Feature Extraction and Identification of Alzheimer's Disease based on Latent Factor of Multi-Channel EEG

Kai Li, Jiang Wang, *Member, IEEE*, Shanshan Li, Haitao Yu, Lin Zhu, Jing Liu, and Lingyun Wu

***Abstract*— Alzheimer's disease is a neurodegenerative disease in old age, early diagnosis will help to delay the progression of the disease. Presently, the features of brain functional diseases can be obtained with EEG analysis, but the relationship between characteristics of EEG and Alzheimer's disease has not been clearly clarified. In this work, we hypothesize that there exist default brain variables (latent factors) across subjects in disease processes, decoding latent factor from brain activity contributes to the study of cognitive impairment. To that end, this work proposes to extract characteristics of Alzheimer's disease by combing latent factors of EEG with variational auto-encoder to realize disease identification. Primarily, power spectrum characteristics is investigated and it is found that the dominant frequency of two groups is different. Further analysis reveals that latent factor distribution of Alzheimer's disease exists obvious differences with normal group in the theta frequency band. Moreover, the latent factors are projected onto the three-dimensional state space and the transient rotation of neural state is found, which shows the dynamic characteristics of latent factors. In addition, Takagi-Sugeno-Kang classifier is adopted and multiple latent factors are fed into Takagi-Sugeno-Kang classifier for decoding. Compared with linear classifier, Takagi-Sugeno-Kang fuzzy classifier has better performance in classification of energy feature from sub-frequency bands of latent factors. The accuracy of identification could up to 98.10% when the combination of energy features of four frequency bands is used as model input. Collectively, this work provides a feasible tool for identification of neurological dysfunction from the view of latent factors, especially contributing to the diagnosis of Alzheimer's disease.**

***Index Terms*— Alzheimer's disease, EEG, latent factor, fuzzy system.**

I. Introduction

ALZHEIMER'S Disease (AD) is an irreversible neurological disease caused by the degenerative loss of neurons in the cerebral cortex, mainly manifested as memory loss, cognitive and language impairment [1]–[3]. To date, the development of disease can only be interfered with by drugs or verbal behavior training, but it cannot be completely cured [4]. Studies have shown that the progression of AD can be slowed, meanwhile the brain function would keep at high level if patients get treated at early stage [5], [6]. However, patients only show slightly cognitive decline and the misdiagnosis rate is high, which causes obstacles to the diagnosis and early intervention of AD [7]. Moreover, the pathological changes in the brain of people at high risk of disease may occur earlier than the clinical appearance of dementia symptoms [8]. Therefore, accurate diagnosis of AD patients combined with neuroimaging technology is the key to improve the therapeutic effect.

Human brain is composed of almost 100 billion neurons via synapses, which has complex network and stochastic fluctuations characteristics [9], [10]. Many kinds of brain signals have been obtained to analyze these features. Compared with other imaging techniques, EEG is a relatively cost-effective and non-invasive technique that provide high temporal resolution data on electrical activity in the brain during neurotransmission [11], [12]. There is increasing evidence that EEG signals may be effective in the differential diagnosis of neural features and cognitive impairment in early AD [13], [14]. In the study of 460 patients with AD, it is found that patients with early onset of AD show more severe diffuse moderating than patients with late onset of AD, which is consistent with the clinical features of AD [15]. Van *et al.* use several effective features extracted from EEG and event-related potential as biomarkers for the identification of AD [16]. However, due to the nature of EEG acquisition, EEG signals have high redundancy and different channels interfere with each other. At the same time, as a complex nonlinear system, the electrophysiological activity of brain has complex dynamic characteristics [17]–[19]. In this manner, an effective decoding method is urgently needed for the analysis of EEG.

The dimensionality reduction technique allows us to identify neural states as principal component modes of interneuron co-variables on the low dimensional surface [20], [21]. Common linear dimensionality reduction methods include Laplacian Eigenmaps, Principal Component Analysis *et al.* Kouropteva *et al.* point out that population dynamics does not focus on exploring the entire high-dimensional neural space, but on the low-dimensional level within the whole space [22].

Manuscript received March 24, 2021; revised June 21, 2021; accepted July 27, 2021. Date of publication July 30, 2021; date of current version August 9, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62071324 and in part by the Natural Science Foundation of Tianjin under Grant 19JCYBJC18800. *(Corresponding author: Haitao Yu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Ethics Committee of Tangshan Gongren Hospital, China, and performed in line with the Declaration of Helsinki.

