Identification of Alzheimer's EEG With a WVG Network-Based Fuzzy Learning Approach

Haitao Yu[1]*, Lin Zhu[1], Lihui Cai[1], Jiang Wang[1], Jing Liu[2]*, Ruofan Wang[3] and Zhiyong Zhang[4]

[1] School of Electrical and Information Engineering, Tianjin University, Tianjin, China, [2] Department of Neurology, Tangshan Gongren Hospital, Tangshan, China, [3] School of Information Technology Engineering, Tianjin University of Technology and Education, Tianjin, China, [4] Department of Pathology, Tangshan Gongren Hospital, Tangshan, China

A novel analytical framework combined fuzzy learning and complex network approaches is proposed for the identification of Alzheimer's disease (AD) with multichannel scalp-recorded electroencephalograph (EEG) signals. Weighted visibility graph (WVG) algorithm is first applied to transform each channel EEG into network and its topological parameters were further extracted. Statistical analysis indicates that AD and normal subjects show significant difference in the structure of WVG network and thus can be used to identify Alzheimer's disease. Taking network parameters as input features, a Takagi-Sugeno-Kang (TSK) fuzzy model is established to identify AD's EEG signal. Three feature sets—single parameter from multi-networks, multi-parameters from single network, and multi-parameters from multi-networks—are considered as input vectors. The number and order of input features in each set is optimized with various feature selection methods. Classification results demonstrate the ability of network-based TSK fuzzy classifiers and the feasibility of three input feature sets. The highest accuracy that can be achieved is 95.28% for single parameter from four networks, 93.41% for three parameters from single network. In particular, multi-parameters from the multi-networks set obtained the best result. The highest accuracy, 97.12%, is achieved with five features selected from four networks. The combination of network and fuzzy learning can highly improve the efficiency of AD's EEG identification.

Keywords: Alzheimer's disease, EEG, TSK fuzzy model, weighted visibility graph, feature select, multiple network

INTRODUCTION

Currently, Alzheimer's disease (AD) is becoming a common and serious disease due to organic neurodegenerative and progressive lesions in the brain. The patients always show some typical clinical presentations, particularly in the aspect of cognitive dysfunction such as deficient episodic memory and disabled remembering (Smailovic et al., 2018). The clinical diagnosis of AD currently adopts scale assessment, such as Mini-mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), activities of daily living (ADL), and physiological detection of cerebrospinal fluid. Patients with severe AD can be observed to have changes in brain structure, such as encephalatrophy, through brain functional imaging. Yang et al. applied magnetic resonance imaging (MRI) to detect the cerebral changes of blood flow and oxygenation in AD and mild cognitive impairment (MCI) subjects, and showed its

powerful ability to distinguish from normal controls (Yang et al., 2010). Hiroshi's study has demonstrated progression of atrophy mapping upstream to Braak's stages of neurofibrillary tangle deposition in AD. The main cause of organic brain lesions in AD is considered to be the loss of neurons and synapses (Brenner et al., 1988). It has been suggested that the loss of both synapses and neural pathways leads to a decrease in brain functional connectivity and influences electrical signals of the brain, so it is feasible to diagnose neurotic disease by electroencephalogram (EEG). EEG, which can measure the brain's voltage fluctuations with high temporal resolution, contains plenty of physiological information, and there is growing evidence that EEG may contribute to early recognition of AD patients.

The conventional EEG visual inspection is one of methods widely used in neurological assessment. Numerous previous studies have reported the disappearance of alpha EEG activities, particularly in posterior brain regions, through unaided viewing (Matsuda, 2013; Wang et al., 2015; Horvath et al., 2018). It has also been reported that visual EEG scores of ADs show a strong correlation with dementia severity (Kowalski et al., 2001). In the study of de Waal et al. (de Waal et al., 2011), AD patients with early onset are more likely to show severe diffuse slowing characteristic than those with later onset, which is consistent with the clinical manifestations of AD. In addition, studies have quantified the complexity of electrophysiological activities and reported declined complexity of EEG in AD patients (Cao et al., 2016). The change on the AD brain is also reflected in the perturbations of EEG synchronization. As EEG signals are irregular and non-stationary complex signals, traditional visual inspection is not sufficient for AD EEG identification (Buzsaki and Draguhn, 2004; de Waal et al., 2011; Cao et al., 2016). To address this issue, complex network theory is introduced into AD diagnosis, which aims to describe human brain from a global perspective (Palop et al., 2006; Nimmrich et al., 2015; Cao et al., 2016; Gao et al., 2019).

Over the past few years, more and more researchers have begun to adopt the attractive idea of using complex network methods to characterize the dynamic features of complex systems (Zou et al., 2019). This novel approach is the thorough combination of two frontier research fields, analysis methods of non-linear time series (Hively et al., 2000; Costa et al., 2002, 2005) and complex networks theory (Brown et al., 2004; Boccaletti et al., 2006). Zhang et al. have constructed complex networks with strength of temporal correlation between time series and reported that the behavior information (chaotic or fractal) of time series directly correlate with the topological structures (Zhang and Small, 2006). As an effective tool to get insight into the brain function, the brain network analysis has been widely applied in AD research. The healthy brain was found to work with network properties such as small-worldness, hubness, and rich-clubs, while the AD brain operated with less optimal network topologies (Meunier et al., 2010; Blinowska and Kaminski, 2013; Martijn and van den Heuvel, 2013; Wang et al., 2014, 2016; Deng et al., 2015). Loss of small-world features (toward random network topology) can be observed in functional network constructed from EEG and functional magnetic resonance imaging data (Stam et al., 2007; He and Evans, 2010; Tahaei et al., 2012; Reid and Evans, 2013). Numerous EEG studies have consistently demonstrated

decreased functional connections in the higher frequency bands of AD patients compared to controls (Tijms et al., 2013; van Straaten et al., 2014).

Compared to other approaches of constructing complex networks through time sequence, visibility graph (VG) algorithms can better integrate the basic features of time series. Lacasa et al. and Liu et al. converted time series into graphs and extracted the topological features using graph theory methods (Lacasa et al., 2008; Liu et al., 2010). They pointed out that the irregularity of time sequence can be characterized by the network topology. For instance, the periodic sequence can be transformed into regular lattice, while the chaotic series corresponds to random graphs. Subsequent researches began to introduce VG method into the EEG study of neurological diseases, and found features extracted from VG networks can be effectively used as mathematical markers in neurodegenerative diagnosis. VG algorithm was first applied in related research in AD by Ahmadlou et al. They reported that complexity of EEGs computed by VGs can be used in the distinguishing between AD and control EEGs (Ahmadlou et al., 2010).

The VG can only express the existence of edges between different time nodes, but not the strength of the edges. Therefore, Supriya et al. have proposed to combine the weighted edge with the horizontal visibility graph, which are not applicable to all complex network graphs (Supriya et al., 2016). Addressing the limitations of above approaches, Zhu et al. have improved the weighted visibility graph (WVG) algorithm by specifying radian function as the criterion for calculating edge weights in all kinds of complex network, and obtained promising results in the detection of epilepsy (Zhu et al., 2014). Also, studies have shown that the visualization method can preserve the characteristics like reduction of complexity (Polikar et al., 2007; Czigler et al., 2008) and slowing of rhythm (Dauwels et al., 2011; Cao et al., 2015; McBride et al., 2015) in patients with AD. WVG networks retain more structural information of the time series, which is more conducive for AD identification, compared to connectivity networks. Therefore, we apply the WVG method on the feature extraction of Alzheimer's disease. A variety of different parameters are extracted from the visibility graph, and used to further investigate which parameter can be used for diagnosing AD.

After quantitative analysis of complex WVG networks, the valuable information about the time series has been extracted. The machine learning generally approaches the extracted features for training the model and then applies them in signal detection. Traditional machine learning methods, including decision tree, random forests, k-nearest neighbor (KNN), Naive Bayes (NB), logistic regression, and so on (Siegelmann and Holzman, 2010; Hramov et al., 2019), have been widely used in the detection of neurological diseases. However, for systems with highly non-linear characteristics, models that built based on these methods do not characterize real models and be utilized in classification well. With this consideration, a rule-based fuzzy model is proposed and has been widely used in many fields like computer vision, natural language processing, and enhanced learning, achieving remarkable results (Gu et al., 2017). Takagi-Sugeno-Kang (TSK) method is proposed to build a model established by using fuzzy mathematics language

to describe some characteristics and internal relations of fuzzy phenomena. Compared with traditional classifiers that lack transparency, TSK can be used in multiple features classification and shows a superior model interpretability, which is defined as the ability to better understand the decision strategies of response functions in a human-interpretable manner in order to interpret internal relationships (Deng et al., 2018). In current applications of machine learning, such interpretability has received wide attention and is considered to be crucial.

In this paper, multiple networks are constructed based on multi-channel EEG, with each EEG channel able to be transformed into one-layer network. Then a number of different network features are extracted from them, which is too much for input feature vectors. In order to explore this problem, some feature selection approaches will be utilized to choose features, and the influence of different screening methods on the final classification results will also be tried. The parameters will be divided into three groups—single parameter from multi-network, multi-parameter from single network, and multi-parameter from multi-network—to observe the difference between the classification results of fuzzy models trained with different types of features. The structure of rest paper is as follows: section Methods and materials is devoted to describing the experimental design, including data collection and subject condition. Meanwhile, the principle of mathematical graph methods and Takagi-Sugeno-Kang (TSK) model adopted in paper are also explained in this part. In section Experimental Results, we performed a statistical analysis of the results and implemented AD recognition based on the proposed framework. Section Conclusion and Discussion includes a discussion of the application and advantages of the proposed model, as well as future work.

