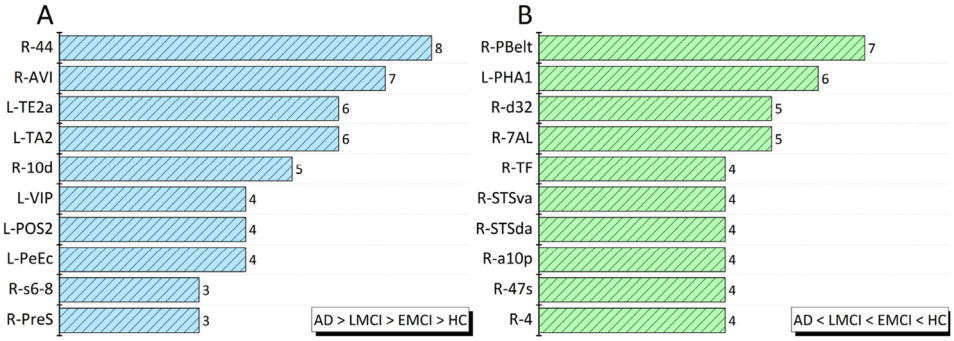


Figure 4. Top half of the histogram of areas that satisfy the hypothesis. A) Areas present progressive relationship of AD $>$ LMCI $>$ EMCI $>$ HC. B) The reverse order.

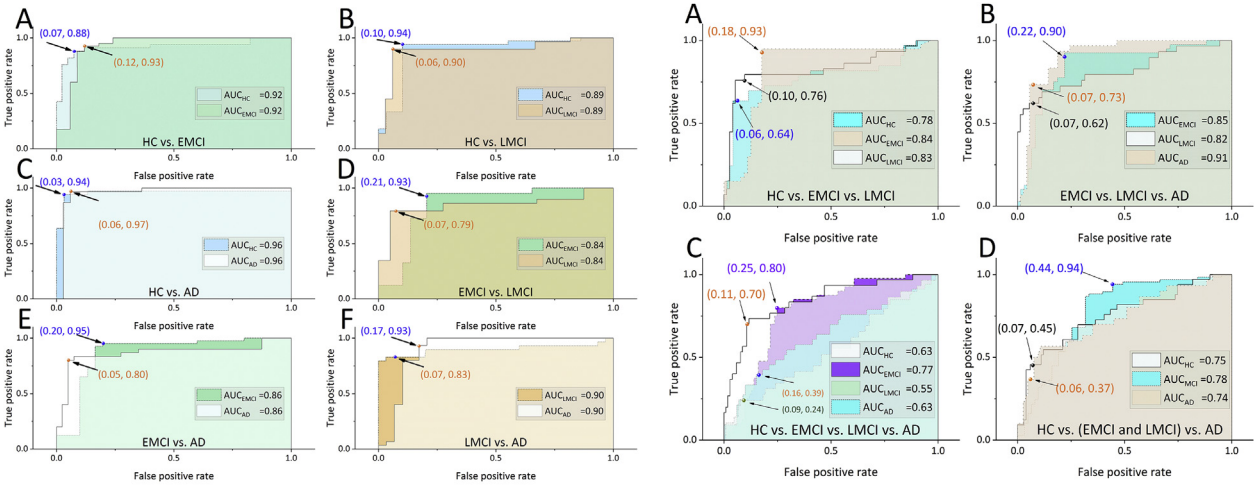
Table 1. HC<EMCI<LMCI<AD or HC>EMCI>LMCI>AD with P-value<0.1 or 0.05 (only underline).

Area	BCT	Order	Network	P-value
R-44	S	Positive	Weighted	0.031
	LE	Positive		0.011
	EC	Positive		0.008
	D	Positive		0.025
L-IP0	EC	Positive		0.091
R-55b	LE	Positive		0.043
R-FOP5	EC	Positive		0.072
L-IP1	CC	Positive		0.060
L-V3B	EC	Positive		0.035
L-p47r	BC	Positive		0.052
L-6V	BC	Positive	Weighted	0.039
R-44	S	Positive	Binary	0.006
	EC	Positive		0.041
L-V3B	S	Positive		0.029
L-FOP5	BC	Positive		0.044
L-10pp	BC	Positive		0.079
L-MI	BC	Positive		0.092
L-IP1	CC	Positive		0.093
R-7AM	EC	Positive		0.082
R-6d	BC	Positive		0.072
R-FOP5	EC	Positive		0.095
L-7pm	BC	Positive		0.060
L-IFJp	BC	Positive	Binary	0.085
R-AAIC	CC	reverse	Weighted	0.024
L-AVI	LE	reverse		0.063
L-PHA1	LE	reverse		0.080
L-VVC	BC	reverse	Weighted	0.099
R-PBelt	PC	reverse	Binary	0.059
L-33pr	LE	reverse		0.074
L-IP1	FC	reverse		0.093