Kai Li, Jiang Wang, Shanshan Li, Haitao Yu, and Lin Zhu are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: htyu@tju.edu.cn).

Jing Liu and Lingyun Wu are with the Department of Neurology, Tangshan Gongren Hospital, Tangshan 063000, China.

Digital Object Identifier 10.1109/TNSRE.2021.3101240

Chunmei *et al.* investigate a new method for the extraction and recognition of epileptic-like activity features in EEG [23]. In addition to the traditional dimensionality reduction method, Variational Autoencoder realizes information compression and reconstruction in the form of unsupervised learning [24], [25]. In this method, Variational Autoencoder is used to reduce the dimension of the original EEG, and approximate entropy algorithm is applied to sub-band signals at different wavelet scales to realize the identification of epileptic-like activity. Using the encoder to describe the probability distribution of each potential attribute can also realize the extraction of latent variables.

As an external manifestation of individual neural activity with high correlation and redundancy, EEG can also be regarded as a kind of coded signal formed by high dependence on latent factors [26]–[28]. The latent factors of this hypothesis can ignore the differences between individuals and interpret the higher functions of the brain, such as learning and memory, by summarizing the activity of a large number of neurons. Studies have shown that when performing a conscious motor task, the brain needs relatively few independent signals to control behavior, and required only a few of independent neural signals [29]. These neural signals are the latent factors to describe the dynamics of neural models. In addition, it has been suggested that majority of neurons participate in neural pattern during cognitive task in the brain, and the neural activity of a single neuron reflect the latent factors of neural patterns [30]–[32]. Therefore, in the absence of effective features in EEG, we propose decoding EEG signals in the form of acquiring latent factors and looking for features that can be used to distinguish AD. The state information of latent factors in EEG contains lots of relevant information to characterize the state of brain activity, which is helpful for the analysis and estimation of neurologic diseases with brain function. Therefore, inferring the latent factors of brain interaction from external EEG can help researchers investigate internal correlations.

By extracting latent factors, we can extract decoded information from highly redundant EEG signals and quantify more meaningful EEG features. Nowadays, the machine learning method is further applied in neuroscience research to solve complex classification problems for the characteristics of EEG, such as complex networks and latent factor [33]–[35]. It reduces the subjectivity that may occur during artificial diagnosis, and does not require relevant prior knowledge. Traditional learning methods, including decision tree, random forests, support vector machine, Naïve Bayes classification, and so on, have been extensively used in the diagnosis of mental diseases [36], [37]. However, for systems with highly nonlinear characteristics, these methods cannot be well modeled and classified. In this way, Takagi-Sugeno-Kang classifier has good performance interpretability in multiple features classification task [38], [39]. The interpretability of model is vital in many diagnosis applications. Rule-based fuzzy methods have been widely used in many filed, such as mental diagnosis, image processing and so on [34], [40]–[44]. The advantage of using fuzzy method in classification tasks is that the interpretability and accuracy of judgment can be synchronously taken in the resultant systems [43]. The interpretability of rule-based fuzzy classifiers is related to the comprehensibility of fuzzy partitions and the simplicity of rule-based fuzzy classifiers [44]. Therefore, the application of Takagi-Sugeno-Kang can better explain the brain system with complex nonlinear dynamic characteristics and obtain a better classification effect of disease characteristics.

The main purpose of this study is to extract the feature of AD and identify the neurological dysfunction between disease and normal groups from the view of latent factors. The remainder of this paper is arranged as follows: in Section II, the EEG experiments and decoding method are introduced; in Section IIIA, the time-frequency dynamical analysis between disease and normal groups are demonstrated; in Section IIIB, the dynamical analysis of latent factor in EEG is studied; in Section IIIC, classification of EEG latent factors investigated. Based on the Variational Autoencoder and Takagi-Sugeno-Kang classifier methods, the latent factors and feature extraction for EEG signals have important explanatory significance to the diagnose of Alzheimer's disease.