MATERIALS AND METHODS

Subjects and EEG Recordings

EEG recordings are collected from AD subjects and control subjects, respectively. The AD group included 30 confirmed AD patients who are diagnosed with mini-mental state examination (MMSE) scores are between 12 and 15. The diagnosis results meet the National Institute on Aging-Alzheimer's Association criteria. All of them have not used antipsychotic drugs, antidepressants, dopamine blockers, or excessive amounts of alcohol, and don't have other neurological or psychiatric disorders or any other serious illness. The AD group includes 18 females and 12 males, whose ages range from 74 to 78. The control group consisted of 30 healthy subjects of matched ages, ranging from 70 to 76 years old, and includes 10 females and 20 males. The MMSE scores of them are between 28 and 30. In order to avoid the impact on EEG activity, all subjects will be prohibited from using neuroactive drugs before the experiment. The data adopted in this paper is from our previous study (Wang et al., 2016), which is approved by the Ethics Committee of Tangshan Gongren Hospital and was conducted in accordance with the Declaration

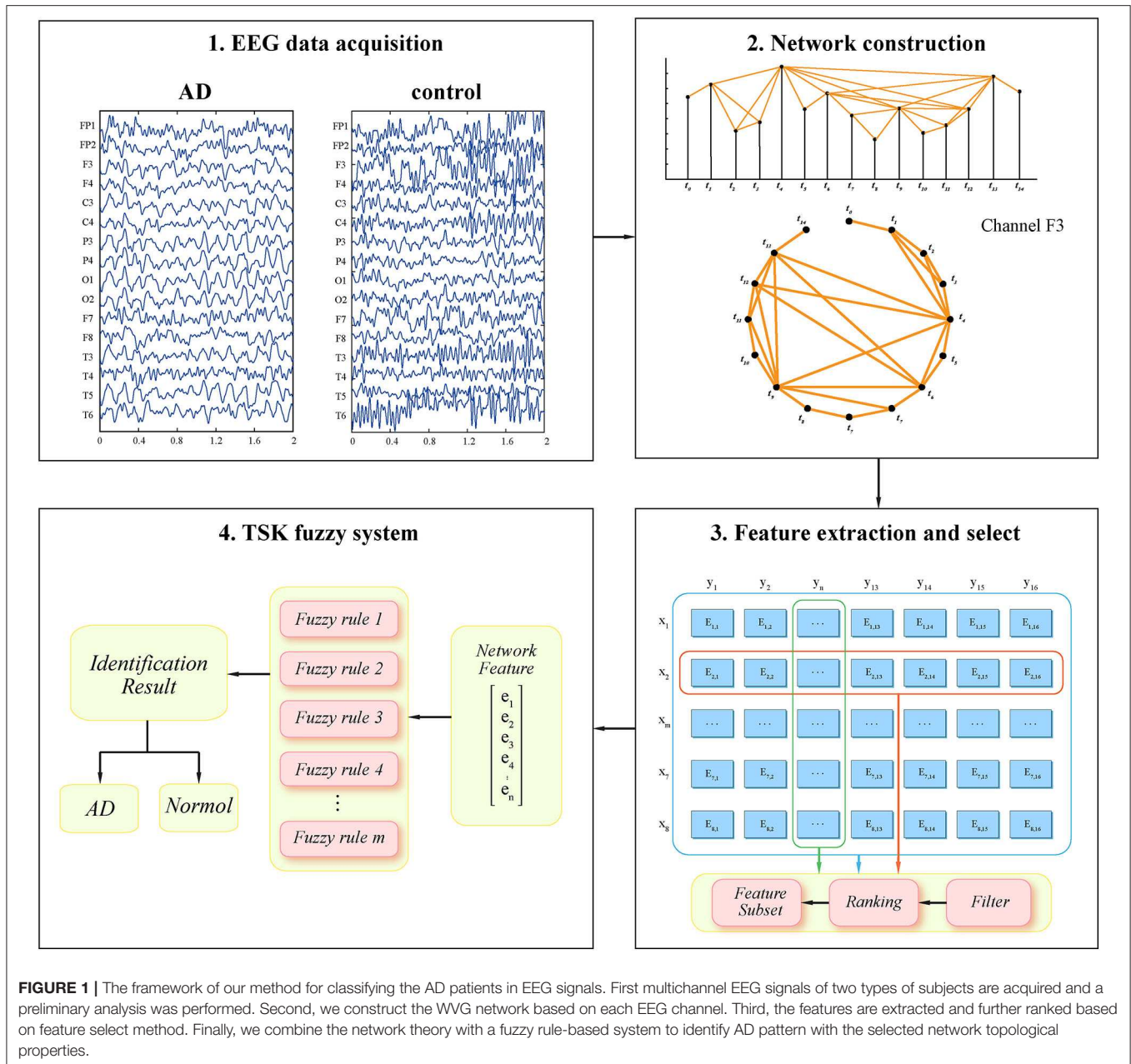of Helsinki. In addition, all the subjects in this experiment gave informed consent.

A 16-channel EEG monitoring system (Solar2000B) is adopted. The EEG channels have 10 MΩ input impedance with bandwidth as 0.08–300 Hz. In order to obtain low-frequency signals that meet the analysis requirements, the low-pass filtering range is set to 0.08–50 Hz. Studies have demonstrated that the EEG amplitude across different bands tends to stabilize when the scalp-electrode impedance is <10 kΩ, so electrode impedance in our experiments is set to 3 kΩ. The international 10–20 system, which consists of 16 electrodes, is adopted as electrode distribution in the scalp (surface) EEG recordings, and the linked earlobe A1 and A2 are used as reference. EEGs are recorded by Symtop amplifier (model: UEA–B; frequency: 1,024 Hz; electrode impedance: 3 k).

During the experiment, the subjects stayed in a semi-dark quiet room and were told to keep awake with eyes closed. The EEG recording process was kept to at least 30 min for each subject. In order to eliminate the impact of nervousness, anxiety, and head movement, a 10-min EEG is selected from each recorded EEG epoch. Sharp transient artifacts caused by eye movement and muscle artifacts, as well as segments with voltage exceeding 150uV, are also removed. Next, fifteen epochs without artifacts with an 8-s long duration for each (15 * 8 s = 120 s) were chosen for each subject's EEG, which are suitable for weighted visibility graph construction.

WVG Methods

The EEG signal is the electrical signal of the brain neurons measured on the surface of the cerebral cortex or scalp. It has obvious non-stationary, non-linear, and dynamic characteristics. The VG method provides a way to research the underlying dynamics of EEG data (Lacasa et al., 2008; Deng et al., 2018). Since the VG can inherit the dynamic nature of creating time series data, this technique has the characteristics of describing time series from the perspective of graph theory. The VG algorithm was originally applied in the field of robot motion planning, architectural design, and topographic descriptions of geographical space (Lozano-Pérez and Wesley, 1979; Turner et al., 2001; Lacasa et al., 2009; Jiang et al., 2017; Zou et al., 2019). This algorithm combined the mutual visible relationship of the point and obstacles in the two-dimensional landscape with the computational geometry framework. The literature study reveals that WVG can also be used in EEG data analysis to convert non-stationary, one-dimensional time series into two-dimensional viewable views for analysis. Different channels of EEG signals can reflect the electrophysiological information from different regions of the brain, so each single channel can obtain single complex network and multi-layer networks can be obtained through multi-channel EEG. The schematic diagram of constructing brain network by WVG method is shown in **Figure 1**.

In the construction of a WVG from a univariate EEG data $\{x_i\}_{i=1}^{N}$ with $x_i = x(t_i)$, individual observations are considered as vertices. Thus, the weighted adjacency matrix **W** can be obtained

FIGURE 1 | The framework of our method for classifying the AD patients in EEG signals. First multichannel EEG signals of two types of subjects are acquired and a preliminary analysis was performed. Second, we construct the WVG network based on each EEG channel. Third, the features are extracted and further ranked based on feature select method. Finally, we combine the network theory with a fuzzy rule-based system to identify AD pattern with the selected network topological properties.

with size of $N \times N$. Nodes of WVG network are defined by time points $\{t_i\}$, $i = 1, 2, ...N$ and each edge in this network is defined by the connection between two time points (Zou et al., 2019). Two nodes are defined as connected if the criterion

$$\frac{x(t_i) - x(t_k)}{t_k - t_i} > \frac{x(t_i) - x(t_j)}{t_k - t_i} \quad (1)$$

is fulfilled for all time points t_k with $t_i < t_k < t_j$. Then the absolute value of edge weight between two nodes are determined as

$$w_{i,j} = \arctan \frac{x(t_i) - x(t_j)}{t_i - t_j}, i < j \quad (2)$$

Feature Extraction and Select

The topology of the network is quantified based on the multiple complex networks obtained with WVG method. In order to statistically analyze the characteristics of AD networks and control networks, we calculate the clustering coefficient, average weighted degree, graph index complexity, network entropy, degree distribution index, modularity, local efficiency, and average path length as eight different topological characteristics.

Clustering Coefficient

The clustering coefficient is a measure to quantify how tightly connected the neighbor is around a node (Rubinov and Sporns, 2010). For a network G with N nodes, the connectivity between

nodes i and j is $a_{i,j}$ ($a_{i,j}=1$ if the connection exists or $a_{i,j}=0$ if not), the weight of connection are $w_{i,j}$ ($w_{i,j} \in [0, 1]$). For a weighted network, the local clustering coefficient of node i is defined as:

$$C(i) = \frac{1}{s_i(K_i - 1)} \sum_{j,h \in G} \frac{(w_{i,j} + w_{i,h})}{2} a_{i,j} a_{i,h} a_{j,h} \qquad (3)$$

where s_i, the strength of the node i, is defined as:

$$s_i = \sum_{j \in G_i} w_{i,j} \qquad (4)$$

And G_i represents the nodes set of node i neighborhoods. Further define the clustering coefficient of the whole network as:

$$C = \frac{1}{N} \sum_{i \in G_i} C(i) \qquad (5)$$

Average Weighted Degree

Average Weighted Degree is an important parameter for distinguishing networks with different topologies. The average weighted degree of the network can be obtained through averaging weights of the incident links on all the nodes in the network (Supriya et al., 2016):

$$wd = \frac{1}{N} \sum_{i \in G_i} s_i \qquad (6)$$

where s_i is described above in function (4).

Graph Index Complexity

Kim et al. have introduced graph index complexity as a new feature into the diagnosis of patients with AD by quantifying the complexity of the image graph (Kim and Wilhelm, 2008; Wang et al., 2016). With the largest eigenvalue of the adjacency matrix of a graph with n nodes presented as $\lambda_{\max}$ (Blinowska and Kaminski, 2013). The graph index complexity is defined as follows:

$$c_{\lambda_{\max}} = 4c(1 - c) \qquad (7)$$

where

$$c = \frac{\lambda_{\max} - 2\cos(\pi/(n+1))}{n - 1 - 2\cos(\pi/(n+1))} \qquad (8)$$

Degree Distribution Index

The degree distribution $P_{\deg}(k)$ is often used to classify complex networks, which can be formed by counting how many nodes have each degree. In this paper, a probability distribution object is obtained by fitting the Poisson distribution to the degree distribution vector. The degree distribution $P_{\deg}(k)$ is defined as

$$P_{\deg}(k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad (9)$$

The degree distribution index is defined as the λ values of the fitting distribution (Stephen and Toubia, 2009).