*Strength (S), Local Efficiency (LE), Eigenvector Centrality (EC), Degree (D), Clustering Coefficient (CC), Betweenness Centrality (BC), Page-rank Centrality (PC), Flow Coefficient (FC).

temporarily holding information and having that information rapidly accessible [35]. In our study, a large part of BA6, including the superior premotor 6d, area 55b and the inferior premotor 6v, are found significant changes. This is not only in agreement with the previous results, but also further reveals the affected cortical subdivisions. Other researchers mentioned that the regional atrophy of the insular cortex was associated with neuropsychiatric symptoms in AD patients [36], and some of the behavioral abnormalities in AD might reflect insular pathology [37]. We elaborate the regions that alter the most in network-based measures in the framework of HCP multi-modal parcellation. As far as we know, it is the first time to take advantage of such a fine-grained brain parcellation to analyze Alzheimer's disease.

Binary and multi-group classification results show the advantage of using the 21 chosen brain areas for recognition in machine learning. Area under the ROC is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one [38]. For the binary classification, the highest accuracy rate (95.5%) was found in the task of HC vs. AD, and the lowest value (86%) was in EMCI vs. LMCI. It can be interpreted that the smaller the difference between groups, the harder to separate them with machine learning. Both three-group classification results show an average accuracy rate about 76%. For the most difficult four-group classification, the highest AUC value reaches 0.77 for the recognition of EMCI group, and the lowest AUC value is 0.55 for the group of LMCI. The average AUC is 0.65. If groups of EMCI and LMCI are treated as one group of MCI, the AUC of MCI has the highest value of 0.78. We realize that the AUC fluctuation may be related to the number of subjects acquired (HC = 33, EMCI = 40, LMCI = 29 and AD = 30). As mentioned previously, the imbalance of training data affects classification results [39]. Classification of HC, MCI and AD with balanced training data is carried out (Figure 5R (D)). The accuracies in this special grouping are all floating around 60%. Compared with two-class recognition, it is closer to the result of four-class recognition (53.3%) rather than those in three classes recognition (above 75%), although we manually label the patients into three groups: HC, MCI, and AD. The ADNI database introduced EMCI in 2009 and LMCI in 2011 by the neuropsychological assessment of MMSE or Wechsler Memory Scale (WMS) to measure cognitive impairment. The better recognition accuracy of classifying EMCI and LMCI as two groups rather than one MCI group suggests that there exists still a gap between early and late MCI patients, and it is inappropriate to classify them into one group in general.



L: Binary Recognition R:Multi-class Recognition

Figure 5. L: ROC curves that illustrated the behavior of trained model in binary recognition. (A) HC vs. EMCI, (B) HC vs. LMCI, (C) HC vs. AD, (D) EMCI vs. LMCI, (E) EMCI vs. AD, (F) LMCI vs. AD; R: ROC curves that illustrated the behavior of trained model in multi-class recognition. (A) HC vs. EMCI vs. LMCI, (B) EMCI vs. LMCI vs. AD, (C) HC vs. EMCI vs. LMCI vs. AD, (D) Groups of EMCI and LMCI were treated as one class: MCI.

