Journal of Neuroscience Methods 361 (2021) 109265

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Classification of patients with AD from healthy controls using entropy-based measures of causality brain networks

Yuanchen Wu [a,1], Yuan Zhou [b,*,1], Miao Song [a]

[a] School of Information Engineering, Shanghai Maritime University, Shanghai, China
[b] School of Logistics Engineering, Shanghai Maritime University, Shanghai, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Granger causality
Sample entropy
Causality pattern

ABSTRACT

Background: Machine learning and pattern recognition have been widely used in rs-fMRI data to investigate Alzheimer's disease (AD). However, many previous methods extracted discriminative features based on functional correlations, which may ignore the asynchronous causality influence of neural activities.
New method: We propose a novel method for AD diagnosis using Sample Entropy to measure the neural complexity of the brain causality network. Granger Causality analysis with a sliding time window was applied on rs-fMRI data of 29 AD patients and 30 cognitive normal (CN) controls to compute the whole brain's causality series. We further grouped these causality series into clusters by agglomerative hierarchical clustering algorithm and computed Sample Entropy of the clusters as the classification features.
Results: We explored four different classifiers, i.e., XGBoost, SVM cluster, Random Forest, and SVM, based on the above features. An accuracy of 89.83%, with a sensitivity of 90.00% and a specificity of 89.66%, was achieved with the optimal feature subsets using the SVM classifier.
Comparison with existing methods: With the same dataset, the performances of the proposed method were generally higher than those of conventional methods for AD classification based on Pearson's correlation network, dynamic Pearson's correlation network, High-order correlation network, and causality correlation network.
Conclusions: Our method demonstrates the measure of Sample Entropy with causality connection as a powerful tool to classify AD patients from CN controls, and provides a deep insight into the neuropathogenesis of AD.

1. Introduction

Alzheimer's Disease (AD) is a severe brain disorder among older adults, resulting in slow and irreversible intellectual deteriorations (Mega et al., 1996). Until now, about 44 million people worldwide suffer from this disease. As it has no current cure, early diagnosis is the only way to curb progression and improve the quality of life for AD patients. Unfortunately, the symptoms of AD vary from person to person, from memory impairments to wording confounding, to vision/spatial perceptual disturbance, to reasoning problems. This volatility and complexity of AD symptoms greatly challenge the routine detection methods based on patients' medical history, genetic history, and clinical observations. Thus, machine learning algorithms have been gradually introduced in medical practice to detect AD in terms of functional magnetic resonance (fMRI), cerebrospinal fluid, and blood indicators.

The resting-state fMRI, which uses blood oxygen level-dependent (BOLD) signal as a neurophysiological indicator, is widely used to study AD. By calculating the correlations of the BOLD signal of different brain regions, the functional connectivity (FC) brain network can be constructed to successfully classify AD/MCI patients (for a review, see (Dennis and Thompson, 2014)). So far, many FC-based methods have been proposed in terms of various network structures and correlation computations. For instance, researchers have introduced some graph theories, such as visibility graph (Gao et al., 2020) and "small world" graph (Sanz-Arigita et al., 2010), to discover topological features of low-level FC network structure. On the other hand, some researchers have explored the correlation computation methods, e.g., sparse representation (Lv et al., 2015), and High-order Pearson's correlation based on dynamic low-level FC (Chen et al., 2016a), to describe the multi-regional and complex interactions among different brain areas.

* Corresponding author.
E-mail address: zyuan@shmtu.edu.cn (Y. Zhou).
[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.jneumeth.2021.109265
Received 24 December 2020; Received in revised form 26 May 2021; Accepted 17 June 2021
Available online 24 June 2021

Most of the above methods achieved good classification results, suggesting that FC is an effective and robust method to characterize disrupted neural activities of AD/MCI. However, it also should be noted that FC, as a non-directional measure, cannot capture the causality influences of different brain areas, which may lead to ambiguous interpretations of information flow for the brain.

To overcome the shortcoming of FC, Granger Causality connectivity, which can determine whether one neural signal series would influence another, has been introduced to examine neural alterations between MCI/AD patients and normal controls (Liu et al., 2012; Liang et al., 2014; Khazaee et al., 2017). For instance, Liu et al. (2012) found that the number of causality connections decreases in AD patients compared to those in the CN controls. In contrast, Liang et al. (2014) observed both increments and decrements of causality connections in different brain regions in MCI patients. Recently, Granger Causality has been further applied to identifying patients with MCI/AD using machine learning methods. Khazaee et al. (2017) are the first to construct directional graphs of causality connectivity for classification between AD patients and CN controls. They calculated topological measures of the directional graph (e.g., degree, betweenness centrality, and flow efficiency) as features, and finally achieved 71.95% classification accuracy on automatic anatomical labeling (AAL) template. Wang et al. (2017a, 2017b). further extended Khazaee's work based on causality network, and they systematically compared the effect of various graphical features and corresponding classification methods on classification results. These studies proved the validity of identifying AD patients by the causality connectivity. However, they only focus on topological and static features in terms of a directional graph without sufficiently characterizing the neural complexity and dynamic reconfiguration in the brain.

Previous studies have applied entropy metrics to quantify the complexity of the BOLD signals (Wang et al., 2017a, 2017b; Niu et al., 2018; Toussaint et al., 2014). Wang et al. (2017a, 2017b) observed the decreasing trend of permutation entropy from CN controls to AD patients, suggesting the lower neural complexity of AD patients than that of normal controls. Consistent with Wang et al.'s research, Niu et al. (2018) performed a multi-scale entropy analysis and found significant differences at multiple time scales in multi-regions (e.g., thalamus, insula, lingual gyrus, and inferior occipital gyrus, middle temporal gyrus) between CN, early MCI, late MCI, and AD. Moreover, Toussaint et al. (2014) combined the entropy measures with a graph-theoretical approach, and reported a decreased antero-posterior integration in AD patients compared to elderly controls, specifically in the precuneus–posterior cingulate region. These studies demonstrated the efficacy of entropy to represent the complexity of the brain.

Inspired by causality connectivity and entropy, we propose a novel method combining the Sample entropy (SampEn) (Richman and Moorman, 2000) with dynamic causality analysis for AD classification:

1. We calculated whole-brain causality connections for every pairwise regions-of-interests (ROIs). Different from the conventional methods that the causality connection was computed based on the entire time series, and a fixed scalar value was obtained regardless of time, we used a sliding window approach to model the dynamic causality connections. In this way, we expected to capture more rich temporal variations of the causality connectivity network.
2. We grouped the obtained causality connections into different clusters by a clustering algorithm and calculated SampEn values for every cluster.
3. The SampEn values of all clusters were concatenated to form a feature vector, and this vector was subject to various machine learning algorithms to explore the optimal classification results.

Compared with most AD classification methods based on the non-directional functional connectivity (Gao et al., 2020; Sanz-Arigita et al., 2010; Chen et al., 2016a; Wee et al., 2012), the proposed method has better neurological interpretations for the observed biomarkers by introducing the causality connections to account for the directional interactions of the brain. Moreover, in the proposed method, the features are extracted based on clusters of causality connections instead of the original causality connectivity network, which will help to better capture the pattern differences of neural activity and lower the computation load in the classification process.

2. Materials and methods

2.1. Materials and methods

All subjects selected in this study were from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database, including 30 AD subjects (15 females, 15 males, 75.3 ± 5.9 years) and 30 CN controls (15 females, 15 males, 73.7 ± 7.6 years). One female subject was excluded because of large head movements (more than 2.5 mm), leaving 59 subjects for further analysis. Both groups of subjects were age-matched ($p = 0.48$) and acquired with 3T Siemens Trio Tim scanner under the same protocol (T2*-weighted, TR = 3000 ms, TE = 30 ms, Flip Angle = 90°, imaging matrix = 448 pixels * 448 pixels, 48 slices, 197 volumes, and Slice Thickness = 3.4 mm). For every subject, his/her T1-weighted sequence (TR = 2300 ms, TE = 3.0 ms, Flip Angle = 9.0°) was acquired for co-registration in the preprocessing stage.