Network Entropy

The network entropy can be computed straightfordly based on the degree distribution as

$$S = -\sum_k P_{\deg}(k) \log P_{\deg}(k) \qquad (10)$$

Modularity

Modularity is a quality feature that can measure the quality of the clusters (communities), which are obtained by dividing the network partition (Supriya et al., 2016). The modularity Q of this weighted network is defined as:

$$Q = \frac{1}{2m} \sum_{i,j} \left(a_{i,j} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_i) \qquad (11)$$

where $m = \frac{1}{2} \sum_{i,j \in G} w_{i,j}$ is the sum weights of all links in the network, $k_i = \sum_{j \in G} w_{i,j}$ is the sum weight of the links attached to node i, C_i represents the community which vertex i is assigned to, the function $\delta(C_i, C_j)$ is 1 if nodes i and j belong to the same community and 0 otherwise. In this paper, we used the Louvain method (Blondel et al., 2008) to distribute nodes into different communities. This method is divided into two steps. In the first step, each node is added into the neighbor communities to determine the one which can maximize the modularity gain ΔQ. In second step, a new network is reconstructed whose node is defined as the small community found in the first step, and whose weights of new links are given by the sum weight of the links between nodes in the corresponding two old communities. Those two steps will be repeated iteratively until the maximum of modularity is accomplished and there is no more movement of nodes. The modularity gain ΔQ is defined as (Zhaohong et al., 2013):
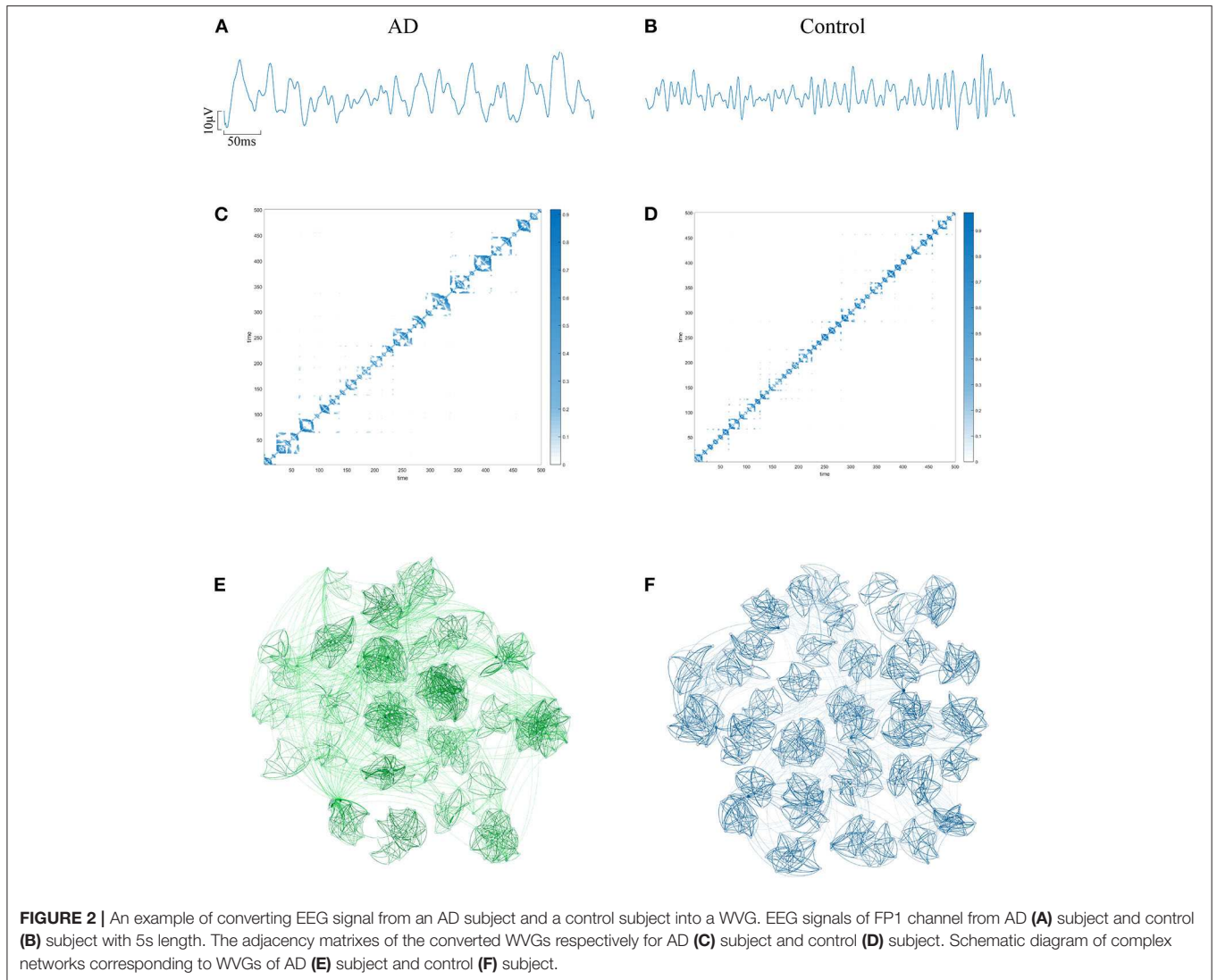
$$\Delta Q = \left[\frac{\Sigma_{in} + k_{i,in}}{2m} - \left(\frac{\Sigma_{tot} + k_i}{2m} \right)^2 \right]$$
$$- \left[\frac{\Sigma_{in}}{2m} - \left(\frac{\Sigma_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \qquad (12)$$

where Σ_{in} represents the sum of all the links weights inside community C, Σ_{tot} is the sum of the weights of the links attached to nodes in C, k_i is the sum of the weights of the links attached to node i, $k_{i,in}$ is the sum of the weights of the links from i to nodes in C, and m is the sum weights of all links in the network.

Local Efficiency

Local efficiency, as a node-specific measure, is defined to measure the density of the subnetwork composed of the neighborhood of the node i. Local efficiency of ith node is given as

$$E_{loc}(i) = \frac{1}{N_{G_i}(N_{G_i} - 1)} \sum_{i,j \in G, i \neq j} l_{i,j} \qquad (13)$$

FIGURE 2 | An example of converting EEG signal from an AD subject and a control subject into a WVG. EEG signals of FP1 channel from AD **(A)** subject and control **(B)** subject with 5s length. The adjacency matrixes of the converted WVGs respectively for AD **(C)** subject and control **(D)** subject. Schematic diagram of complex networks corresponding to WVGs of AD **(E)** subject and control **(F)** subject.

Where $l_{i,j}$ is the shortest distance between i and j, and N_{G_i} is the number of the neighborhood of node i. Local network efficiency is the average of the local efficiency of all nodes

$$E_{loc} = \frac{1}{N} \sum_i E_{loc}(i) \qquad (14)$$

Average Path Length

Average path length is a vital index to measure information transmission ability of networks. It can be used to evaluate the connectivity of the global functional network, including local and remote connection. The average path L is defined as:

$$L = \frac{1}{N(N-1)} \sum_{i,j,i \neq j} l_{i,j} \qquad (15)$$

TSK Fuzzy Model

Given an original input dataset $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\} \in \mathbf{R}^d$ and the corresponding class label $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n\}$ ($y_{i,j} = 1$ when the ith sample belongs to jth class; otherwise, $y_{i,j} = 0$), the kth fuzzy inference rules are often defined as

$$R^k : \text{IF } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \ldots \wedge x_d \text{ is } A_d^k, \text{THEN}$$
$$f_k(\mathbf{x}) = \beta_0^k + \beta_1^k x_1 + ... + \beta_d^k x_d, \text{k} = 1, ..., \text{K}$$

Where $\mathbf{x} = [x_1, x_2, ..., x_d]^T$ is input vector of each rule, K is the number of fuzzy rules, A_i^k are Gaussian antecedent fuzzy sets subscribed by the input variable x_i of Rule k, $\wedge$ is a fuzzy conjunction operator, $f_k(\mathbf{x})$ is a linear function about the inputs, and β_i^k are linear parameters.

With each rule is premised on the sample vector $\mathbf{x}$, the output of a TSK fuzzy system is expressed as
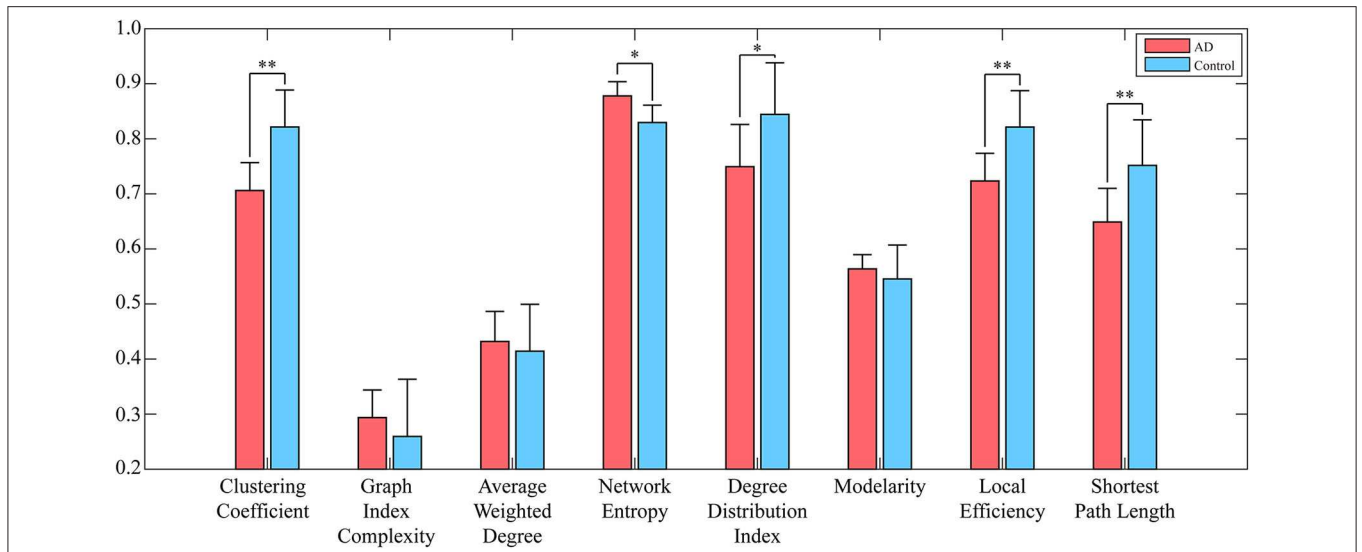
FIGURE 3 | Network parameters (averaged across subjects) of both AD networks and control networks. Error bars represent standard error across subjects. The degree of significant difference is calculated by Analysis of Variance (ANOVA) across all subjects. **A significant correlation ($p_c \leq 0.01$ corrected for multiple comparisons across tiles). *A trend ($p_c \leq 0.05$).