II. Mathematical Model and Methods

A. Experiment Design and EEG Recording

In this study, depending on the severity of the diseases, 40 subjects are enlisted and separated into 2 groups: (1) 20 AD patients (8 males and 12 females, age: 74-78 years old). (2) 20 age-similar normal groups (10 males and 10 females, age: 70-76 years old). To assess the brain states of two groups, we perform a clinical neuroimaging and neurological examination. None of the subjects has any neurological symptom, neuroactive drugs or other factors that might affect EEG activity. Our study is approved by the Ethics Committee of Tangshan Gongren hospital in China and the experiments are conducted in accordance with the Declaration of Helsinki. In addition, all subjects have obtained formal consent with full agreement of the target and procedure of the study.

EEG of 10 min is collected for each subject. During the process of acquisition, subjects sit in a semi-dark room, stay awake and close their eyes. Besides, they are told not to take needless body movements. The EEG is recorded by the signal amplifier in accordance with the international 20 systems. The linked earlobe A1 and A2 are used as a reference, and the 16 Ag-AgCl scalp electrodes are channels Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6. To eliminate the power line interference, all EEG recordings are first filtered from 0.5Hz to 30Hz by a band passed finite impulse digital filter based on wavelet package (Morlet wavelet). Then average of all channel signals is calculated as the reference potential. EEG data of each channel is re-referenced to remove systematic effects which might be caused by referencing to a particular channel. Besides, interference signals caused by blinking and muscle shaking are eliminated via independent population analysis (ICA) method. In addition, each channel of EEG recording is decomposed into four sub-band: (1-4Hz), (4-8Hz), (8-12Hz), (12-30Hz) by a band-passed finite impulse digital filter based on Morlet wavelet. The mother wave of

complex Morlet wavelet has a Gaussian shape near the center frequency. The details are as follows: The continuous wavelet transforms of EEG signals ($x^{c,k}(\tau)$) is defined as:

$$W^{c,k}(a,t) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x^{c,k}(\tau)\psi(\frac{\tau - t}{a})d\tau \tag{1}$$

where t means time migration, a represents scale parameter, ψ represents the wavelet function, $W^{c,k}(a,t)$ represents the wavelet transform coefficient of the $x^{c,k}(\tau)$ data segment. c is the location of EEG channel, while k is the number of $x^{c,k}(\tau)$.

Although there are many wavelet functions, Morlet wavelet is considered to be the most suitable for analyzing the time-frequency distribution of EEG data. Besides, a relationship between scale parameters a and frequency f exits as: $a = F_c/(fT)$. Then we get the form of wavelet transform with frequency as parameters:

$$\hat{W}^{c,k}(f,t) = W^{c,k}(\frac{F_c}{fT}, t) \tag{2}$$

where $\hat{W}^{c,k}(f,t)$ represents the wavelet transform coefficient of k's EEG signal segment $x^{c,k}(\tau)$ at channel c when the time is t and frequency is f. F_c is the central frequency of wavelet, and T is the signal sampling period, which set as 64.

Wavelet transform can transform signals in time domain and frequency domain without the energy changing, and is widely used in the frequency division of EEG signals. The scale parameter a is 1, while the central frequencies of wavelet F_c are 2Hz, 6Hz, 10Hz, 18Hz, which are coherent with the four frequency bands of EEG and latent factors.

B. Extraction of Latent Factor via VAE

Variational auto-encoder (VAE) methods assume that there is a continuous dynamic system can describe the state of neurons electrophysiological activity, EEG data can be collected as input. Then after repeated "code" and "decoding" to implement the EEG signal feature extraction and compression, the extraction of latent factors and the estimation of EEG hidden activity state can be realized. Suppose that there exists a high-dimensional space $\mathcal{X}$ in which all the data X are generated by the random process of invisible latent factors z and the sample X is subject to an unknown distribution $P_{gt}(X)$. It is hoped that the generated model parameter vector θ can be optimized so that an approximate model $P_{gt}(X)$ can be learned. $P_\theta(X|z)$ describes the model X generated with z, from which the distribution formula can be obtained according to the Bayesian formula as follows:

$$P(X) = \int P_\theta(X|z)P(z)dz \tag{3}$$

When the implied variables z are fixed, theoretically the generated X is also the only confirmed, but in actual engineering there are many factors to disturb, so VAE method assumes that the posterior distribution P_θ obeys the normal distribution, thus $P_\theta(X|z) \sim \mathcal{N}(\mu, \varepsilon)$, the goal of VAE is to maximize the above formula. To this end, a new function $Q(z|X)$ is further defined to estimate the z distribution of X that may be generated, $P(X)$ can be further represented by the following formula:

$$P(X) \approx E_{z\sim Q} P(X|z) \tag{4}$$

In order to maximize the value of $P(X)$, it is necessary to make the calculated $Q(z|X)$ as close to the ideal $P(z|X)$ as possible. Therefore, KL divergence is introduced to constrain the function. The KL divergence formula of the continuous probability distribution is:

$$KL(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx \tag{5}$$

Based on the application of Bayesian formula, it can be further obtained as follows:

$$\begin{aligned} &\mathcal{D}\left[Q(z|X)||P(z|X)\right] \\ &= E_{z\sim Q}\left[\log Q(z|X) - \log P(X|z) - \log P(z)\right] + \log P(X) \end{aligned} \tag{6}$$

Since $\log P(X)$ is independent of the implied variable z, the following equation can be obtained by deforming the above equation:

$$\begin{aligned} &\log P(X) - \mathcal{D}\left[Q(z|X)||P(z|X)\right] \\ &\qquad = E_{z\sim Q}\left[\log P(X|z)\right] - \mathcal{D}\left[Q(z|X)||P(z)\right] \end{aligned} \tag{7}$$

$P(X)$ is the optimization goal of VAE method, and $\mathcal{D}\left[Q(z|X)||P(z|X)\right]$ is the error terms between the real posterior distribution $Q(z|X)$ and the ideal distribution $P(z|X)$, VAE aims to maximize $P(X)$ and minimize the error terms. When selecting the Q that meets the requirements, the right side of the equation can be optimized by using gradient descent. At this time, $Q(z|X)$ can be approximately regarded as the encoder from X to z. In addition, the latent factors are decomposed into four sub-band: by a band-passed finite impulse digital filter based on wavelet package (Morlet wavelet). The frequency energies are used for classification.

C. Latent Factors Analysis via Neural Models

In neural dynamic modeling (NDM) and representational modeling (RM), the dynamics of neural activity are modeled in terms of a latent factor that constitutes the state in the model and is learned via the data and agnostic to the behavior. The description are as follows:

$$\begin{bmatrix} x_{k+1}^{(1)} \\ x_{k+1}^{(2)} \end{bmatrix} = \begin{bmatrix} A_{11}0 \\ A_{21}A_{22} \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} + \begin{bmatrix} w_k^{(1)} \\ w_k^{(2)} \end{bmatrix} \tag{8}$$

$$y_k = \begin{bmatrix} C_{y1}C_{y2} \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} + \upsilon_k \tag{9}$$

$$z_k = \begin{bmatrix} C_{z1}0 \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} + \varepsilon_k \tag{10}$$

where $x_k^{(1)} \in \Re^{n_1}$ is the minimal set of states that affect the specific measured behavior of interest and whose dimension n_1 is the rank of the behavior matrix. In this manner, we refer to $x_k^{(1)}$ as the behaviorally relevant latent factors and $x_k^{(2)} \in \Re^{n_2}$ with $n_2 = n_x - n_1$ as the behaviorally irrelevant latent factors

and n_x is the total number of latent factors. We analyze the decoding performance of VAE by comparing the data decoding of RM and NDM in terms of the number of latent factors via cross-validated correlation coefficient (CC).

D. Takagi-Sugeno-Kang (TSK) Fuzzy Model

Given an original input dataset $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \in \mathbf{R}^d$ and the corresponding class label set $\mathbf{Y} = \{y_1, y_2, \ldots, y_n\}$, where $y_i, i = 1, 2, \ldots, N$ represent the corresponding class, the kth fuzzy inference rules are often defined as

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

where $\mathbf{x} = [x_1, x_2, \ldots, x_d]^T$ 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 of 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 can be expressed as

$$\tilde{y} = \sum_{k=1}^{K} \frac{\mu_k(\mathbf{x}) f_k(\mathbf{x})}{\sum_{k'=1}^{K} \mu_{k'}(\mathbf{x})} = \sum_{k=1}^{K} \tilde{\mu}_k(\mathbf{x}) f_k(\mathbf{x}) \tag{11}$$

where

$$\mu_k(\mathbf{x}) = \prod_{i=1}^{d} \mu_{A_i^k}(x_i) \tag{12}$$

$$\tilde{\mu}_k(\mathbf{x}) = \mu_k(\mathbf{x}) / \sum_{k'=1}^{K} \mu_{k'}(\mathbf{x}) \tag{13}$$