For each subject, the first 17 volumes of each subject were discarded for magnetic field's instability at the beginning and allowing for their adaption to the experimental circumstance. On the rest 180 volumes, we used SPM8 (https://www.fil.ion.ucl.ac.uk/spm/software/spm8) and REST (http://restfmri.net/forum/RESTplusV1.2) run on MATLAB R2013b (The MathWorks, Inc., Natick, MA, USA) to perform the standard preprocessing, which includes slicing time correction, head movement alignment using a six-parameter rigid-body spatial transformation (Friston et al., 1995), spatial standardization to the Montreal Neurological Institute (MNI) space, spatial smoothing with a Gaussian filter of 4 mm full width at half-maximum (FWHM), and low-band pass filtering ($0.01\ \text{Hz} \le f \le 0.08\ \text{Hz}$). After that, we removed the following possible nuisance covariates through linear regression: (1) the average whole-brain signal over the entire brain, (2) white matter and cerebrospinal fluid signal, (3) linear drift. Finally, the preprocessed images were parcellated into 90 ROIs based on AAL template and extracted corresponding mean BOLD signals.

2.2. Proposed framework

As shown in Fig. 1, we summarize our proposed method into three steps: brain network analysis, causality pattern clustering, and classification. The first step includes the matrix construction of a dynamic causality network (Section 2.2.1). The second step involves pattern extraction and entropy calculation (Section 2.2.2). The final step focuses on the concatenation of feature vectors, optimal parameters selection of classification, and performance evaluation (Section 2.2.3).

2.2.1. Brain network analysis

To ensure the stationarity of the time series before Granger Causality analysis, we employed differential processing (Granger, 1969) to smooth the BOLD signal series. Meanwhile, considering the possibility of the original sequence's autocorrelation, we selected the lag order corresponding to the smallest Akaike information criterion (AIC) (Akaike, 1974) as the optimal lag order.

Then, we adopted sliding time window approach for dynamic causality analysis. Each preprocessed BOLD signal series was divided into $K = (\lfloor (n-w)/s \rfloor + 1)$ subseries, where n is the number of all image volumes, w is the window width, and s is the step size of each move. We denote the l-th subject's k-th segmented BOLD subseries of the i-th ROI of as $x_i^l(k)$.

For each sliding window, the causality connection between X_t and Y_t

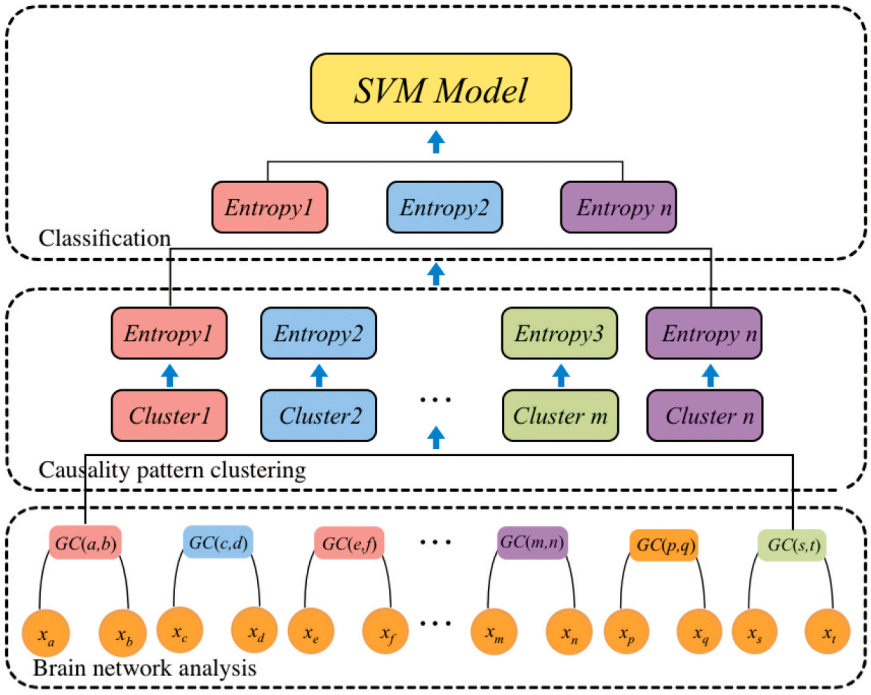

Fig. 1. The proposed framework for AD classification.

ROIs at time t was calculated by Bivariate Granger Causality using Statsmodels 2.1.0 (https://www.statsmodels.org/stable/index.html) on Python 3.8, with the autoregressive model as shown in Eq. (1):

$$Y_t = \sum_{p=1}^{p'} A_p X_{(t-p)} + \sum_{p=1}^{p'} B_p Y_{(t-p)} + CZ_t + \varepsilon_t$$

$$X_t = \sum_{p=1}^{p'} A_p' Y_{(t-p)} + \sum_{p=1}^{p'} B_p' X_{(t-p)} + CZ_t + \varepsilon_t' \quad (1)$$

where p' is the model order, A_p, A_p',B_p,B_p', and C are model coefficients, Z_t is exogenous variables, and ε_t and ε_t' are the prediction residual errors. If A_p is nonzero, which means X helps to predict Y, then it is said X causes Y, and vice versa (Granger, 1969). The F test was used to assess the strength of causality connection between X and Y ROIs.

We computed the dynamic causality connections of every pairwise ROIs for each subject. The causality connection of the k-th sliding window between the i-th and the j-th ROIs of the l-th subject is denoted as Eq. (2):

$$C_{i,j}^{l}(k) = GC\left(x_i^l(k), x_j^l(k)\right) \quad (2)$$

The dynamic causality series $S_{i,j}$ was obtained by concatenating all windows' causality connections between the i-th and the j-th ROIs of 59 subjects, as shown in Eq. (3), and the length of $S_{i,j}$ is $59 * K$.

$$S_{i,j} = \left[C_{i,j}^{1}(1), \ldots, C_{i,j}^{1}(K), C_{i,j}^{2}(1) \ldots, C_{i,j}^{2}(K), \ldots, C_{i,j}^{59}(1) \ldots, C_{i,j}^{59}(K)\right] \quad (3)$$

Finally, we stacked all $S_{i,j}$ to form a large-scale matrix Z of 90*89 rows and $59 * K$ columns as shown in Eq. (4), which contains the dynamic causality connections of all pairwise ROIs of 59 subjects.

$$Z = \begin{pmatrix} S_{1,2} \\ S_{1,3} \\ \vdots \\ S_{90,89} \end{pmatrix} = \begin{pmatrix} C_{1,2}^{1}(1), \ldots, C_{1,2}^{1}(K), \ldots, C_{1,2}^{59}(1) \ldots, C_{1,2}^{59}(K) \\ C_{1,3}^{1}(1), \ldots, C_{1,3}^{1}(K), \ldots, C_{1,3}^{59}(1) \ldots, C_{1,3}^{59}(K) \\ \vdots \\ C_{90,89}^{1}(1), \ldots, C_{90,89}^{1}(K), \ldots, C_{90,89}^{59}(1) \ldots, C_{90,89}^{59}(K) \end{pmatrix} \quad (4)$$

2.2.2. Causality pattern clustering

In order to capture the distinctive characteristics of neural activity pattern, agglomerative hierarchical clustering (Ward, 1963) was applied to group similar causality connections (i.e., the dynamic causality connections between the i-th and the j-th ROIs of all subjects) from the large-scale Z into the small-scale clusters. The advantages of agglomerative hierarchical clustering include no initialization, less dependence on hyperparameters, and robustness of clustering processing (Chen et al., 2016b). However, it should be noted that the clustering algorithm may lose the directionality of the causality connection. It is likely that the concatenated diagonal symmetrical connections (e.g., $S_{i,j}$ and $S_{j,i}$) will be clustered in the same group, which will bring about the confusion about the causality connection's directionality. To solve this problem, we separated Z into two matrices: $L = \left(S_{1,2}, S_{1,3,\ldots,}S_{i,j}\right), i \in [1,89], j \in [2,90], i < j$ and $U = \left(S_{2,1}, S_{3,1,\ldots,}S_{i,j}\right), i \in [2,90], j \in [1,89], i > j$, and then grouped them to u clusters, respectively. For both U and L, the initial number of clusters is (8100-90)/2 = 4005, where 8100 is the total number of causality connections, and 90 is the number of excluded pairwise causality connections between the same ROI. In the above procedure, we finally generated different numbers of clusters from 100 to 400 to achieve the optimal clustering effect for further classification, which is detailed in Section 3.1.