$$\tilde{y} = \sum_{k=1}^{K} \frac{\mu_k(\mathbf{x}) f_k(\mathbf{x})}{\sum_{k'=1}^{K} \mu_{k'}(\mathbf{x})} = \sum_{k=1}^{K} \tilde{\mu}_k(\mathbf{x}) f_k(\mathbf{x}) \qquad (16)$$

where

$$\mu_k(\mathbf{x}) = \prod_{i=1}^{d} \mu_{A_i^k}(x_i) \qquad (17)$$

is the fuzzy membership function and

$$\tilde{\mu}_k(\mathbf{x}) = \frac{\mu_k(\mathbf{x})}{\sum_{k'=1}^{K} \mu_{k'}(\mathbf{x})} \qquad (18)$$

is the normalized fuzzy membership function of the antecedent parameters of the kth fuzzy rule. While $\mu_{A_i^k}(x_i)$ is Gaussian membership function for fuzzy set A_i^k that can be expressed as

$$\mu_{A_i^k}(x_i) = \exp\left(-\frac{\left(x_i - c_i^k\right)^2}{\delta_i^k}\right) \qquad (19)$$

where c_i^k is kth cluster center parameters, which can be calculated with the classical fuzzy c-means (FCM) clustering algorithm (Bezdek et al., 1984):

$$c_i^k = \frac{\sum_{j=1}^{N} u_{jk} x_{ji}}{\sum_{j=1}^{N} u_{jk}} \qquad (20)$$

and the width parameter δ_i^k can be estimated by (Zhaohong et al., 2013):

$$\delta_i^k = \frac{h \cdot \sum_{j=1}^{N} u_{jk}\left(x_{ji} - c_i^k\right)^2}{\sum_{j=1}^{N} u_{jk}} \qquad (21)$$

where the element $u_{jk} \in [0, 1]$ denotes the fuzzy membership of nth input sample $\mathbf{x}_n$ to the kth cluster ($k = 1, 2, ..., K$), h is a constant called the scale parameter.

For an input sample $\mathbf{x}_n$, let

$$\mathbf{x}_{n,e} = \left(1, \mathbf{x}_n^T\right)^T \qquad (22)$$

$$\tilde{x}_n^k = \tilde{\mu}^k(\mathbf{x}_n) \mathbf{x}_e \qquad (23)$$

$$\rho(\mathbf{x}_n) = \left(\left(\tilde{x}_n^1\right)^T, \left(\tilde{x}_n^2\right)^T, ..., \left(\tilde{x}_n^K\right)^T\right)^T \in R^{K(d+1)} \qquad (24)$$

TABLE 1 | Classification results with each single parameter from single network layer is taken as input feature.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
y_1	0.6218	0.5546	0.5618	0.6543	0.5654	0.5457	0.5721	0.5904
y_2	0.5925	0.6550	0.5086	0.6368	0.5275	0.5364	0.5721	0.6893
y_3	0.6225	0.5482	0.5454	0.6454	0.6214	0.5596	0.5996	0.5243
y_4	0.6718	0.4779	0.5514	0.5632	0.6464	0.5229	0.5800	0.6146
y_5	0.5754	0.5918	0.5507	0.6343	0.5443	0.5382	0.5868	0.5982
y_6	0.6600	0.5718	0.5582	0.5486	0.5536	0.5625	0.5675	0.6386
y_7	0.5721	0.5793	0.5086	0.7893	0.6211	0.5779	0.5964	0.5293
y_8	0.5461	0.5550	0.6279	0.5957	0.6207	0.5679	0.5818	0.5679
y_9	0.7925	0.6671	0.6446	0.54	0.5454	0.7446	0.7182	0.5768
y_{10}	0.5271	0.3804	0.5811	0.6625	0.5504	0.5693	0.5225	0.5479
y_{11}	0.5157	0.425	0.5304	0.5107	0.5257	0.5275	0.5179	0.5075
y_{12}	0.6996	0.6832	0.6193	0.7432	0.6789	0.7904	0.745	0.7386
y_{13}	**0.7996**	0.6411	0.5664	0.5586	0.6325	0.7218	0.69	0.5343
y_{14}	0.5279	0.515	0.5082	0.5218	0.5314	0.5464	0.5379	0.5107
y_{15}	0.5625	0.4011	0.5071	0.5529	0.365	0.6868	0.3814	0.3789
y_{16}	0.7489	0.5479	0.5082	0.5486	0.54	0.6979	0.6175	0.4589

$$\beta^k = \left(\beta_0^k, \beta_1^k, ..., \beta_d^k \right)^T \tag{25}$$

$$\beta_g = \left(\left(\beta^1 \right)^T, \left(\beta^2 \right)^T, ..., \left(\beta^K \right)^T \right)^T \tag{26}$$

then the output value $\tilde{y}_n$ of a TSK fuzzy classifier for sample $\mathbf{x}_n$ can be expressed as

$$\tilde{y}_n = \beta_g^T \rho \left(\mathbf{x}_n \right) \tag{27}$$

Learning Algorithm

Given a training dataset $D_S = \left\{ \mathbf{x}_i, \mathbf{y}_i | \mathbf{x}_i \in R^d, \mathbf{y}_i \in R^C, i = 1, ..., N_S \right\}$, where C is the number of classes, the consequent parameter β_g can be learned by using generalized hidden-mapping ridge regression (GHRR) (Deng et al., 2014; Tian et al., 2019). The objective function is:
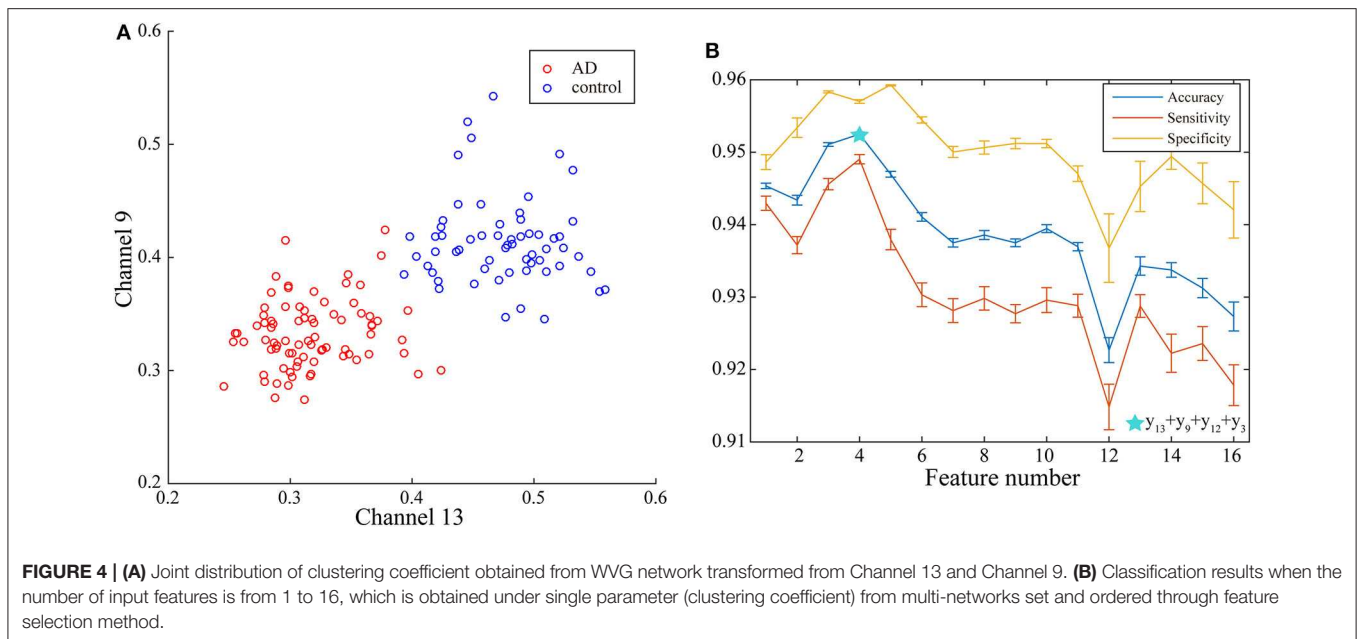
$$\min_{\beta_g} J \left(\beta_g \right) = \frac{1}{2} \sum_{j=1}^{C} \sum_{i=1}^{N_S} \left\| \beta_{g,j}^T \mathbf{x}_{g,i} - y_{i,j} \right\|^2 + \frac{\lambda}{2} \sum_{j=1}^{C} \beta_{g,j}^T \beta_{g,j} \tag{28}$$

where is the consequent parameter vector of the jth class is represented by $\beta_{g,j}$, λ is a regularization parameter controls the complexity of the classifier, and the tolerance of error λ can be set manually or determined by cross-validation. The optimal consequent parameters, $\beta_{g,j}$, can be computed by setting the derivatives of J with regard to each $\beta_{g,j}$ is 0 and the solution is (Yu et al., 2020):

TABLE 2 | Classification results with the set of single parameter from multiple networks is taken as input feature vector.

Method	Best input vector length	Accuracy	Sensitivity	Specificity
CFS	4	0.9211	0.9375	0.9091
DGUFS	3	0.7971	0.8974	0.6667
Fisher	4	0.9528	0.9491	0.9583
FSV	6	0.9430	0.9406	0.9449
LLCFS	9	0.9431	0.9223	0.9600
mRMR	6	0.8706	0.7971	0.9091

$$\beta_{g,j} = \left(\lambda_1 \mathbf{I}_{(d+1)*K \times (d+1)*K} + \sum_{i=1}^{N_S} \mathbf{x}_{g,i} \mathbf{x}_{g,i}^T \right)^{-1} \sum_{i=1}^{N_S} \mathbf{x}_{g,i} y_{i,j} \tag{29}$$

EXPERIMENTAL RESULTS

The EEG of AD patients implies a large amount of information that cannot be visually expressed from the waveform. Research shows that the visualization algorithm can express the hidden information in the form of images. In order to verify whether the AD brain's electrical features can be represented by WVG, we first select the same channel EEG from an AD patient and a control subject. Two episodes with a length of 500 data points (as shown in **Figures 2A,B**) are further intercepted, and converted to WVG. The result is shown in **Figures 2C,D**. Studies have reported that it's easy to detect a diffuse slowing in the EEG of AD patients with the naked eye (Micanovic and Pal, 2014). This diffuse slowing feature is well-preserved in WVG, and WVG of AD patients can be clearly observed in more communities, indicating the feasibility of WVG method for AD detection. For further observation of the topological feature of the WVG network, the two adjacency matrixes are represented as network structure diagrams that shown in **Figures 2E,F**. The dots in figure represent all network nodes and the network edges are represented by curves, and the shade of the curve color can directly reflect the weighted value of the edges. It can be observed that the different communities in the WVG network of normal people are generally similar in size and the distributions of connections are uniform. The community structure of the networks obtained by the WVG method is more irregular for AD patients. Most nodes are concentrated in a small part of communities, and the connection between communities is also closer. The result indicates that the electrophysiological signals of AD brains are more unstable, with fluctuations that are stronger. Research on single channel reveals that the WVG network of AD and normal people are significantly difference. Next, we will transform all 16 channels into multi-networks ($\{y_n\}(1 \leq n \leq 16)$) and each layer of network can be obtained from each channel. We further considered which parameters are selected to quantify this difference.