are the fuzzy membership function and 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

$$\mu_{A_i^k}(x_i) = \exp\left(-\left(x_i - c_i^k\right)^2 / \delta_i^k\right) \tag{14}$$

where c_i^k is kth cluster center parameters that is obtained by the classical fuzzy c-means (FCM) clustering algorithm:

$$c_i^k = \sum_{j=1}^{N} u_{jk} x_{ji} / \sum_{j=1}^{N} u_{jk} \tag{15}$$

and the width parameter δ_i^k can be estimated by

$$\delta_i^k = h \cdot \sum_{j=1}^{N} u_{jk} \left(x_{ji} - c_i^k\right)^2 / \sum_{j=1}^{N} u_{jk} \tag{16}$$

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

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

$$\mathbf{x}_{n,e} = \left(1, \mathbf{x}_n^T\right)^T$$
$$\tilde{\mathbf{x}}_n^k = \tilde{\mu}^k(\mathbf{x}_n)\, \mathbf{x}_e$$
$$\rho(\mathbf{x}_n) = \left(\left(\tilde{\mathbf{x}}_n^1\right)^T, \left(\tilde{\mathbf{x}}_n^2\right)^T, \ldots, \left(\tilde{\mathbf{x}}_n^K\right)^T\right)^T \in R^{K(d+1)}$$

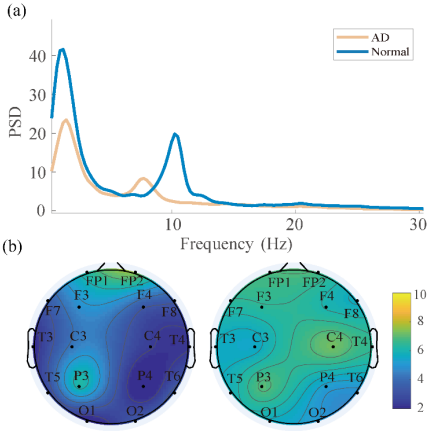

Fig. 1. Power spectrum analysis between AD and normal group. (a) EEG data power spectrum characteristics. The PSD of normal groups has two peaks at 2Hz and 10Hz, while the PSD of AD groups has peaks at 2Hz and 8Hz. (b) Topographic map of power spectral density distribution for AD (left) and normal (right). The overall power spectrum value of AD groups is low, while the values from Frontal and Parietal lobes is relatively strong.

$$\boldsymbol{\beta}^k = \left(\beta_0^k, \beta_1^k, \ldots, \beta_d^k\right)^T$$
$$\boldsymbol{\beta}_g = \left(\left(\boldsymbol{\beta}^1\right)^T, \left(\boldsymbol{\beta}^2\right)^T, \ldots, \left(\boldsymbol{\beta}^K\right)^T\right)^T \tag{17}$$

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

$$\tilde{y}_n = \boldsymbol{\beta}_g^T \rho(\mathbf{x}_n) \tag{18}$$

III. Numerical Results

A. Dynamic Analysis Between AD and Normal Groups

In this section, high-dimensional power characteristics of EEG are analyzed and the rich rhythmic information in the EEG is investigated via relative power of 19 EEG channels in different frequency bands.

In the first section of study, power spectrum analysis is performed and relative power of two groups is calculated. The average power spectrum density (PSD) over all channels is shown in Fig. 1(a). In the figure, the yellow curve represents AD group, and the blue one represents the normal group. Comparing the two curves, we can draw that the power curve of two groups has the same fluctuation trend. It can be concluded that the power of normal group has two peaks at 2 Hz and 10 Hz. Identically, the PSD of AD has two lower peaks than normal groups, and the peak of AD at the high frequency band shows at 8 Hz. However, compared with normal groups, the power of AD groups is significantly higher in the low frequency band (1-3 Hz), while it is lower in high frequency band and the high peak of AD groups shift down from 10 Hz to 8 Hz. It is indicated that the brain function of AD groups has deteriorated and the dominant frequency has changed from alpha to theta. This abnormal brain neuron oscillating activity may be related to the cognitive impairment of early AD. Furthermore, the distribution of PSD in different brain regions between two groups is analyzed. Fig. 1(b) is the topographic map of average PSD distribution of two groups. It is shown that the subjects in both groups have lower PSD