Then, we calculated SampEn (Yentes et al., 2013) value of each cluster to construct the feature vector for each subject. SampEn is a modification of approximate entropy and has the merit of independence of data length. It can be used to measure the complexity of physiological time-series signals, such as heart rate variability and EEG data. Here, we used it to assess the neural complexity of each cluster with similar causality connections, and the SampEn value of one of the l-th subject's clusters was calculated as follows:

1) Suppose this cluster includes q causality series and each series has K subseries partitioned by sliding windows, we concatenate them into a periodic time series P:

$$P = \left(C_1^l(1), \ldots, C_1^l(K), C_2^l(1), \ldots, C_2^l(K), \ldots, C_q^l(1), \ldots, C_q^l(K)\right) \quad (5)$$

Then, we denote P as Q for easy demonstration:

$$Q = (c(1), c(2), \ldots, c(q * K)) \quad (6)$$

Here, the $c(1), .., c(q * K)$ corresponds to each element in the series P.

2) Reconstruct Q with m consecutive data points into $N = (q * K - m + 1)$ subseries. Each subseries can be denoted as $Q(v)$:

$$Q(v) = (c(v), c(v+1), c(v+2), \ldots, c(v+m-1)), v = 1, 2, \ldots, N \quad (7)$$

3) Calculate the distance $d_{vv'}$ between $Q(v)$ and other $Q(v'), v' = 1, 2, \ldots, N$ by Chebyshev distance, that is,

$$d_{vv'} = \max|c(v+k) - c(v'+k)|, k = 0, 1, \ldots m-1, v \neq v' \quad (8)$$

4) Given a threshold $f = r * SD$ (r is the similarity criterion and SD is the standard deviation of the original Q), D_v^m represents the proportion of $d_{vv'} < f$ under m data points.

5) By averaging across all D_v^m, we get

$$S^m = \frac{1}{N-m} \sum_{i=1}^{N-m} D_v^m \quad (9)$$

and

$$S^{m+1} = \frac{1}{N-m} \sum_{i=1}^{N-m} D_v^{m+1} \quad (10)$$

6) The SampEn value of this cluster is calculated as

$$SampEn = -\ln\left(S^{m+1}/S^m\right) \quad (11)$$

Following the above steps, we can obtain all clusters' SampEn values of every subject, with a larger SampEn value corresponding to a more complex neural pattern in that cluster and a smaller one representing the decreased complexity. Here, the m and r are two key parameters to calculate SampEn values, thus we discussed the selection of these two parameters in Section 3.2.

2.2.3. Classification

We used SampEn values of causality connection clusters obtained in Section 2.2.2 as the potential features for the subsequent AD classification. Considering that excessive features may increase the training load and decrease the generalization performance of models, we adopted variable importance measures (VIM) to reduce redundant features by Random Forest (Bi et al., 2018), with Gini index evaluating the VIM value. A higher VIM value corresponds to a more discriminative feature, which means it is more helpful for the AD classification. In this process, we ranked the VIM value of every feature from high to low and set a threshold a to remove the redundant features below a. The a was varied in the range $[0.0001, \ldots, 0.001]$ to get an optimal feature subset where features are discriminative for classification. Under the optimal condition ($u = 300, w = 20$) (see Section 3.1), we selected 13 SampEn values as final features (the corresponding sample-to-feature ratio is 59:13) after the VIM dimensionality reduction.

After dimensionality reduction, we used Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and SVM cluster (Bi et al., 2018) by scikit-learn 0.23 (https://scikit-learn.org/) to train the AD classification model. To obtain the optimal performance, it is crucial here to determine the optimal parameters (i.e., the window size w and cluster number u) and the corresponding hyperparameters of each method (e.g., the number of estimators of the classifier). We selected the optimal combination of parameters and hyperparameters by grid search cross-validation, with the models' optimal parameters being varied in the following ranges: $w \in [20, \ldots, 80]$, $u \in [100, \ldots, 400]$, and the optimal hyperparameters being determined for each method as described below.

Specifically, for SVM, we examined different kernel functions (i.e., Gaussian Radial Basis Kernal, Poly Kernel, and Sigmoid Kernal) with their corresponding hyperparameters (e.g., the penalty term coefficient, intercept, and gamma). For Random Forest classifier, we determined the estimator number, the max depth, and the maximum leaf node, while other hyperparameters were set to default values. For SVM cluster, we used 150 estimators (the optimal number), and separately trained them with randomly selected partial samples under their optimal hyperparameters (the same way as SVM classifier). To reduce the effect of poor classifiers, each classifier was set a voting weight by a z-score transformation in terms of its classification accuracy. The final result is determined by the majority of classifiers' predictions. For XGBoost, we selected hyperparameters for the tree booster (e.g., the learning rate, the estimator number, the maximum depth, and the minimum child weight), and tuned them respectively. In the above classifiers, we introduced regularization for SVM, SVM cluster, and XGBoost classifiers with squared L2 penalty.

All these methods were evaluated by LOOCV (Leave-one-out Cross Validation), and their performances are shown in Table 1. The accuracy (ACC), sensitivity (SEN), specificity (SPE), and F1-score were used to evaluate the performances of the four methods. Among the four methods, SVM achieved the best performance, SVM cluster and Random Forest obtained slightly lower performances than SVM, and XGBoost obtained the lowest performances. We plotted the learning curves (Perlich, 2010), which show the predictive generalization performance as a function of the number of training examples, with 5-fold cross-validation to explore the differences between the four methods. It is found that SVM (Fig. 2a), SVM cluster (Fig. 2b), and Random Forest (Fig. 2c) achieve good learning effect, with the variance of their training and test scores gradually decreases with the increase of training samples. In contrast, XGBoost shows a relatively high variance between training and test scores, indicating the poor learning effect in comparison with the other three methods. We further plotted the curve of loss against training epochs for XGBoost (Fig. 3). It can be seen that the training loss gradually decreases while the test loss always maintains a high level, suggesting the XGBoost suffers from overfitting. We suggest that XGBoost, as a complex iterative ensemble model, may require a large number of samples to achieve good performances. The SVM method is more suitable for the fMRI classification task where the sample size is generally small.

3. Experimental results and discussion

3.1. Effect of sliding window width and clustering number on accuracy

The width of the sliding window w and the number of clusters u are key factors to affect the classification performance. If a small window width w is selected, it is high possibility that the noise signals are mistakenly taken as causality variations, though the possible dynamic causality can be better detected. Conversely, with a larger width w, it can get a more stable causality connection but easily lose causality variations. Meanwhile, the number of clusters u would also modulate the classification accuracy. If we divide the causality network into a big number u of clusters, each cluster will only include few causality series and cannot well represent a neural pattern. Conversely, with a small

Table 1
Performance comparison among the four methods.