To reduce the computing time and to retain as much information as possible, the EEG signal is divided into many

FIGURE 4 | (A) Joint distribution of clustering coefficient obtained from WVG network transformed from Channel 13 and Channel 9. **(B)** Classification results when the number of input features is from 1 to 16, which is obtained under single parameter (clustering coefficient) from multi-networks set and ordered through feature selection method.

episodes through sliding windows with lengths of 500 data points. Since the size of the converted WVG network is consistent with the length of EEG series, a series of adjacency matrixes of size 500×500 are finally obtained. Next, we calculate clustering coefficient (x_1), graph index complexity (x_2), average weighted degree (x_3), network entropy (x_4), degree distribution index (x_5), modularity (x_6), local efficiency (x_7), and average path length (x_8) of each WVG network of both AD and control. Above parameters can be obtained from each different network layers, which can be considered as different features. Since there is a considerable difference in the magnitude of the values of different parameters, the calculated result is normalized to 0~1. All windows of each person were further averaged, and then a statistical analysis was performed based on each person. As shown in **Figure 3**, parameters of all subjects are statistically analyzed and the parameters that are significantly different for AD group and control group are marked with ∗. The values of clustering coefficient, local efficiency, and shortest path length of the AD group are significantly lower than that of controls with $p < 0.01$. Meanwhile, the degree distribution entropy of AD group is higher than that of controls with $p < 0.05$ while the degree distribution lambda of AD group is lower than that of controls with $p < 0.05$. The obtained results demonstrate that network topological parameters can be used to detect AD.

Through statistical analysis, it's obvious that some of the above parameters can clearly distinguish AD from the control group. In order to further verify the effect of these parameters on AD recognition, these parameters will be used as input features of the training fuzzy classifier. In each training process, we randomly select 80% of the original data to form training datasets which can be used for ten-fold cross-validation (10-CV), with 90% (90% $\times$ 80%) utilized for model training and 10% (10% $\times$ 80%) for constructing a validation set. The above procedure is repeated 10 times to cover the entire training set and finally determine the optimal hyperparameters of the TSK model. The remaining

20% of all data is tested as the testing data with determined hyperparameters. For each different input feature or feature vector, the classification results (accuracy, sensitivity, specificity) are averaged after training for 50 times.

The construction of each WVG network is based on a single time series, so 16 WVGs are obtained from 16-channel EEG used in this paper. These WVG networks contain different electrophysiological information of neurons in different brain regions. However, in the existing studies, the parameters extracted from WVG networks constructed by different brain regions' EEG were usually regarded as the same class of features, so the differences between brain regions were ignored. Therefore, we consider the 16 WVG networks as different networks and combine them into a multi-layer network. In order to verify whether the underlying dynamic information of these network layers are different, the classification is first performed with a single feature as input. Each parameter extracted from each single network layer transformed from different channels is used as the single input feature for model training, and the classification results are shown in **Table 1** with optimal classification result is bolded. It can be observed that for the same network parameter extracted from different network layers, the classification results are significantly different. The difference in classification accuracy of the same parameter from different network can even reach 28.39% for average weighted degree ($\{(x_3, y_k)\}, k = 1, ..., 16$), indicating that the dynamic information that contained in EEG of different brain regions does have significant differences and parameters of different layers maybe independent from each other. This finding shows that the network characteristics of the multi-network composed of WVG network layers can be used as independent input features for the classifier.

The input feature vector consisting of multiple parameters is used for fuzzy system training. The classification will be performed based on the following three feature sets [as shown

TABLE 3 | Classification results with the set of multiple parameters from single network is taken as input feature vector.

Method	Best input vector length	Accuracy	Sensitivity	Specificity
CFS	3	0.9259	0.9434	0.9091
DGUFS	4	0.8734	0.8772	0.8696
Fisher	2	0.8969	0.9259	0.8696
FSV	3	0.9341	0.9379	0.9321
LLCFS	6	0.9009	0.9091	0.8929
mRMR	5	0.7937	0.7692	0.8197

in **Figure 1**(3)]: (1) Single parameter from multi-networks: When ensuring that the classifier input is the same parameter, select different network layers for parameter extraction and combination. (2) Multi-parameters from single network: In the case of one single network layer, different parameters are extracted and selected for combination as a classifier input. (3) Multi-parameters from multi-network: All parameters extracted from all network layers are used as different input features to the classifier. Then for each set, various feature select methods including Correlation-based Feature Selection (CFS) (Guyon et al., 2002), Dependence Guided Unsupervised Feature Selection (DGUFS) (Zhu et al., 2017), Fisher (Gu et al., 2012), Feature Selective Validation (FSV) (Bradley and Mangasarian, 1999), Locality-Constrained Linear Coding Feature Select (LLCFS) (Zeng and Cheung, 2011), and minimum-redundancy maximum-relevance (mRMR) (Peng et al., 2005) are used to sort the features to obtain the feature sequence for each set. According to the obtained feature sequence, select the different number of features in order (i.e., the first one feature, the first two features, the first three features...) to component the input vectors for the TSK model training process. In the feature select process (as shown in **Figure 1**), the methods of Feature Selection Library (FSLib) are adopted for determining feature input vectors of TSK. All the algorithms are implemented with MATLAB 2016b.

First, case 1 is described as an example, and the structure of TSK is also described in details in following. As the clustering coefficient ($\{(x_1, y_{13})\}$) reached a highest accuracy of 79.96% in **Table 1**, local efficiency from all network layers ($\{(x_1, y_k)\}, k = 1, ..., 16$) is adopted for feature selection and multi-input classification. The orders of the features are obtained by various sorting feature selection algorithms. After the ranking of network parameters, we choose input feature vectors with different lengths as inputs of TSK model and calculate classification results (accuracy, sensitivity, and specificity), respectively. The optimal length of input vectors and classification results are shown in **Table 2**. It can be observed that with different feature select methods, the length of the feature vectors with the optimal classification result is different. Besides, the sensitivity is higher than the specificity for the feature vectors filtrated by CFS and DGUFS methods, while the others are opposite. It shows that the change of the feature used for training will affect the properties of the trained model. As for the parameter set of clustering coefficients extracted from multiple networks, the Fisher method

can be used to achieve the optimal classification result. The classification process with Fisher method are further explored.
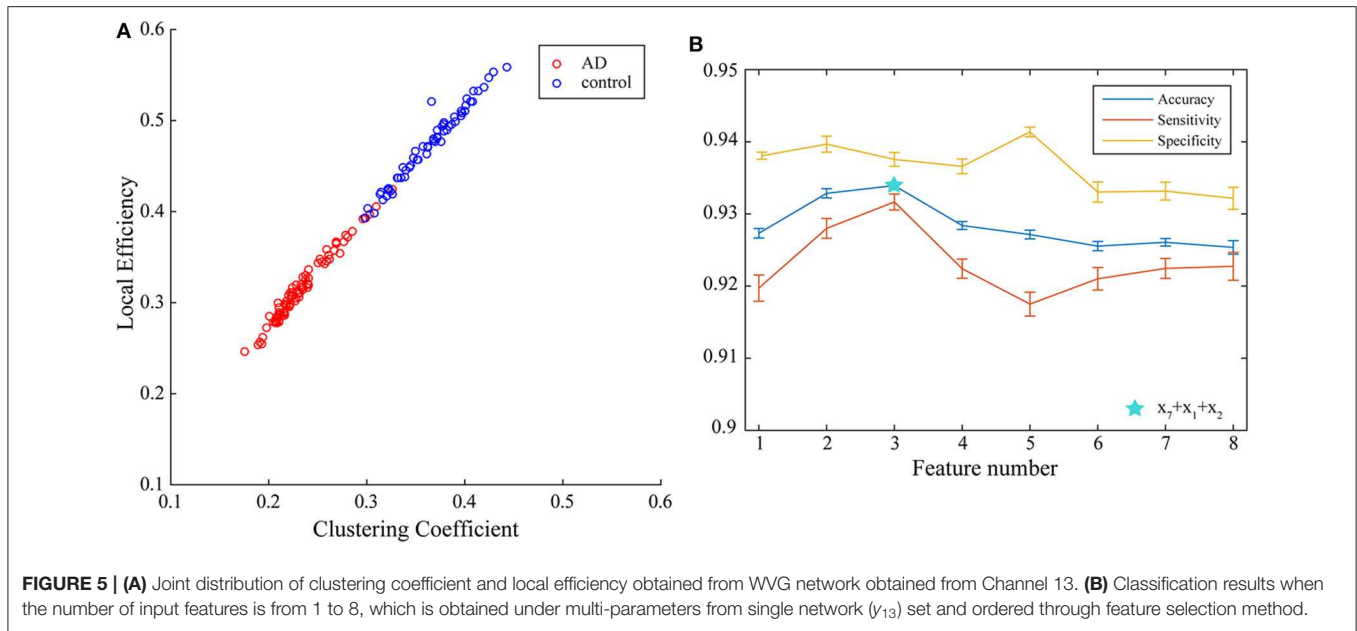
With the applying of Fisher algorithm, the order of the parameters is obtained as (x_1, y_{13}), (x_1, y_9), (x_1, y_{12}), (x_1, y_3), (x_1, y_1), (x_1, y_2), (x_1, y_1), (x_1, y_6), (x_1, y_{10}), (x_1, y_8), (x_1, y_5), (x_1, y_{15}), (x_1, y_7), (x_1, y_4), (x_1, y_{14}), (x_1, y_6). The joint distribution of the first two channels under the ranking is illustrated to verify the effectiveness of the same network parameter of WVG network transformed from different channels as the multi-input for classification. The result is shown in **Figure 4A** with each point represents a subject. It's obviously that AD subjects display significant differences from controls, which also demonstrate that local efficiencies, respectively, of channel 9 and channel 13 are effective to classify AD and controls. These two parameters can also get the best classification results when multi-network clustering coefficient is taken as single parameter input. However, the optimal parameters obtained by feature selection are not completely consistent with those that are optimal for the classification result when a single parameter is used as input. This indicates that the information of a single brain region cannot be used as a direct feature to distinguish patients with AD, but the implicit information of different brain regions can complement each other. In the above ranking order, five rules TSK classifiers are used with the number of classifier inputs is from 1 to 16 in order, and the final classification results under cross-validation are listed in **Figure 4B**. As the length of input feature vector increases, the accuracy reaches a maximum of 95.28% at four inputs and then begins to decrease.