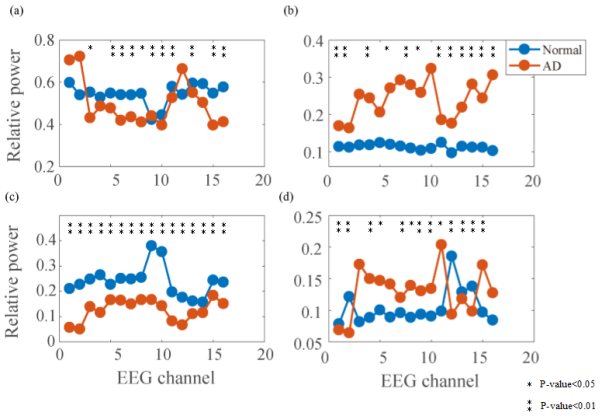

Fig. 2. Relative power of 16 EEG channels in the delta (a), theta (b), alpha (c), beta (d) band for AD (red) and normal (blue) groups. The sequence of EEG number is the same as the electrode position: from left to right, from top to bottom. Please refer to Fig 1 for detailed electrode distribution. * indicates significant different between two groups with $p<0.05$ after ANOVA; ** indicates significant difference with $p<0.01$.

in O2, T6 and P4, with higher PSD in the frontal and parietal lobes. Besides, PSD values of AD groups are significantly lower in temporal and occipital lobe in contrast with frontal lobe, which indicates that weaker neural oscillation activity in these regions and the frontal lobe area is abnormal.

In order to study the rich rhythmic information in the EEG signal and analyze the differences in brain activity among frequency bands, we estimate the relative contribution of the relative power estimation to the EEG signal at a specific frequency band, which is calculated by dividing the given frequency band by the total frequency of all bands. Specifically, the relative power of AD and normal group is calculated and the inter-group differences between two groups of each frequency band in every electrode are assessed. The result is shown in Fig. 2, where the red curve represents AD and the blue curve represents normal group. It is observed that the relative power variation trend is roughly the same in the delta bands, and no significant difference is found between two groups. However, the variation of two groups is opposite in some electrodes, such as Fp1, Fp2 and P3. In the theta band, the relative power of AD groups in almost all electrodes is higher than that of normal group. In contrast, AD groups have lower relative power in the alpha band. Through the analysis of significant differences among groups of electrodes, it is found that the alpha and theta bands have more significant differences. These results indicate that there are valid features in the electrical activity of different frequency bands, and the increase of electrical activity in the theta band and the decrease in alpha bands might be an early marker of cognitive decline in AD.

B. Dynamic Analysis of Latent Factor in EEG

High-dimensional characteristics of EEG are analyzed in the previous section. In this section, the dynamic latent factors of EEG between AD and normal groups are analyzed based on VAE methods. VAE is a kind of neural network which uses the back-propagation algorithm to make the input and output similarly. The framework is shown as Fig. 3. First, the appropriate number of latent factors should be selected to fit EEG. The correlation between the input EEG sequence and the predicted data is chosen as the standard to measure the accuracy of the model. Only when the input EEG is similar to the decoded output data can the VAE model be considered effective. It is necessary to ensure the effectiveness of the model while simplifying the model complexity as much as possible. EEG data of AD patients and normal group are used as input for EEG model training. During training, 10% samples are kept randomly as the test set, and the remaining EEG data are used as the training set. Meanwhile, each segment of data would be iterated for 30 times to achieve better training effect. The test data get input into the VAE model after training, and the time series of implied variable output by the encoder part is called the latent factor of the observed data. Besides, the time series output by the decoder is called the predicted value. The number of latent factors is set from 1 to 32 successively.

During each training, 1/10 of the data segments are kept randomly as the test set. The remaining data are trained separately, and the test set is used to test the VAE model accuracy after training. The data of each channel in the predicted data obtained after the test are connected into the same time series. Besides, the correlation between the original EEG and the predicted data is calculated using Pearson Correlation Coefficients (PCC). The results obtained are shown in Table I. It can be observed that with the increase of the number of latent factors, the accuracy of the model gradually increases. When the number of latent factors exceeds 9, the increase of the accuracy decreases to less than 1%. When the number of latent factors reaches 16, the correlation between the input data and the predicted data exceeds 95%. At the same time, the accuracy of EEG data reconstruction in AD patients is better than that of normal people under the same set of latent factors.