	ACC (%)	SEN (%)	SPE (%)	F-Score
XGBoost	79.66	70.00	89.66	0.7777
SVM cluster	84.75	86.67	82.76	0.8525
Random Forest	86.44	86.66	86.20	0.8667
SVM	89.83	90.00	89.66	0.8852

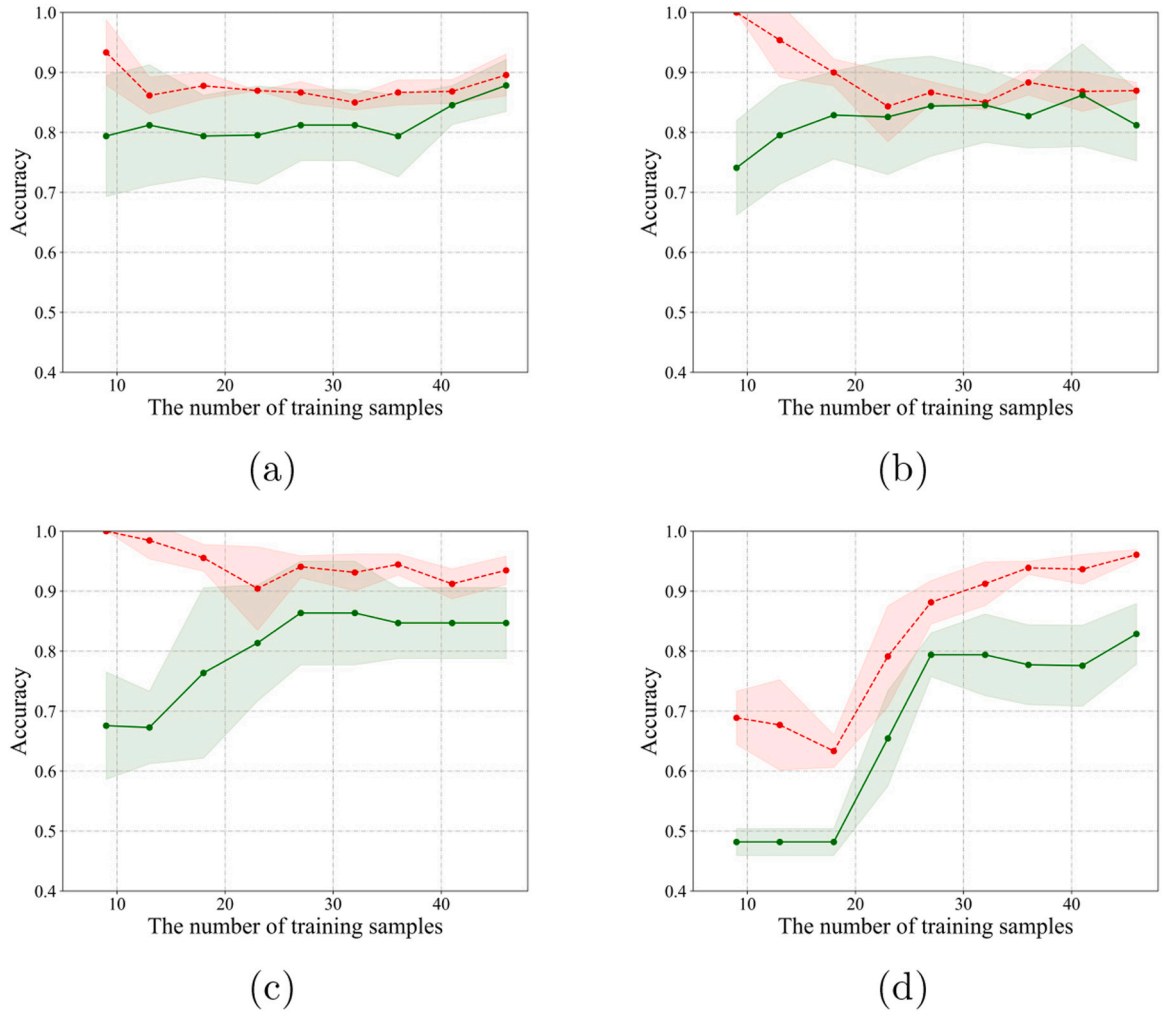

Fig. 2. The learning curves with 5-fold cross-validation of (a) SVM, (b) SVM clusters, (c) Random Forest and (d) XGBoost (training accuracy: red; test accuracy: green; standard deviation: shaded area). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

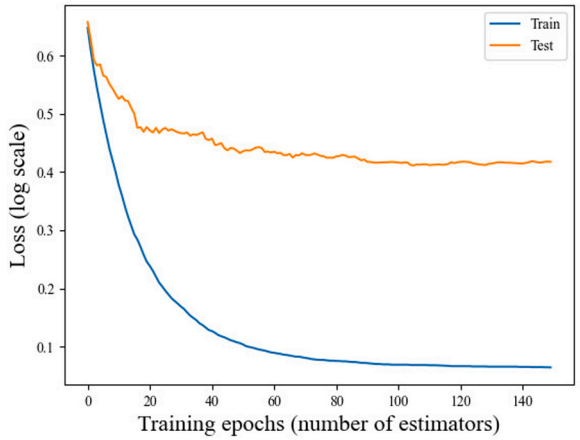

Fig. 3. The loss against training epochs curve of XGBoost. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

number u of clusters, each cluster would contain more causality series and obscure its unique characteristics.

In this study, we used grid search cross-validation to find the optimal w and u (mentioned in Section 2.2.3). To be specific, we changed the cluster number u from 100 to 400 with a step 100, and width w from 20 to 80 with a step 20 to find the best classification performance. Fig. 4 shows the classification accuracy under various parameter combinations with SVM. The best performance is 89.83% when $w = 20$ and $u = 300$, and the worst performance is 67.8% when $w = 60$ and $u = 100$. We note that the cases of $w = 80$ show stable and high performances. This may be due to that the sliding window with a big width catches more global features, thus obtaining relatively better performance.

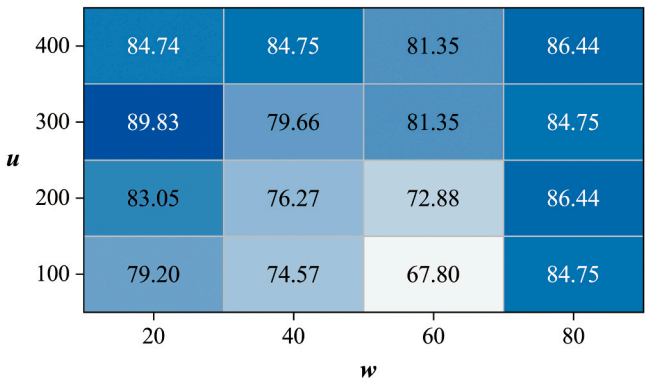

Fig. 4. The classification accuracy (%) in terms of the number of clusters u and width of the sliding window w with SVM model.

3.2. Effect of the parameters of SampEn on accuracy

The number of data points m and the similarity criterion r are two essential parameters for SampEn's calculation, and further influence the classification accuracy. So, it is important to select suitable parameters of m and r. Some researchers have discovered that $m = 2$ and $0.1 < r < 0.25$ show good statistical power when the series length is larger than 200 (Yentes et al., 2013; Wee et al., 2016; Echávarri et al., 2011). Considering the length of our concatenated series is larger than 200, we fixed m at 2 and examined r in $[0.13, 0.25]$. Fig. 5 shows the effect of r on accuracy under SVM classifier with its optimal hyperparameters. It is found that the optimal result has been achieved at $r = 0.2$. Therefore, $m = 2$ and $r = 0.2$ are fixed when calculating each clusters' SampEn values.

3.3. Classification performance

In this paper, we compared the performance of our framework with conventional methods which are based on the Pearson's correlation network (PCN) (Wee et al., 2016), dynamic Pearson's correlation network (DPCN) (Wee et al., 2016), High-order correlation network (HOCN) (Chen et al., 2016a), and causality correlation network (CCN) (Khazaee et al., 2017), respectively. Among them, both PCN and DPCN methods used the low-level brain functional network for classification, with Pearson's correlation representing pairwise functional connectivity among brain areas. The distinction between them is that, in PCN, the entire BOLD time series was directly used to calculate the functional connectivity; while in DPCN, the time sliding window is introduced to characterize the dynamic variations of the functional connection network. In contrast, HOCN is a recently proposed method based on DPCN, which further used the high-level correlation between two dynamic functional connection series to construct a high-order network. Finally, different from the above three methods using the non-directional functional connection network, the CCN-based method constructed a directional causality connection network and extracted its graphical features, such as indegree and outdegree of nodes, for the automatic classification.