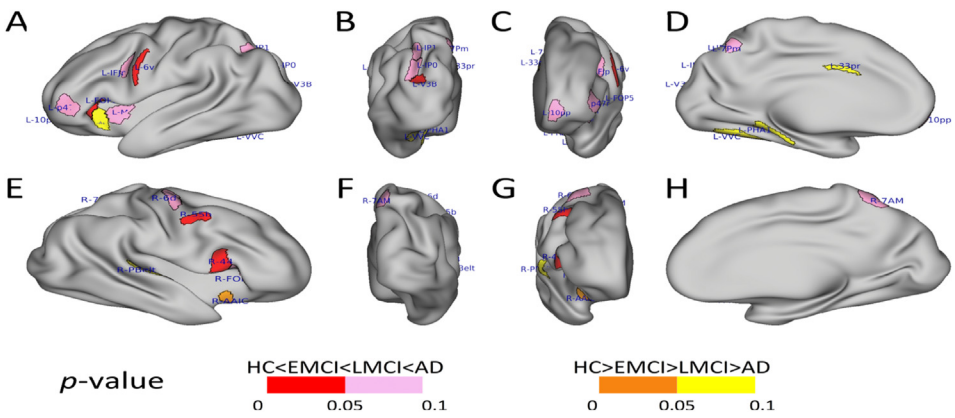


Figure 6. Distribution of areas that showed significant alterations with progressive relationship. All these areas formed the set U, which is described in Section: Analysis of variability above. A)-D) are for the left hemisphere with views from the lateral, posterior, anterior and medial perspectives. E)-H) are for the right hemisphere with views from the lateral, anterior, posterior and medial perspectives. (Drawn in HCP Connectome Workbench v1.3.2, https://www.humanconnectome.org/software/connectome-workbench).

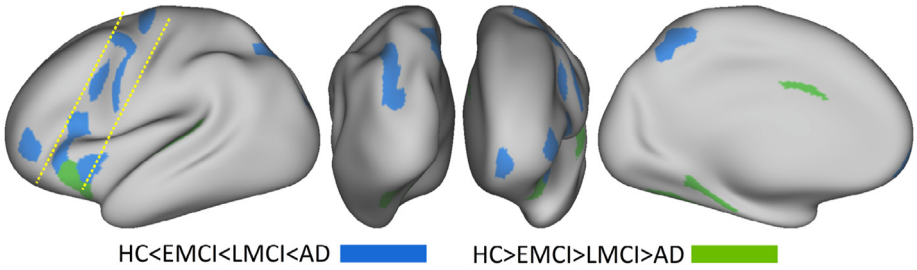


Figure 7. All the chosen 21 areas are putted together in each hemisphere (here shows the left), and a distinct banding area can be seen.

The sample size in machine learning is a crucial factor that impacts the model performance. To ensure the reproducibility of the results, the uniformity of data acquisition protocols was imposed in this study, resulting in fewer samples. It is different from the common machine learning research in which plenty samples are engaged in training and testing. A limited sample size usually leads to model underfitting. Specifically, both the accuracies in training and testing will not perform well enough. This requires additional data preprocessing methods like bootstrapping to enhance the robustness of results. Second, only two kinds of monotonic progressive relationship among brain areas are taken into account. However, the deterioration of Alzheimer's disease is such a complex process, there must be more manifestations of brain functional degeneration besides these two relationships in a longitudinal scale. Hence, other criteria of core feature selection can be introduced to reflect the variation of network measures in the progression of Alzheimer's disease and help to differentiate HC, EMCI, LMCI, and AD.

5. Conclusion

A non-HCP preprocessing method was used to prepare and parcellate structural and functional brain MRI data into 360 areas. Through analysis of variability which combined hypothesis and significance tests, we found a set of areas in which the network measures significantly changed in the order of HC > EMCI > LMCI > AD or in the reverse order. Areas in this set mainly locate in the frontal lobe and insular. It is the first time a distinct banding area to be observed as far as we know. We call these network measures the ordered core features (OCF). By preserving the OCF set as keep-in features in classifiers, superior performance in multi-class recognition was obtained. Our results also support the EMCI/LMCI inclusion criteria introduced by ADNI based on neuropsychological assessments.

Declarations

Author contribution statement

Jinhua Sheng: Conceived and designed the experiments; Wrote the paper.

Bocheng Wang: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Qiao Zhang: Conceived and designed the experiments.

Rougang Zhou, Luyun Wang and Yu Xin: Analyzed and interpreted the data.

Funding statement

This work was supported by the National Natural Science Foundation of China (No. 61871168). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI), United States and U.S. Department of Defense (DOD) ADNI.

Data availability statement

Data included in article/supplementary material/referenced in article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] World Health Organization, Dementia, 2020. https://www.who.int/news-room/fact-sheets/detail/dementia.