In this part, the framework of the TSK is also described in details based on the selected optimal combination feature. The input vector **x** consists of the clustering coefficients of channel 13((x_1, y_{13})), channel 9((x_1, y_9)), channel 12((x_1, y_{12})), and channel 3((x_1, y_3)). Membership functions can be linguistically expressed using a fuzzy linguistic description including "*very low*," "*low*," "*medium*," "*high*," and "*very high*." Each membership function of different features corresponds to different description in ascending order of the values of centers. To provide further explanation, the clustering coefficient of channel 13 is interpreted as an example. We define the gaussian model as a membership function, and each rule will get a set of antecedent parameter (centers, standard variance), respectively, which are (0.3990 0.0031) for Rule 1, (0.3956 0.0030) for Rule 2, (0.4165 0.0032) for Rule 3, (0.4052 0.0031) for Rule 4, and (0.4040 0.0030) for Rule 5. By the permutation of these five centers of each rule, membership functions can be described with fuzzy linguistic description: Rule 1 is "*very low*," Rule 2 is "*very high*," Rule 3 is "*low*," Rule 4 is "*medium*," and Rule 5 is "*high*." The other four features can also be fuzzy and described similarly. Therefore, with the linguistic expressions and the corresponding linear function the fuzzy rule can be given as follows:

R^1 : IF y_{13} is ***very low*** $\wedge$ y_9 is ***very low*** $\wedge$ y_{12} is ***medium*** $\wedge$ y_3 is ***very low***,

THEN $f_1(\mathbf{x}) =$

$$\begin{bmatrix} 0.4975 - 0.1872y_{13} + 0.1615y_9 + 0.1515y_{12} + 0.2134y_3 \\ -0.1385 - 0.0959y_{13} - 0.0738y_9 - 0.0514y_{12} - 0.1147y_3 \end{bmatrix},$$

FIGURE 5 | (A) Joint distribution of clustering coefficient and local efficiency obtained from WVG network obtained from Channel 13. **(B)** Classification results when the number of input features is from 1 to 8, which is obtained under multi-parameters from single network (y_{13}) set and ordered through feature selection method.

R^2 : IF y_{13} is **very high** $\wedge$ y_9 is **very high** $\wedge$ y_{12} is **very low** $\wedge$ y_3 is **very high**,

THEN $f_2(\mathbf{x}) =$

$$\begin{bmatrix} -1.99\text{e-}4 + 1.59\text{e-}5y_{13} - 9.79\text{e-}5y_9 - 2.13\text{e-}5y_{12} + 2.03\text{e-}4y_3 \\ 0.0013 + 3.34\text{e-}4y_{13} + 4.12\text{e-}4y_9 + 2.65\text{e-}4y_{12} + 1.15\text{e-}4y_3 \end{bmatrix},$$

R^3 : IF y_{13} is **low** $\wedge$ y_9 is **low** $\wedge$ y_{12} is **high** $\wedge$ y_3 is **low**,

THEN $f_3(\mathbf{x}) =$

$$\begin{bmatrix} 0.2508 - 0.0485y_{13} + 0.0555y_9 + 0.0167y_{12} + 0.1049y_3 \\ -0.0039 - 0.0195y_{13} - 0.0095y_9 - 0.0572y_{12} - 0.0341y_3 \end{bmatrix},$$

R^4 : IF y_{13} is **medium** $\wedge$ y_9 is **high** $\wedge$ y_{12} is **very high** $\wedge$ y_3 is **medium**,

THEN $f_4(\mathbf{x}) =$

$$\begin{bmatrix} -0.0536 - 0.0429y_{13} - 0.0341y_9 - 0.0765y_{12} - 2.05\text{e-}4y_3 \\ 0.2071 + 0.0872y_{13} + 0.0759y_9 + 0.1244y_{12} + 0.0450y_3 \end{bmatrix},$$

R^5 : IF y_{13} is **high** $\wedge$ y_9 is **medium** $\wedge$ y_{12} is **low** $\wedge$ y_3 is **high**,
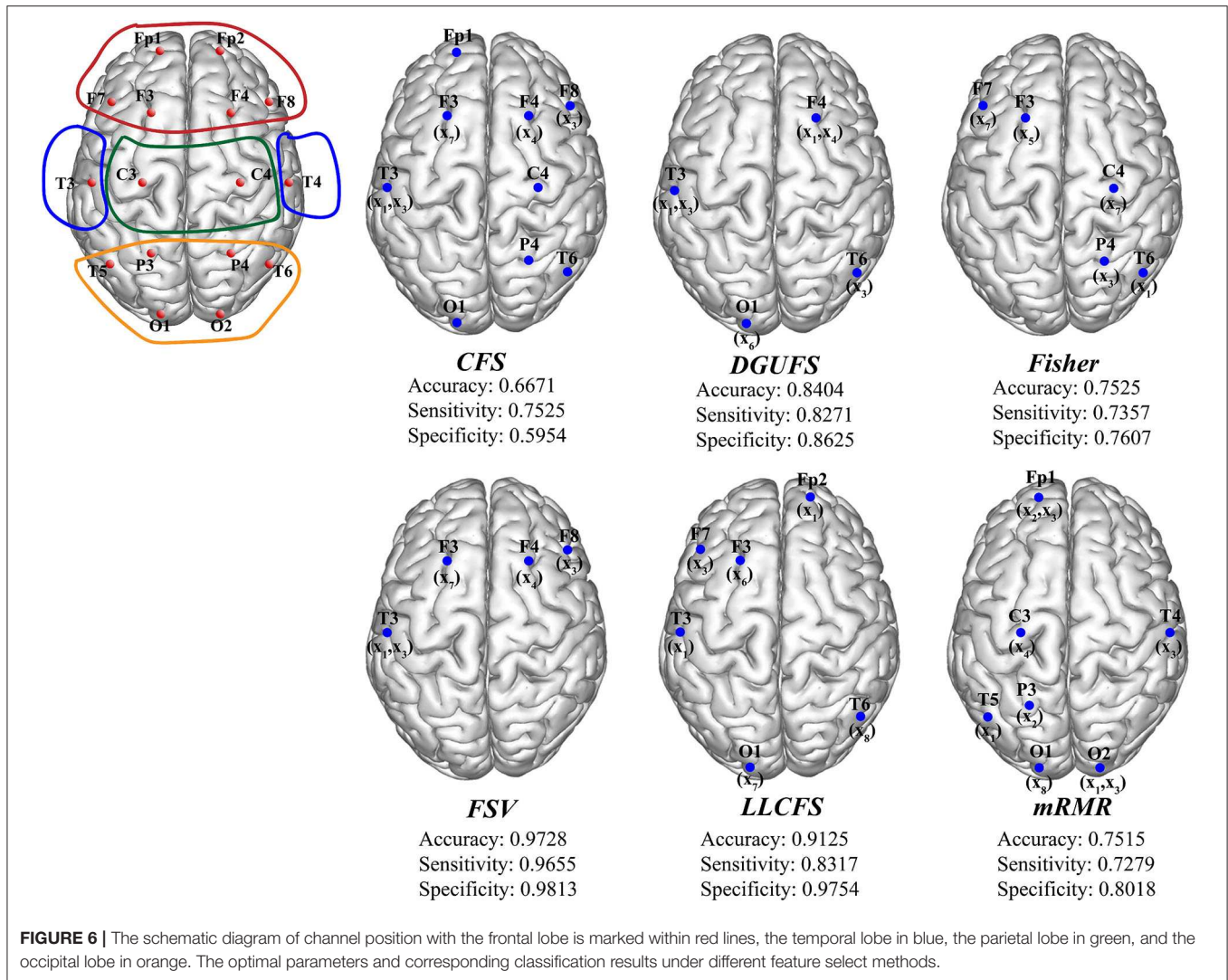
THEN $f_5(\mathbf{x}) =$

$$\begin{bmatrix} 0.0130 + 0.0016y_{13} - 0.0020y_9 + 0.0025y_{12} + 0.0041y_3 \\ 2.47\text{e-}4 + 0.0026y_{13} + 0.0017y_9 + 0.0011y_{12} + 4.33\text{e-}5y_3 \end{bmatrix}.$$

The fuzzy system that has been learned based on these five rules above, the example with an input of [0.2098 0.2106 0.3585 0.2264] is given to further explain the mechanism of testing process. Inputs of the identification process based on the trained fuzzy system are the network features of an AD patient, and the decision output is the prediction of label vector. The sum of the five calculated rule-based outputs is $f = [0.8940 \ 0.00956]^T$, then the maximal

element in f is set to 1 while others to 0 for handling the decision output. Finally, AD patient can be identified based on the final value of the output $y = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$.

Next, multi-parameters from single network are also used as input set for the classifier together. The classification results obtained by various feature select methods and the optimal lengths of input feature vectors are shown in **Table 3**. The parameters selected by FSV method can be used to form the vector to obtain the optimal classification result, and the sorted parameters are further analyzed in detail. The features in order obtained through the FSV algorithm is (x_7, y_{13}), (x_1, y_{13}), (x_2, y_{13}), (x_3, y_{13}), (x_8, y_{13}), (x_5, y_{13}), (x_6, y_{13}), (x_4, y_{13}). Clustering coefficients $((x_7, y_{13}))$ and local efficiencies $((x_1, y_{13}))$ are chosen to verify the feasibility of the classification, and the image is shown in **Figure 5A**. It is clear that there is a significant difference between the AD and the control group. The TSK classification is applied to all feature input groups. As shown in **Figure 5B**, the classification accuracy reaches a maximum value of 93.41% when the first three features are taken as input vector. The optimal combination obtained by the feature sorting method is local efficiency (x_7), clustering coefficient (x_1), and graph complexity index (x_2). The graph complexity index has a low discrimination between AD and the control group, and the TSK models trained with graph complexity index extracted from each network layer as single input have low classification accuracy. However, the image complexity index can supplement the clustering coefficient and local efficiency, indicating that the redundancy between some parameters from same network layer is small, which is of great significance as a feature of model training. Through the above classification results, multi-parameters, and multi-networks can both be applied to the TSK classification, and they are not the same type as input sets for model training.