To facilitate the comparison, we implemented the above four methods using Python 3.8 and trained them with the same dataset described in Section 2.1. As we can see from Table 2, the performance of DPCN is higher than those of PCN, suggesting that dynamic connectivity analysis captures the neuropathological biomarker for the AD classification. The CCN method using causality connections achieves higher performance than that of DPCN. Moreover, the HOCN achieves higher performance than that of DPCN and CCN. The overall performance of our method is slightly higher than those of HOCN, with the approximately same accuracy and sensitivity, and a 3% increase in specificity.

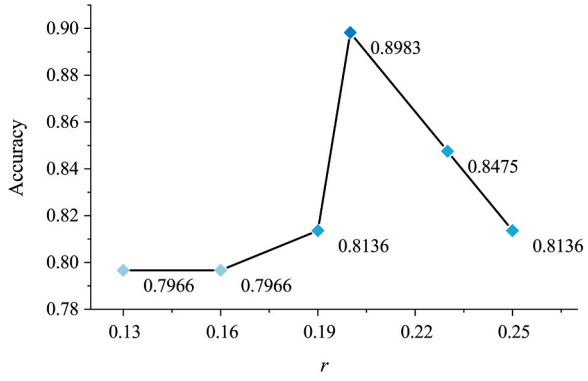

Fig. 5. The classification accuracy in terms of the similarity criterion r with SVM model.

Table 2
Performance comparison between our proposed framework and other methods.

Method	ACC (%)	SEN (%)	SPE (%)	F1-Score
PCN	71.92	74.07	70.00	0.7421
DPCN	79.66	82.14	77.42	0.8000
HOCN	88.14	89.66	86.67	0.8814
CCN	74.58	83.33	65.51	0.7692
Our proposed method	89.83	90.00	89.66	0.9000

3.4. Temporal complexity

We analyzed the temporal complexity of our proposed method. As described above, it includes three steps of feature extraction, i.e., brain network analysis, causality pattern clustering, and SampEn computation. Among the three steps, the SampEn computation is most time-consuming and therefore dominates the temporal complexity. Because the SampEn needs to be computed for every cluster of all subjects, the overall temporal complexity of SampEn computation can be written as $O(u * S) * O(f(SampEn))$, where u denotes the number of clusters (see Section 2.2.3), S denotes the number of subjects, and $O(f(SampEn))$ denotes the complexity of one SampEn computation.

In terms of the definition of SampEn, the complexity of one SampEn computation is determined by the number M of causality connection series in the cluster and the length of causality connection K (see Section 2.2.1). Thus, $O(f(SampEn))$ can be written as $O(K * M)$.

Interestingly, the M here is inverse proportion to u, because the increasing number u of clusters will inevitably lead to less causality connection series in each cluster, and vice versa. Considering there is a total of 8100 causality connection series (90 ROIs*90 ROIs), the average M in each cluster here can be represented as $8100/u$, and the overall temporal complexity T of SampEn computation is $O(8100 * S * K)$ as derived in Eq. (12):

$$T = O(u * S) * O(K * M) = O(u * S * K * M)$$
$$= O(u * S * K * 8100/u) = O(8100 * S * K) \quad (12)$$

Considering that 8100 is an invariant constant, the temporal complexity of the proposed method is therefore linearly related to the number of subjects S and the length of causality connection K.

3.5. Merits and limitations

One of the merits is that, in comparison with the existed methods, the proposed method has a fine temporal sensitivity by introducing the sliding window. This can be verified by comparing the static causality network to the dynamic causality network. Fig. 6a shows a static causality network of an AD patient, which provides a global view for the causality connections between pairwise ROIs. Fig. 6b–g show the first six dynamic causality networks of the same patient in the same way. It can be clearly seen that the dynamic causality network in Fig. 6b–g shows much richer temporal variations of causality connections than that in Fig. 6a (Yellow indicates the strong causality connections in Fig. 6), with the strength of static causality connections fluctuating in a relatively limited range ($2.14 \sim 30.72$), and the strength of dynamic causality connections varies in a wide range ($0.01 \sim 40.61$). This is due to the difference of computation methods between static and dynamic causality connections, leading to different temporal sensitivities. Static causality connections are computed in terms of an entire BOLD signal series, which would mix the various types of fast neural activities, and obscure the detailed causality information. In contrast, dynamic causality connections are computed in terms of BOLD signal in a short sliding window, which can easily capture the spikes of causality strength in fine-grain time level. We can further observe the variations of causality connections as the sliding window move afterward. This would help to clarify the temporal properties of pathological biomarkers for AD patients. Compared with methods using entire BOLD signals, we can

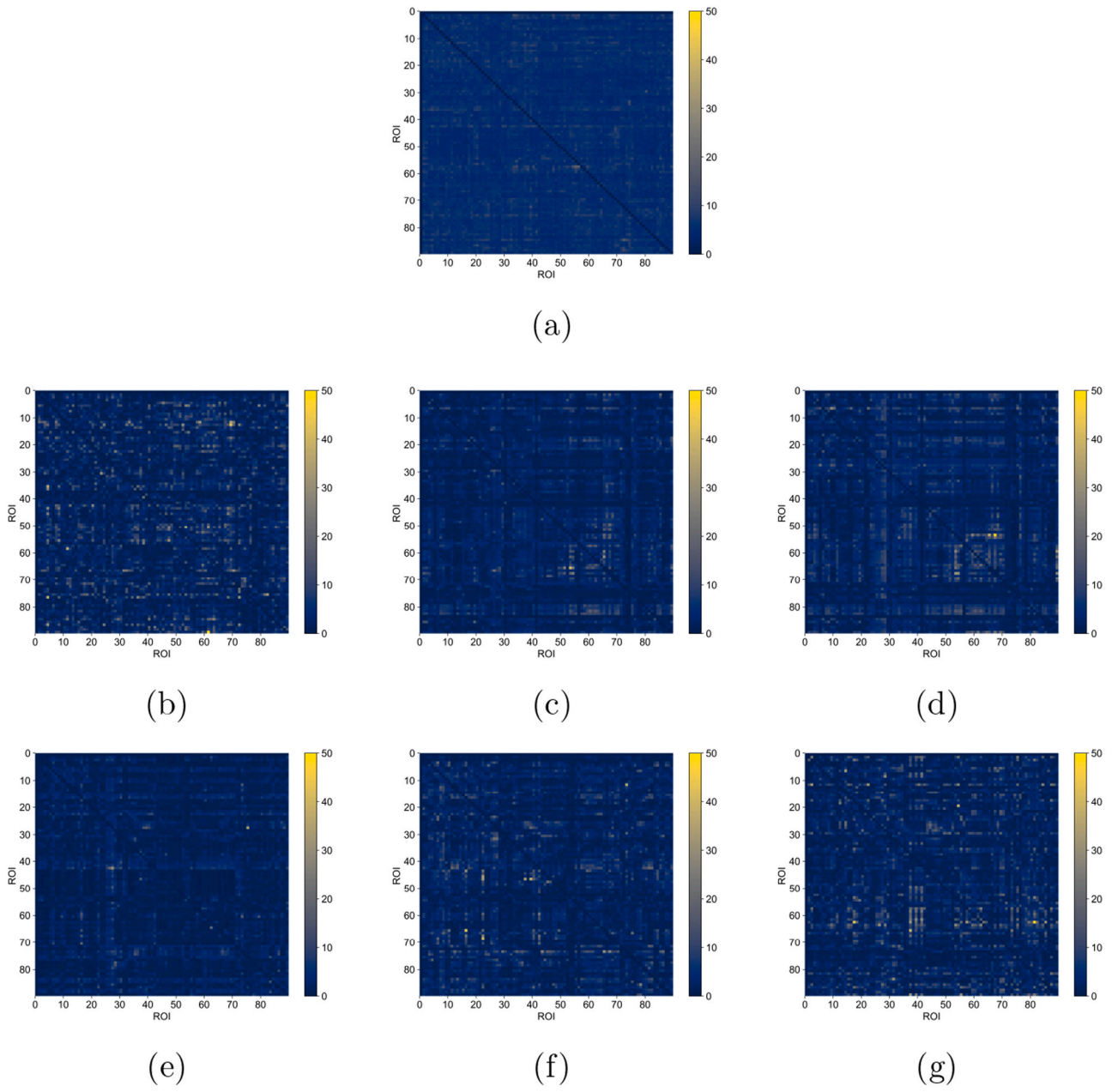

Fig. 6. (a) The static connectivity network and (b–g) the first six dynamic connectivity networks partitioned by a sliding window for one AD subject. The yellow denotes strong causality connection strength, and the blue denotes weak causality connection strength. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

capture the richer signal variations of neural activity in the brain, which helps to clarify the AD pathological biomarkers.