For EEG signals with 16 channels, when the number of latent factors is 10, the overall reconstruction of multi-derivative data can be better achieved. The reconstruction ability of VAE model for each channel can be further considered. In order to observe the respective reconstruction of the 16 channels, the similarity of original data and decoded output data of the 16 channels in the test set are calculated respectively, and the results are shown in Figure 4. When the number of latent factors is 1, the input-output similarity of different EEG channels varies greatly, ranging from 0.27 to 0.75, and the VAE model obtained at this time is not reliable. As the increase in the number of latent factors set, the reconstruction of differences precision between channel increases. When the number is close to 10, the reconstruction accuracy gets into stabilization, and the similarity based on EEG is high. This is consistent with the distribution of topographic maps at different stages, which demonstrate the outcome of VAE model. When the number is 9, the power spectrum of latent factors shows in Fig. 4c. It can be concluded that the power spectrum analysis of normal groups has two peaks at 2 Hz and 10 Hz for latent factors of EEG, while the power spectrum analysis of AD groups has peaks at 2 Hz and 8 Hz. This result is coherent with that of EEG analysis. Besides, the reconstructed data is basically the same as the original data when the number is over 10. The high similarity between the input EEG and the output signal can indicate

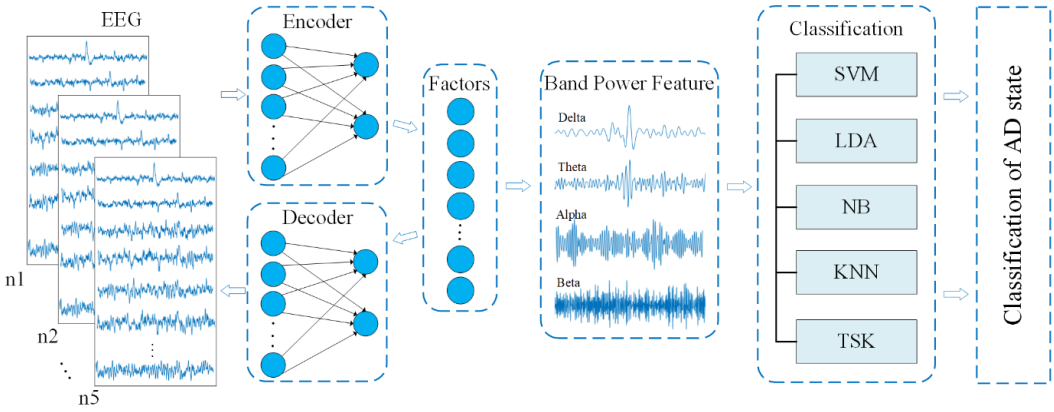

Fig. 3. The framework for extraction of latent factor and classification of AD states. EEG is sent trial – trail to the encoder and decoder, and the latent factors are obtained. Furthermore, the latent factors are divided into 4 frequency bands and energy is extracted to determine the characteristics of each frequency band. Meanwhile, the results of latent factors are fed into multiple classification methods to distinguish AD state, such as Support Vector Machine (SVM), Latent Dirichlet Allocation (LDA), Naïve-Beyesian (NB), k-nearest-neighbor (KNN) and TSK classification.

TABLE I
CORRELATION BETWEEN ORIGINAL EEG AND PREDICTED DATA USING PCC

Number of latent factors	AD	Normal	Number of latent factors	AD	Normal
1	0.5905	0.5293	17	0.9574	0.9167
2	0.6444	0.6218	18	0.9621	0.9169
3	0.7321	0.7107	19	0.9654	0.9252
4	0.7765	0.7474	20	0.9689	0.9247
5	0.8111	0.7809	21	0.9701	0.9259
6	0.8368	0.8041	22	0.9679	0.9261
7	0.8506	0.8204	23	0.9700	0.9271
8	0.8708	0.8404	24	0.9749	0.9345
9	0.8800	0.8537	25	0.9751	0.9383
10	0.8962	0.8710	26	0.9733	0.9388
11	0.9062	0.8822	27	0.9768	0.9392
12	0.9107	0.8843	28	0.9769	0.9414
13	0.9217	0.8885	29	0.9771	0.9427
14	0.9359	0.9023	30	0.9769	0.9428
15	0.9379	0.9031	31	0.9779	0.9430
16	0.9530	0.9132	32	0.9779	0.9441