FIGURE 6 | The schematic diagram of channel position with the frontal lobe is marked within red lines, the temporal lobe in blue, the parietal lobe in green, and the occipital lobe in orange. The optimal parameters and corresponding classification results under different feature select methods.

Finally, the multi-parameters from multi-networks are used for training. We further applied different feature select methods on this input set, and find the best feature input vectors, respectively. **Figure 6** provides the methods and corresponding classification results. The brain area enclosed by the red line is the frontal lobe, the blue is the temporal lobe, the green is the parietal lobe, and the orange is the occipital lobe. It can be observed that the parameters that are filtered by different methods are more common to be extracted from the network layers of the frontal EEG. This suggests that information in the frontal lobe is more effective in identifying AD patients. Damage to the frontal lobe of the brain, which plays a prominent role in thinking and behavior, can lead to forgetfulness, delayed behavior, and distraction. Meanwhile, signals from other brain regions also play an important role in AD recognition, indicating that AD disease has a global impact on the brain. The best result of multi-parameters from multi-networks set are selected through FSV method, which up to 97.28%. The

combination is $\{(x_7, y_3), (x_3, y_{13}), (x_1, y_{13}), (x_3, y_{12}), (x_4, y_4)\}$. The accuracy rate with set 3 is improved compared with set 1 and set 2, indicating that it is of certain significance to take multiple parameters extracted from multiple networks as different features.

CONCLUSION AND DISCUSSION

This paper proposes a multi-input machine learning method that combines fuzzy classifier and WVG to identify AD patient's EEG. In order to improve the interpretability and recognition accuracy of the model, complex network theory and TSK fuzzy system model is adopted. A WVG network layer is constructed using a single channel EEG. The multi-parameters obtained from multiple networks can be used as independent input features for model training, and the TSK model based on fuzzy rules is used to classify AD EEG with better interpretability.

We considered three types of classification input sets: multi-parameters from single network, single parameter from multi-networks, and multi-parameters from multi-networks. These three types of inputs are, respectively, applied as the training set of the learning of the TSK model. The experimental results show that the fuzzy model-based system model can achieve optimal performance with multi-parameters from multi-networks as classification input set, and the accuracy is up to 97.83%. Meanwhile, the optimal input numbers are different for the three types of input sets proposed in this paper. The best input combination is 5 input features in the input set of multi-parameters from multi-networks.

The current clinical techniques of AD identification, mainly including the scale assessing, cerebrospinal fluid examination, and the observation of atrophy of gray matter through the brain functional imaging, are difficult to obtain reliable diagnostic markers. It is also difficult to find obvious organic changes in the early stage of AD. We propose an AD diagnostic model that combines the TSK fuzzy model with complex network obtained by WVG method and propose three different kinds of training input sets, which provides a new method for the search of AD EEG biomarkers. Compared with traditional methods, the AD identification approach proposed in this paper, has lower implementation difficulty and higher accuracy.

EEG, which is commonly considered to have significant chaotic characteristics, cannot be well-evaluated with linear analysis. The WVG method used in this paper can transform the one-dimensional time series into images and extract the underlying information contained in electrophysiological activities of different brain regions. In contrast with other network construction methods like synchronous network, the WVG networks obtained by each EEG channel are independent of each other. Thus, more network features can be found and effective biomarkers can be obtained from kinds of feature sets with WVG (Zhu et al., 2014). The classification results show that this WVG method is very effective for feature extraction of AD recognition. In future works it will be combined with multi-layer network theory, further discussing the correlation between different channels with constructing multi-layer network. In past research we confirmed the feasibility of the multi-layer network scheme, and extracted the multiplex clustering coefficient and multiplex participation coefficient (Cai et al., 2020). Future work will consider both the implicit characteristics of single channels and the information integration between multiple channels.

We propose three different kinds of feature sets and prove that the optimal parameter vectors can be obtained from the set multi-parameters from multi-networks. This finding indicates that simultaneously considering different networks and different parameters as disparate features has obvious help for the acquisition of AD biomarkers. At the same time, the classification results show that the excessive features as input is not conducive to the optimization of the classification model, so it is necessary to reduce the feature dimension. Too much feature increase may lead to the overfitting of the learning model, and even the increase of invalid features may lead to the decrease of the accuracy based on test set (Guyon et al., 2002). Therefore, the application of feature selection plays an important role in improving the accuracy of fuzzy learning models.

In this paper, we combined the identification model combining feature selection approaches with machine learning. Researchers can effectively reduce the number of EEG channels, and the difficulty of data collection will be significantly reduced. Meanwhile, with the reduction of the parameters, it can be easier to improve the efficiency of the AD recognition process. Compared with traditional manual diagnosis, machine learning methods have higher reliability, and improved recognition accuracy. Especially, the TSK method has higher interpretability and robustness by integrating the advantages of fuzzy rules and membership functions. There are still some limitations in our research. We used a variety of feature selection methods, but a feature selection method suitable for the highly interpretable TSK model is necessary to be considered. Future work may focus on how to select features more efficiently and accurately to achieve higher classification accuracy.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics committee of Tangshan Gongren hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HY: article writing, design of methods, and article correction. LZ: article writing, processing and analysis of data, and design of methods. LC: design of methods and data analysis. JW: data analysis and article review. JL: data collection. RW: article review and correction. ZZ: data collection. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Ahmadlou, M., Adeli, H., and Adeli, A. (2010). New diagnostic EEG markers of the Alzheimer's disease using visibility graph. *J. Neural Transm.* 117, 1099–1109. doi: 10.1007/s00702-010-0450-3

Bezdek, J. C., Ehrlich, R., and Full, W. (1984). FCM: the fuzzy c-means clustering algorithm. *Comput. Geosci.* 10, 191–203. doi: 10.1016/0098-3004(84)90020-7

Blinowska, K. J., and Kaminski, M. (2013). Functional brain networks: random, "small world" or deterministic? *PLoS ONE* 8:e78763. doi: 10.1371/journal.pone.0078763

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech Theory Exp.* 10:P10008. doi: 10.1088/1742-5468/2008/10/p10008

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. (2006). Complex networks: Structure and dynamics. *Phys. Rep.* 424, 175–308. doi: 10.1016/j.physrep.2005.10.009

Bradley, P., and Mangasarian, O. (1999). "Feature selection via concave minimization and support vector machines," in *Machine Learning Proceedings of the Fifteenth International Conference* (San Francisco, CA).

Brenner, R. P., Reynolds, C. F., and Ulrich, R. F. (1988). Diagnostic efficacy of computerized spectral versus visual EEG analysis in elderly normal, demented and depressed subjects. *Electroencephalogr. Clin. Neurophysiol.* 69, 110–117. doi: 10.1016/0013-4694(88)90206-4

Brown, K. S., Hill, C. C., Calero, G. A., Myers, C. R., Lee, K. H., Sethna, J. P., et al. (2004). The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys. Biol.* 1, 184–195. doi: 10.1088/1478-3967/1/3/006

Buzsaki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745

Cai, L., Wei, X., Liu, J., Zhu, L., Wang, J., Deng, B., et al. (2020). Functional integration and segregation in multiplex brain networks for Alzheimer's disease. *Front. Neurosci.* 14:51. doi: 10.3389/fnins.2020.00051

Cao, L., Wang, J., Cao,Y., Deng, B., and Yang, C. (2016). "LPVG analysis of the EEG activity in Alzheimer's disease patients," in *2016 12th World Congress on Intelligent Control and Automation (WCICA)* (Guilin: IEEE). doi: 10.1109/WCICA.2016.7578491

Cao, Y., Cai, L., Wang, J., Wang, R., Yu, H., Cao, Y., et al. (2015). Characterization of complexity in the electroencephalograph activity of Alzheimer's disease based on fuzzy entropy. *Chaos* 25:083116. doi: 10.1063/1.4929148

Costa, M., Goldberger, A. L., and Peng, C. K. (2002). Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* 89:068102. doi: 10.1103/PhysRevLett.89.068102

Costa, M. D., Goldberger, A. L., and Peng, C. K. (2005). Broken asymmetry of the human heartbeat: loss of time irreversibility in aging and disease. *J. Electrocardiol.* 38, 1–206. doi: 10.1016/j.jelectrocard.2005.06.076

Czigler, B., Csikos, D., Hidasi, Z., Anna Gaal, Z., Csibri, E., Kiss, E., et al. (2008). Quantitative EEG in early Alzheimer's disease patients - power spectrum and complexity features. *Int. J. Psychophysiol.* 68, 75–80. doi: 10.1016/j.ijpsycho.2007.11.002

Dauwels, J., Srinivasan, K., Ramasubba Reddy, M., Musha, T., Vialatte, F. B., Latchoumane, C., et al. (2011). Slowing and loss of complexity in Alzheimer's EEG: two sides of the same coin? *Int. J. Alzheimers. Dis.* 2011:539621. doi: 10.4061/2011/539621

de Waal, H., Stam, C. J., Blankenstein, M. A., Pijnenburg, Y. A., Scheltens, P., and van der Flier, W. M. (2011). EEG abnormalities in early and late onset Alzheimer's disease: understanding heterogeneity. *J. Neurol. Neurosurg. Psychiatr.* 82, 67–71. doi: 10.1136/jnnp.2010.216432

Deng, B., Liang, L., Li, S., Wang, R., Yu, H., Wang, J., et al. (2015). Complexity extraction of electroencephalograms in Alzheimer's disease with weighted-permutation entropy. *Chaos* 25:043105. doi: 10.1063/1.4917013

Deng, Z., Choi, K. S., Jiang, Y., and Wang, S. (2014). Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods. *IEEE Trans. Cybern.* 44, 2585–2599. doi: 10.1109/TCYB.2014.2311014

Deng, Z., Xu, P., Xie, L., Choi, K. S., and Wang, S. (2018). Transductive joint-knowledge-transfer TSK FS for recognition of epileptic EEG signals. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 1481–1494. doi: 10.1109/TNSRE.2018.2850308

Gao, Z., Wang, X., Yang, Y., Mu, C., Cai, Q., Dang, W., et al. (2019). EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 2755–2763. doi: 10.1109/TNNLS.2018.2886414

Gu, Q., Li, Z., and Han, J. (2012). "Generalized fisher score for feature selection," in *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI'11)* (Arlington, VA: AUAI Press), 266–273.