The other merit of the proposed method lies in its ability to explore similar neural activities from different brain areas, by using agglomerative hierarchical clustering. Fig. 7 shows the visualization of the clustering process, with the original 15 dynamic causality connections in Fig. 7a, and their clusters indicated by three different colors in Fig. 7b. It can be seen that the causality series with different fluctuations consistently fall into the clusters indicated in red, green, and blue, suggesting that the clustering algorithm correctly captures the dynamic characteristics of the causality series, extracting various causality patterns of neural activities.

The other merit of the proposed method lies in its ability to explore similar neural activities from different brain areas, by using agglomerative hierarchical clustering. Fig. 7 shows the visualization of the clustering process, with the original 15 dynamic causality connections in Fig. 7a, and their clusters indicated by three different colors in Fig. 7b. It can be seen that the causality series with different fluctuations consistently fall into the clusters indicated in red, green, and blue, suggesting that the clustering algorithm correctly captures the dynamic characteristics of the causality series, extracting various causality patterns of neural activities.

A limitation of the proposed method is the relatively high computational load. Due to the introduction of sliding window and directional causality connection, the temporal complexity of the proposed method is $O(8100 * S * K)$ as described above. With the increase of S and/or K, the computational load is higher than those of conventional methods without the dynamic sliding window, e.g., PCN ($O(4050 * S)$) using non-directional functional connectivity, and CCN ($O(8100 * S)$) using directional causality connectivity, which is a potential demerit of our

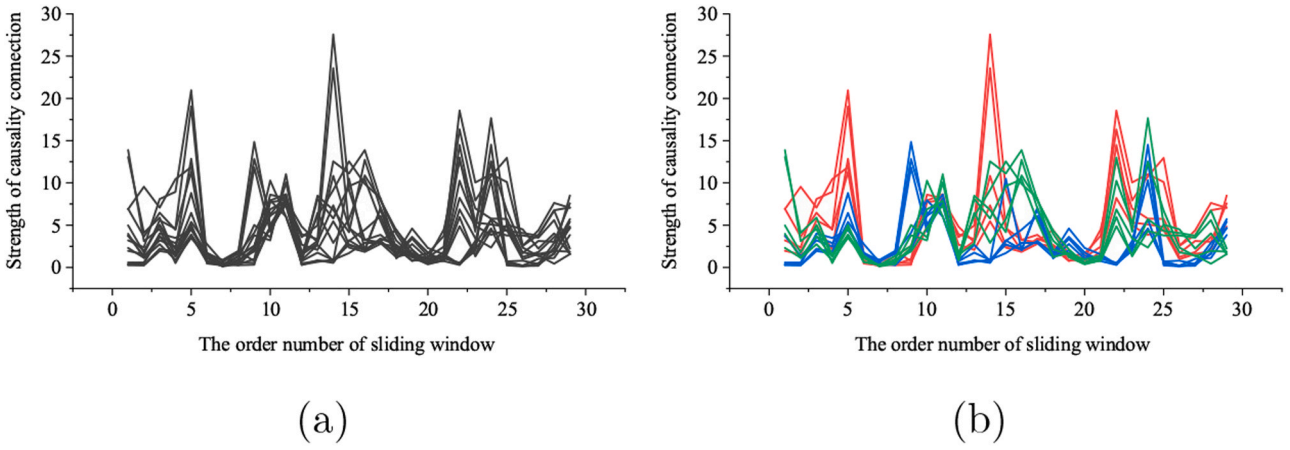

Fig. 7. (a) Original causality time series and (b) their clustering results for one AD subject.

method. In practice, considering a limited number of samples in the medical classification application, this computational load is acceptable, and can be lower by adjusting the length of the sliding window and its step.

3.6. SampEn differences and their pathological implications

To discover the difference of SampEn values between AD and CN, we applied dimensionality reduction by principal component analysis. We can see that the AD and CN groups' features after dimensionality reduction are distributed with obvious distinction on a two-dimensional plane in Fig. 8a. The CN controls' features mainly lie on the left side of the plane, and AD patients' features are mainly located from the middle to the right of the plane. It also can be observed that the features of CN group are more concentratedly distributed than those of AD group. This may be due to that the CN controls have relatively consistent neural activity, while AD patients have poor neural activity consistency due to various brain function impairments. In Fig. 8b, the violin plot combines the box plot with density traces to exploit the difference of mean SampEn values between two groups. By the two-sample t-test using SPSS, we can find that the overall SampEn values of CN group ($Mean \pm SD$:1.31 ± 0.086) are significantly higher than those of AD group ($Mean \pm SD$:1.20 ± 0.095), $t(57) = 4.660$, $p < 0.005$).

The top four discriminative clusters for classifying AD patients and CN controls are shown in Fig. 9. These clusters were selected in terms of their contributions to classification performances measured by the VIM method as described in 2.2.3, in terms of SampEn features of AD and CN groups. To exploit the difference between AD and CN groups of these four clusters, we performed a two-way ANOVA, with the SampEn values of the clusters of each subject as the dependent variable, and group (2 levels, AD group or CN group) and cluster (4 levels, 4 clusters) as fixed factors. There is a significant main effect of group, $F(1, 228) = 80.55$, $p < 0.001, \eta_p^2 = 0.26$, indicating the lower SampEn values of AD group than those of CN group. There is also significant interaction between group and cluster, $F(3, 228) = 3.65$, $p < 0.02$, $\eta_p^2 = 0.05$, i.e., the SampEn difference between CN and AD groups in the cluster 1 ($Mean = 0.41$) is much greater than those in the clusters 2 ($Mean = 0.24$), cluster 3 ($Mean = 0.17$), and cluster 4 ($Mean = 0.19$) (See Fig. 9c), although the SampEn difference between AD and CN groups in four clusters are all significant, revealed by simple effect analysis (See Table 3). In addition to the statistical difference described above, we further observed that, although a cluster generally includes many causality series, only a few ROIs are the leading factors to interact with many other ROIs. For instance, there are only two leading ROIs, i.e., Hippocampus right (38) and Para hippocampal gyrus right (40) in the cluster 1 (Fig. 9(a), coral color), these two ROIs, however, cause causality interactions with eight ROIs (e.g., Superior frontal gyrus, Middle frontal gyrus, and Cingulate gyrus). This shows these specific ROIs may play pivotal roles in certain neural activity patterns, which are potential candidates for pathological biomarkers of AD.