[2] S. Rathore, M. Habes, M.A. Iftikhar, A. Shacklett, C. Davatzikos, A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages, NeuroImage 155 (2017) 530–548.

[3] I. Arevalo-Rodriguez, N. Smailagic, M.R. i Figuls, A. Ciapponi, E. Sanchez-Perez, A. Giannakou, O.L. Pedraza, X.B. Cosp, S. Cullum, Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI), Cochrane Database Syst. Rev. (2015).

[4] J.-C. Lambert, C.A. Ibrahim-Verbaas, D. Harold, A.C. Naj, R. Sims, C. Bellenguez, G. Jun, A.L. DeStefano, J.C. Bis, G.W. Beecham, others, Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease, Nat. Genet. 45 (2013) 1452.
[5] G. Šimić, M. Babić Leko, S. Wray, C. Harrington, I. Delalle, N. Jovanov-Milošević, D. Bažadona, L. Buée, R. De Silva, G. Di Giovanni, C. Wischik, P.R. Hof, Tau protein hyperphosphorylation and aggregation in Alzheimer's disease and other tauopathies, and possible neuroprotective strategies, Biomolecules 6 (2016) 6.
[6] F. Márquez, M.A. Yassa, Neuroimaging biomarkers for Alzheimer's disease, Mol. Neurodegener. 14 (2019) 1–14.
[7] R. Khoury, E. Ghossoub, Diagnostic biomarkers of Alzheimer's disease: a state-of-the-art review, Biomark. Neuropsychiatr. 1 (2019) 100005.
[8] S. Sarraf, G. Tofighi, Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data, in: 2016 Future Technol. Conf. FTC, IEEE, 2016, pp. 816–820.
[9] F. Agosta, M. Pievani, C. Geroldi, M. Copetti, G.B. Frisoni, M. Filippi, Resting state fMRI in Alzheimer's disease: beyond the default mode network, Neurobiol. Aging 33 (2012) 1564–1578.
[10] L. Wang, Y. Zang, Y. He, M. Liang, X. Zhang, L. Tian, T. Wu, T. Jiang, K. Li, Changes in hippocampal connectivity in the early stages of Alzheimer's disease: evidence from resting state fMRI, Neuroimage 31 (2006) 496–504.
[11] A. Khazaee, A. Ebrahimzadeh, A. Babajani-Feremi, Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory, Clin. Neurophysiol. 126 (2015) 2132–2141.
[12] J. Diedrichsen, J.H. Balsters, J. Flavell, E. Cussans, N. Ramnani, A probabilistic MR atlas of the human cerebellum, Neuroimage 46 (2009) 39–46.
[13] M.F. Glasser, T.S. Coalson, E.C. Robinson, C.D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C.F. Beckmann, M. Jenkinson, A multi-modal parcellation of human cerebral cortex, Nature 536 (2016) 171–178.
[14] M. Rubinov, O. Sporns, Complex network measures of brain connectivity: uses and interpretations, NeuroImage 52 (2010) 1059–1069.
[15] F. de Vos, M. Koini, T.M. Schouten, S. Seiler, J. van der Grond, A. Lechner, R. Schmidt, M. de Rooij, S.A.R.B. Rombouts, A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease, NeuroImage 167 (2018) 62–72.
[16] A. Ortiz, J. Munilla, J.M. Gorriz, J. Ramirez, Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease, Int. J. Neural Syst. 26 (2016) 1650025.
[17] J. Sheng, B. Wang, Q. Zhang, Q. Liu, Y. Ma, W. Liu, M. Shao, B. Chen, A novel joint HCPMMP method for automatically classifying Alzheimer's and different stage MCI patients, Behav. Brain Res. 365 (2019) 210–221.
[18] N. Langer, A. Pedroni, L. Jäncke, The problem of thresholding in small-world network analysis, PloS One 8 (2013), e53199.
[19] M.P. van den Heuvel, S.C. de Lange, A. Zalesky, C. Seguin, B.T. Yeo, R. Schmidt, Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: issues and recommendations, Neuroimage 152 (2017) 437–449.
[20] B. Jie, D. Zhang, C.-Y. Wee, D. Shen, Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification, Hum. Brain Mapp. 35 (2014) 2876–2897.
[21] A. Khazaee, A. Ebrahimzadeh, A. Babajani-Feremi, Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease, Brain Imaging Behav. 10 (2016) 799–817.