Gu, X., Chung, F.-L., and Wang, S. (2017). Bayesian Takagi–Sugeno–Kang Fuzzy classifier. *IEEE Trans. Fuzzy Syst.* 25, 1655–1671. doi: 10.1109/tfuzz.2016.2617377

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. doi: 10.1023/a:1012487302797

He, Y., and Evans, A. (2010). Graph theoretical modeling of brain connectivity. *Curr. Opin. Neurol.* 23, 341–350. doi: 10.1097/WCO.0b013e32833aa567

Hively, L. M., Protopopescu, V. A., and Gailey, P. C. (2000). Timely detection of dynamical change in scalp EEG signals. *Chaos* 10, 864–875. doi: 10.1063/1.1312369

Horvath, A., Szucs, A., Csukly, G., Sakovics, A., Stefanics, G., and Kamondi, A. (2018). EEG and ERP biomarkers of Alzheimer's disease: a critical review. *Front. Biosci.* 23, 183–220. doi: 10.2741/4587

Hramov, A. E., Maksimenko, V., Koronovskii, A., Runnova, A. E., Zhuravlev, M., Pisarchik, A. N., et al. (2019). Percept-related EEG classification using machine learning approach and features of functional brain connectivity. *Chaos* 29:093110. doi: 10.1063/1.5113844

Jiang, W., Wei, B., Tang, Y., and Zhou, D. (2017). Ordered visibility graph average aggregation operator: An application in produced water management. *Chaos* 27:023117. doi: 10.1063/1.4977186

Kim, J., and Wilhelm, T. (2008). What is a complex graph? *Phys. A* 387, 2637–2652. doi: 10.1016/j.physa.2008.01.015

Kowalski, J. W., Gawel, M., Pfeffer, A., and Barcikowska, M. (2001). The diagnostic value of EEG in Alzheimer disease: correlation with the severity of mental impairment. *J. Clin. Neurophysiol.* 18, 570–575. doi: 10.1097/00004691-200111000-00008

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., and Nuno, J. C. (2008). From time series to complex networks: the visibility graph. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4972–4975. doi: 10.1073/pnas.0709247105

Lacasa, L., Luque, B., Luque, J., and Nuño, J. C. (2009). The visibility graph: a new method for estimating the Hurst exponent of fractional Brownian motion. *EPL* 86, 30001–30005. doi: 10.1209/0295-5075/86/30001

Liu, C., Zhou, W.-X., and Yuan, W.-K. (2010). Statistical properties of visibility graph of energy dissipation rates in three-dimensional fully developed turbulence. *Phys. Stat. Mech. Appl.* 389, 2675–2681. doi: 10.1016/j.physa.2010.02.043

Lozano-Pérez, T., and Wesley, M. A. (1979). An algorithm for planning collision-free paths among polyhedral obstacles. *Commun. ACM* 22, 560–570. doi: 10.1145/359156.359164

Martijn, P., and van den Heuvel, O. S. (2013). Network hubs in the human brain. *Trends Cogn. Sci.* 17, 683–696

Matsuda, H. (2013). Voxel-based morphometry of brain MRI in normal aging and Alzheimer's disease. *Aging Dis* 4, 29–37. doi: 10.1016/j.tics.2013.09.012

McBride, J. C., Zhao, X., Munro, N. B., Jicha, G. A., Schmitt, F. A., Kryscio, R. J., et al. (2015). Sugihara causality analysis of scalp EEG for detection of early Alzheimer's disease. *Neuroimage Clin.* 7, 258–265. doi: 10.1016/j.nicl.2014.12.005

Meunier, D., Lambiotte, R., and Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Front. Neurosci.* 4:200. doi: 10.3389/fnins.2010.00200

Micanovic, C., and Pal, S. (2014). The diagnostic utility of EEG in early-onset dementia: a systematic review of the literature with narrative analysis. *J. Neural Transm.* 121, 59–69. doi: 10.1007/s00702-013-1070-5

Nimmrich, V., Draguhn, A., and Axmacher, N. (2015). Neuronal network oscillations in neurodegenerative diseases. *Neuromolecular Med.* 17, 270–284. doi: 10.1007/s12017-015-8355-9

Palop, J. J., Chin, J., and Mucke, L. (2006). A network dysfunction perspective on neurodegenerative diseases. *Nature* 443, 768–773. doi: 10.1038/nature05289

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1226–1238. doi: 10.1109/TPAMI.2005.159

Polikar, R., Topalis, A., Green, D., Kounios, J., and Clark, C. M. (2007). Comparative multiresolution wavelet analysis of ERP spectral bands using an ensemble of classifiers approach for early diagnosis of Alzheimer's disease. *Comput. Biol. Med.* 37, 542–558. doi: 10.1016/j.compbiomed.2006.08.012

Reid, A. T., and Evans, A. C. (2013). Structural networks in Alzheimer's disease. *Eur. Neuropsychopharmacol.* 23, 63–77. doi: 10.1016/j.euroneuro.2012.11.010

Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003

Siegelmann, H. T., and Holzman, L. E. (2010). Neuronal integration of dynamic sources: Bayesian learning and Bayesian inference. *Chaos* 20, 037112. doi: 10.1063/1.3491237

Smailovic, U., Koenig, T., Kareholt, I., Andersson, T., Kramberger, M. G., Winblad, B., et al. (2018). Quantitative EEG power and synchronization correlate with Alzheimer's disease CSF biomarkers. *Neurobiol. Aging* 63, 88–95. doi: 10.1016/j.neurobiolaging.2017.11.005

Stam, C. J., Jones, B. F., Nolte, G., Breakspear, M., and Scheltens, P. (2007). Small-world networks and functional connectivity in Alzheimer's disease. *Cereb. Cortex* 17, 92–99. doi: 10.1093/cercor/bhj127

Stephen, A. T., and Toubia, O. (2009). Explaining the power-law degree distribution in a social commerce network. *Soc. Netw.* 31, 262–270. doi: 10.1016/j.socnet.2009.07.002

Supriya, S., Siuly, S., Wang, H., Cao, J., and Zhang, Y. (2016). Weighted visibility graph with complex network features in the detection of epilepsy. *IEEE Access* 4, 6554–6566. doi: 10.1109/access.2016.2612242

Tahaei, M. S., Jalili, M., and Knyazeva, M. G. (2012). Synchronizability of EEG-based functional networks in early Alzheimer's disease. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 636–641. doi: 10.1109/TNSRE.2012.2202127

Tian, X., Deng, Z., Ying, W., Choi, K. S., Wu, D., Qin, B., et al. (2019). Deep multi-view feature learning for EEG-based epileptic seizure detection. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1962–1972. doi: 10.1109/TNSRE.2019.2940485

Tijms, B. M., Wink, A. M., de Haan, W., van der Flier, W. M., Stam, C. J., Scheltens, P., et al. (2013). Alzheimer's disease: connecting findings from graph theoretical studies of brain networks. *Neurobiol. Aging* 34, 2023–2036. doi: 10.1016/j.neurobiolaging.2013.02.020

Turner, A., Doxa, M., O'Sullivan, D., and Penn, A. (2001). From Isovists to Visibility Graphs: A Methodology for the Analysis of Architectural Space. *Environ. Plan. B* 28, 103–121. doi: 10.1068/b2684

van Straaten, E. C., Scheltens, P., Gouw, A. A., and Stam, C. J. (2014). Eyes-closed task-free electroencephalography in clinical trials for Alzheimer's disease: an emerging method based upon brain dynamics. *Alzheimers. Res. Ther.* 6:86. doi: 10.1186/s13195-014-0086-x

Wang, J., Yang, C., Wang, R., Yu, H., Cao, Y., and Liu, J. (2016). Functional brain networks in Alzheimer's disease: EEG analysis based on limited penetrable visibility graph and phase space method. *Phys. A Stat. Mechan. Appl.* 460, 174–187. doi: 10.1016/j.physa.2016.05.012

Wang, R., Wang, J., Li, S., Yu, H., Deng, B., and Wei, X. (2015). Multiple feature extraction and classification of electroencephalograph signal for Alzheimers' with spectrum and bispectrum. *Chaos* 25:013110. doi: 10.1063/1.4906038

Wang, R., Wang, J., Yu, H., Wei, X., Yang, C., and Deng, B. (2014). Decreased coherence and functional connectivity of electroencephalograph in Alzheimer's disease. *Chaos* 24:033136. doi: 10.1063/1.4896095

Yang, W., Xia, H., Xia, B., Lui, L. M., and Huang, X. (2010). "ICA-based feature extraction and automatic classification of AD-related MRI data," in *Sixth International Conference on Natural Computation* (Yantai: IEEE).

Yu, H., Lei, X., Song, Z., Liu, C., and Wang, J. (2020). Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification. *IEEE Trans. Fuzzy Syst.* 28, 60–71. doi: 10.1109/tfuzz.2019.2903753

Zeng, H., and Cheung, Y. M. (2011). Feature selection and kernel learning for local learning-based clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 1532–1547. doi: 10.1109/TPAMI.2010.215

Zhang, J., and Small, M. (2006). Complex network from pseudoperiodic time series: topology versus dynamics. *Phys. Rev. Lett.* 96:238701. doi: 10.1103/PhysRevLett.96.238701

Zhaohong, D., Yizhang, J., Kup-Sze, C., Fu-Lai, C., and Shitong, W. (2013). Knowledge-leverage-based TSK Fuzzy System modeling. *IEEE Trans. Neural Netw. Learn. Syst.* 24, 1200–1212. doi: 10.1109/TNNLS.2013.2253617

Zhu, G., Li, Y., and Wen, P. P. (2014). Epileptic seizure detection in EEGs signals using a fast weighted horizontal visibility algorithm. *Comput. Methods Programs Biomed.* 115, 64–75. doi: 10.1016/j.cmpb.2014.04.001

Zhu, P., Zhu, W., Hu, Q., Zhang, C., and Zuo, W. (2017). Subspace clustering guided unsupervised feature selection. *Pattern Recognit.* 66, 364–374. doi: 10.1016/j.patcog.2017.01.016

Zou, Y., Donner, R. V., Marwan, N., Donges, J. F., and Kurths, J. (2019). Complex network approaches to nonlinear time series analysis. *Phys. Rep.* 787, 1–97. doi: 10.1016/j.physrep.2018.10.005