We select 12 ROIs in the top 4 discriminative clusters, which are distributed in Frontal lobe, Parietal lobe, and Temporal lobe. 10 of them are consistent with previous physiology studies (see Table 3). For

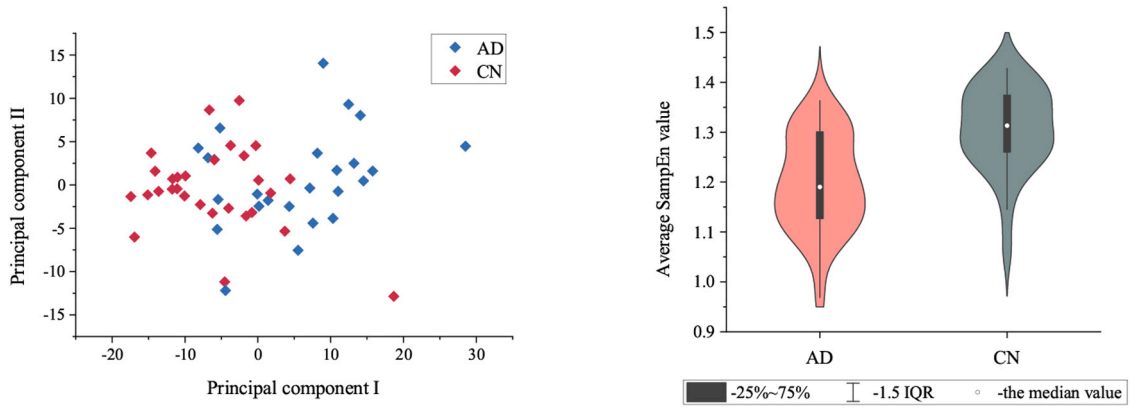

Fig. 8. (a) The feature distributions and (b) the violin plot of AD and CN groups under the optimal parameters (cluster = 300; width of the sliding window = 20).

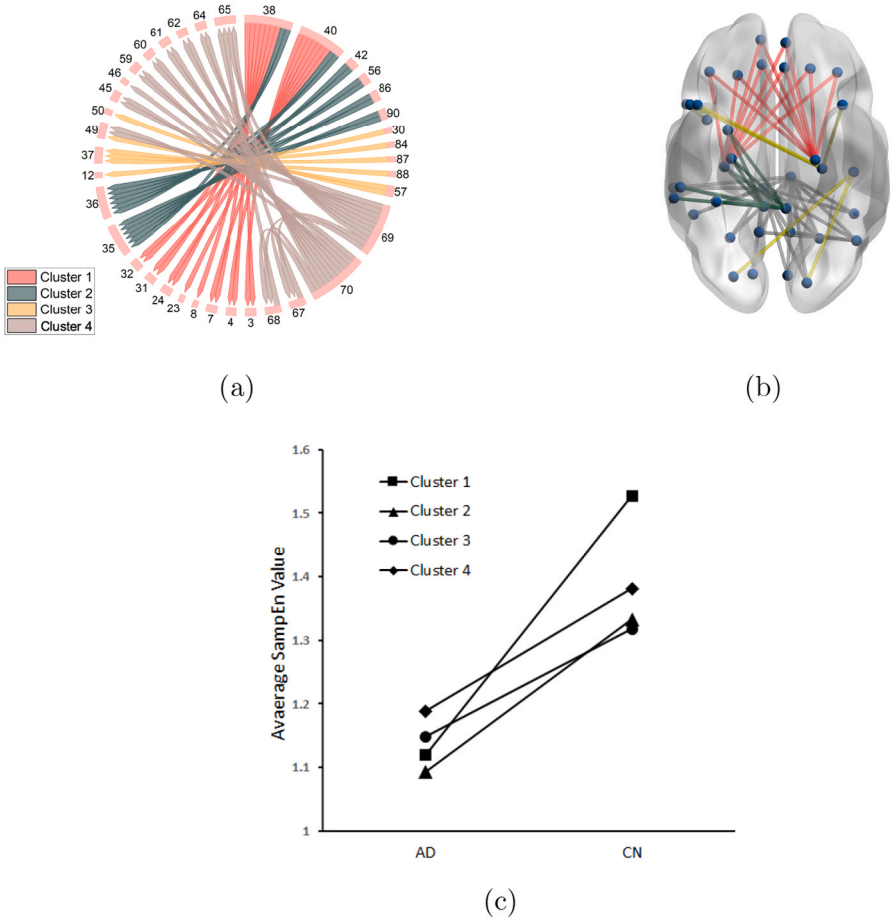

Fig. 9. (a) The top 4 most discriminative clusters described by chord chart (No.1 - coral, No.2 - Green, No.3 - Yellow, No.4 - Gray), (b) their locations in brain, and (c) the comparisons between of AD and CN groups for these four clusters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
The mean and statistical results of the top 4 most discriminative clusters.

No.	Simple effect analysis	CN (Mean ± SD)	AD (Mean ± SD)
1	$F(1,228) = 53.01, p < 0.001$	1.52 ± 0.24	1.11 ± 0.29
2	$F(1,228) = 17.89, p < 0.001$	1.33 ± 0.24	1.09 ± 0.20
3	$F(1,228) = 8.95, p < 0.003$	1.32 ± 0.21	1.15 ± 0.16
4	$F(1,228) = 11.91, p < 0.001$	1.38 ± 0.16	1.19 ± 0.19

example, Echávarri et al. (2011) observed that Para hippocampal gyrus and hippocampus are important biomarkers, which favors the leading factors (i.e., Right Hippocampus, Right Para hippocampal gyrus) observed in the top two distinctive patterns – No.1 and No.2. Foundas et al. (1997) found AD have significant atrophy of insula than healthy controls. These studies further validate the effectiveness of our method and indicate that the physiological changes may cause a decrease in the complexity of neural activities. On the other hand, we also discovered two ROIs (i.e., temporal pole of middle temporal gyrus and left Post-central gyrus) that haven't been found in relevant physiology studies. These two ROIs are reported being related to emotional and social behavior, semantic presentation, memory, and language expression. Considering the AD patients also exhibit the clinical symptoms of these cognitive functions, these two areas are potential biomarkers for AD and require further investigations (Table 4).

Table 4
The ROIs selected from the top 4 discriminative clusters.

ROI index	ROI name	Citations
38 (1,2)	Hippocampus right	Echávarri et al. (2011)
40 (1,2)	Para hippocampal gyrus right	Echávarri et al. (2011)
42 (2)	Amygdala right	Poulin et al. (2011)
56 (2)	Fusiform right	Caramelli et al. (1998)
86 (2)	Middle temporal gyrus right	Bandaru et al. (2009)
90 (2)	Inferior temporal gyrus right	Scheff et al. (2011)
30 (3)	Insula right	Foundas et al. (1997)
84 (3)	Temporal pole: superior temporal gyrus right	Hänggi et al. (2011)
67 (4)	Precuneus left	Rami et al. (2012)
68 (4)	Precuneus Right	Rami et al. (2012)
69 (4)	Paracentral lobule left	Kang et al. (2013)
70 (4)	Paracentral lobule right	Kang et al. (2013)

4. Conclusions

In this study, we proposed an entropy-based measure of causality brain networks on the basis of the rs-fMRI data to classify AD and CN. The proposed method is able to efficiently discriminate AD and CN with an accuracy of 89.83% with SVM, which shows that SampEn of neural patterns can be used as promising features for AD classification. Also, we observed that the overall causality pattern's SampEn values of the AD are significantly reduced compared with those of the CN, indicating a decreased complexity resulting from disrupted neural activities in AD patients. Furthermore, we found that, although the ROIs in the discriminative clusters are widely distributed in different brain regions,

some of them (e.g., Right Hippocampus, Right Para hippocampal gyrus) serve as pivotal nodes in the causality network, and exhibit a one-to-many link to the other ROIs, accounting for neural distinctions between AD patients and CN controls. Our work provides an accurate and light framework for AD classification, and helps to get a better understanding of the neuropathological mechanism for AD.

CRediT authorship contribution statement

Yuanchen Wu: Investigation, Software, Formal analysis, Visualization, Writing - original draft. **Yuan Zhou:** Conceptualization, Methodology, Supervision, Data curation, Writing - review & editing, **Miao Song**: Methodology, Formal analysis, Writing - review & editing.

Acknowledgements

This research was supported by the National Natural Science Foundation of China under Grants No. 31870979 and No. 61872231.