[22] J. Yu, C.L. Lam, T.M. Lee, White matter microstructural abnormalities in amnestic mild cognitive impairment: a meta-analysis of whole-brain and ROI-based studies, Neurosci. Biobehav. Rev. 83 (2017) 405–416.
[23] E.R. Lindemer, D.N. Greve, B. Fischl, J.C. Augustinack, D.H. Salat, Differential regional distribution of juxtacortical white matter signal abnormalities in aging and Alzheimer's disease, J. Alzheimers Dis. 57 (2017) 293–303.
[24] T. Yokoi, H. Watanabe, H. Yamaguchi, E. Bagarinao, M. Masuda, K. Imai, A. Ogura, R. Ohdake, K. Kawabata, K. Hara, others, Involvement of the precuneus/posterior cingulate cortex is significant for the development of Alzheimer's disease: a PET (THK5351, PiB) and resting fMRI study, Front. Aging Neurosci. 10 (2018) 304.
[25] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, T.-F. Yuan, Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning, Front. Comput. Neurosci. 9 (2015) 66.
[26] P. Ghamisi, J.A. Benediktsson, Feature selection based on hybridization of genetic algorithm and particle swarm optimization, IEEE Geosci. Remote Sens. Lett. 12 (2015) 309–313.
[27] J. Pohjalainen, O. Räsänen, S. Kadioglu, Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits, Comput, Speech Lang. 29 (2015) 145–171.
[28] M. Galar, A. Fernández, E. Barrenechea, F. Herrera, DRCW-OVO: distance-based relative competence weighting combination for one-vs-one strategy in multi-class problems, Pattern Recogn. 48 (2015) 28–42.
[29] Y. Liu, J.-W. Bi, Z.-P. Fan, A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm, Inf. Sci. 394 (2017) 38–52.
[30] H.-C. Tai, B.Y. Wang, A. Serrano-Pozo, M.P. Frosch, T.L. Spires-Jones, B.T. Hyman, Frequent and symmetric deposition of misfolded tau oligomers within presynaptic and postsynaptic terminals in Alzheimer's disease, Acta Neuropathol. Commun. 2 (2014) 146.
[31] B.L. Miller, J.L. Cummings, The Human Frontal Lobes: Functions and Disorders, Guilford Publications, 2017.
[32] S. Sacuiu, P.S. Insel, S. Mueller, D. Tosun, N. Mattsson, C.R. Jack Jr., C. DeCarli, R. Petersen, P.S. Aisen, M.W. Weiner, others, Chronic depressive symptomatology in mild cognitive impairment is associated with frontal atrophy rate which hastens conversion to Alzheimer dementia, Am. J. Geriatr. Psychiatr. 24 (2016) 126–135.
[33] J. Koppel, S. Sunday, T.E. Goldberg, P. Davies, E. Christen, B.S. Greenwald, A.D.N. Initiative, Others, psychosis in Alzheimer's disease is associated with frontal metabolic impairment and accelerated decline in working memory: findings from the Alzheimer's disease neuroimaging initiative, Am. J. Geriatr. Psychiatr. 22 (2014) 698–707.
[34] J. Olazarán, P. García-Polo, D. García-Frank, A. Quirós, J. Hernández-Tamames, C. Acedo, J. Álvarez-Linera, A. Frank, Structural correlates of depressive symptoms in prodromal Alzheimer's disease, Br. J. Med. Med. Res. 14 (2016) 1–10.
[35] A. Baddeley, Essentials of Human Memory (Classic Edition), Psychology Press, 2013.
[36] Y. Moon, W.-J. Moon, H. Kim, S.-H. Han, Regional atrophy of the insular cortex is associated with neuropsychiatric symptoms in Alzheimer's disease patients, Eur. Neurol. 71 (2014) 223–229.
[37] D.J. Bonthius, A. Solodkin, G.W. Van Hoesen, Pathology of the insular cortex in Alzheimer disease depends on cortical architecture, J. Neuropathol. Exp. Neurol. 64 (2005) 910–922.
[38] T. Fawcett, An Introduction to ROC Analysis Pattern Recognition Letter, 2006.
[39] A. L'heureux, K. Grolinger, H.F. Elyamany, M.A. Capretz, Machine learning with big data: challenges and approaches, IEEE Access 5 (2017) 7776–7797.