Conflict of Interest

The authors declare no conflict of interest.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723.
Bandaru, V.V., Troncoso, J., Wheeler, D., Pletnikova, O., Wang, J., Conant, K., Haughey, N.J., 2009. ApoE4 disrupts sterol and sphingolipid metabolism in Alzheimer's but not normal brain. Neurobiol. Aging 30 (4), 591–599.
Bi, X.-A., Wang, Y., Shu, Q., Sun, Q., Xu, Q., 2018. Classification of autism spectrum disorder using random support vector machine cluster. Front. Genet. 9, 18.
Caramelli, P., Robitaille, Y., Laroche-Cholette, A., Nitrini, R., Gauvreau, D., Joanette, Y., Lecours, A.R., 1998. Structural correlates of cognitive deficits in a selected group of patients with Alzheimer's disease. Neuropsychiatry Neuropsychol. Behav. Neurol. 11 (4), 184–190.
Chen, X., Zhang, H., Gao, Y., Wee, C.Y., Li, G., Shen, D., Alzheimer's Disease Neuroimaging, I., 2016a. High-order resting-state functional connectivity network for MCI classification. Hum. Brain Mapp. 37 (9), 3282–3296.
Chen, X., Zhang, H., Shen, D., 2016b. Ensemble hierarchical high-order functional connectivity networks fo rMCI classification. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer.
Dennis, E.L., Thompson, P.M., 2014. Functional brain connectivity using fMRI in aging and Alzheimer's disease. Neuropsychol. Rev. 24 (1), 49–62.
Echávarri, C., Aalten, P., Uylings, H.B., Jacobs, H.I., Visser, P.J., Gronenschild, E.H., Verhey, F.R., Burgmans, S., 2011. Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer's disease. Brain Struct. Funct. 215 (3–4), 265–271.
Foundas, A.L., et al., 1997. Atrophy of the hippocampus, parietal cortex, and insula in Alzheimer's disease: a volumetric magnetic resonance imaging study. Neuropsychiatry Neuropsychol. Behav. Neurol.
Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.B., Heather, J.D., Frackowiak, R.S.J., 1995. Spatial registration and normalization of images. Hum. Brain Mapp. 3 (3), 165–189.
Gao, Z., Feng, Y., Ma, C., Ma, K., Cai, Q., Alzheimer's DiseaseNeuroimaging, I., 2020. Disrupted time-dependent and functional connectivity brain network in Alzheimer's disease: a resting-state fMRI study based on visibility graph. Curr. Alzheimer Res. 17 (1), 69–79.
Granger, C.W., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 424–438.
Hänggi, J., Streffer, J., Jäncke, L., Hock, C., 2011. Volumes of lateral temporal and parietal structures distinguish between healthy aging, mild cognitive impairment, and Alzheimer's disease. J. Alzheimer's Dis. 26 (4), 719–734.
Kang, K., Yoon, U., Lee, J.M., Lee, H.W., 2013. Idiopathic normal-pressure hydrocephalus, cortical thinning, and the cerebrospinal fluid tap test. J. Neurol. Sci. 334 (1–2), 55–62.
Khazaee, A., Ebrahimzadeh, A., Babajani-Feremi, A., Alzheimer's Disease Neuroimaging, I., 2017. Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI. Behav. Brain Res. 322, 339–350.
Liang, P., Li, Z., Deshpande, G., Wang, Z., Hu, X., Li, K., 2014. Altered causal connectivity of resting state brain networks in amnesic MCI. PLoS One 9 (3), e88476.
Liu, Z., Zhang, Y., Bai, L., Yan, H., Dai, R., Zhong, C., Wang, H., Wei, W., Xue, T., Feng, Y., You, Y., Tian, J., 2012. Investigation of the effective connectivity of resting state networks in Alzheimer's disease: a functional MRI study combining independent components analysis and multivariate Granger causality analysis. NMR Biomed. 25 (12), 1311–1320.
Lv, J., Jiang, X., Li, X., Zhu, D., Chen, H., Zhang, T., Zhang, S., Hu, X., Han, J., Huang, H., Zhang, J., Guo, L., Liu, T., 2015. Sparse representation of whole-brain fMRI signals for identification of functional networks. Med. Image Anal. 20 (1), 112–134.
Mega, M.S., Cummings, J.L., Fiorello, T., Gornbein, J., 1996. The spectrum of behavioral changes in Alzheimer's disease. Neurology 46 (1), 130–135.
Niu, Y., Wang, B., Zhou, M., Xue, J., Shapour, H., Cao, R., Cui, X., Wu, J., Xiang, J., 2018. Dynamic complexity of spontaneous bold activity in Alzheimer's Disease and mild cognitive impairment using multiscale entropy analysis. Front. Neurosci. 12, 677.
Perlich, C., Learning Curves in Machine Learning. 2010.
Poulin, S.P., Dautoff, R., Morris, J.C., Barrett, L.F., Dickerson, B.C., Alzheimer's Disease Neuroimaging, I., 2011. Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. Psychiatry Res. Neuroimaging 194 (1), 7–13.
Rami, L., Sala-Llonch, R., Solé-Padullés, C., Fortea, J., Olives, J., Lladó, A., Peña-Gómez, C., Balasa, M., Bosch, B., Antonell, A., Sanchez-Valle, R., Bartrés-Faz, D., Molinuevo, J.L., 2012. Distinct functional activity of the precuneus and posterior cingulate cortex during encoding in the preclinical stage of Alzheimer's disease. J. Alzheimer's Dis. 31 (3), 517–526.
Richman, J.S., Moorman, J.R., 2000. Physiological time-series analysis using approximate entropy and sample entropy. Am. J. Physiol. Heart Circ. Physiol. 278, 2039–2049.
Sanz-Arigita, E.J., Schoonheim, M.M., Damoiseaux, J.S., Rombouts, S.A., Maris, E., Barkhof, F., Scheltens, P., Stam, C.J., 2010. Loss of 'small-world' networks in Alzheimer's disease: graph analysis of FMRI resting-state functional connectivity. PLoS One 5 (11), e13788.
Scheff, S.W., Price, D.A., Schmitt, F.A., Scheff, M.A., Mufson, E.J., 2011. Synaptic loss in the inferior temporal gyrus in mild cognitive impairment and Alzheimer's disease. J. Alzheimer's Dis. 24 (3), 547–557.
Toussaint, P.-J., Maiz, S., Coynel, D., Doyon, J., Messé, A., de Souza, L.C., Sarazin, M., Perlbarg, V., Habert, M.O., Benali, H., 2014. Characteristics of the default mode functional connectivity in normal ageing and Alzheimer's disease using resting state fMRI with a combined approach of entropy-based and graph theoretical measurements. NeuroImage 101, 778–786.
Wang, B., Niu, Y., Miao, L., Cao, R., Yan, P., Guo, H., Li, D., Guo, Y., Yan, T., Wu, J., Xiang, J., Zhang, H., 2017. Decreased complexity in Alzheimer's disease: resting-state fMRI evidence of brain entropy mapping. Front. Aging Neurosci. 9, 378.
Wang, J., Wilson, R.C., Hancock, E.R., 2017b. Detecting Alzheimer's disease using directed graphs. In: International Workshop on Graph-Based Representations in Pattern Recognition, Springer.
Ward Jr., J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58 (301), 236–244.
Wee, C.-Y., Yap, P.T., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D., 2012. Resting-state multi-spectrum functional connectivity networks for identification of MCI patients. PLoS One 7 (5), e37828.
Wee, C.-Y., Yang, S., Yap, P.T., Shen, D., Alzheimer's Disease Neuroimaging, 2016. Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification. Brain Imaging Behav. 10 (2), 342–356.
Yentes, J.M., Hunt, N., Schmid, K.K., Kaipust, J.P., McGrath, D., Stergiou, N., 2013. The appropriate use of approximate entropy and sample entropy with short data sets. Ann. Biomed. Eng. 41 (2), 349–365.