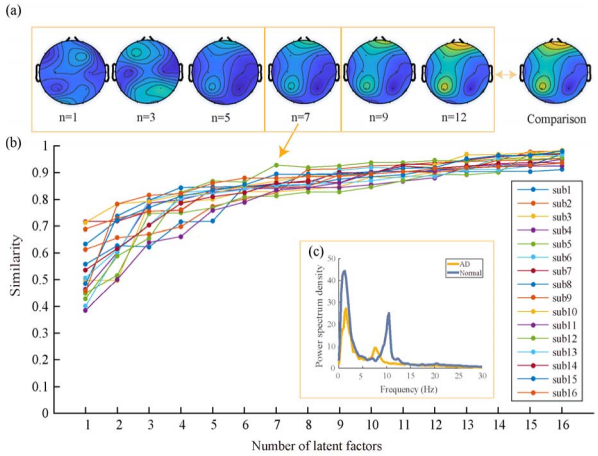

Fig. 4. The reconstruction performance (mean Pearson correlation coefficient) of VAE model when assuming a different number of latent factors. (a) Topographic maps of different stages (left) and comparison normal groups (right). With the increase of latent factors' number, the similarity between decoding maps and original maps gets enhanced. (b) Similarity varies with the number of latent factors. By comparing the topographic maps and similarity, we can draw the conclusion that the reconstruction performance reach stable when the number of latent factors passes 7. (c) Power spectrum analysis of latent factors between AD and normal groups when number is 9.

the effectiveness and robustness of the VAE model in the encoding and reconstruction of the multi-channel EEG. For the multi-channel EEG, a more ideal model can be obtained when the number of latent factors is between 1/2 and 1 of the number of channels.

VAE reveals that behaviorally relevant neural dynamics are much lower-dimensional than implied using standard methods (Fig. 5a). Furthermore, VAE identifies these dynamics more accurately than standard methods both at that dimension (Fig. 5b) and even when standard methods used much higher-dimensional states (Fig. 5c). The recovery performance of the latent factors obtained by VAE is compared with that of NDM and RM, which is shown in Fig. 5. In terms of the optimal performance of decoding, the dimensions of VAE realization decoding are the lowest, around 10, while the dimensions of NDM and RM are 12 or 27. In terms of optimal VAE decoding dimension, VAE decoding performance is better than NDM. At the same time in the best decoding performance, VAE decoding effect is the best, and the CC value is about 28. The "encoder" part of the VAE model is to generate EEG data that is very similar to the input data based on the latent factors. The latent factors obtained at this time can be considered to represent the latent factors used by the brain to generate EEG data. We extract the frequency band energy characteristics of 4 subbands (Delta, Theta, Alpha, Beta) based on each latent component in VAE model. Combined with the TSK system with fuzzy rules, biomarkers under multi-frequency bands are used as input

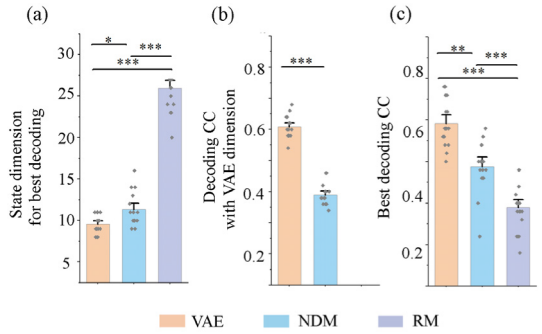

Fig. 5. The difference between VAE and other methods (dynamic relevant). (a) State dimension for best decoding compared in VAE, neural dynamic modeling (NDM) and representational modeling (RM). Among them, the besting decoding state for VAE, NDM and RM is 9,12, and 26, respectively. (b) Decoding CC with VAE dimension between VAE and NDM. By comprising the decoding CC with VAE dimension, VAE reaches 0.6 while NDM reaches 0.4. (c) Besting decoding CC among VAE, NDM and RM. It can be drawn that the best decoding CC for VAE, NDM and RM is 0.62, 0.48 and 0.39, respectively. Asterisks indicate significance of statistical tests with *** $P<0.0005$.

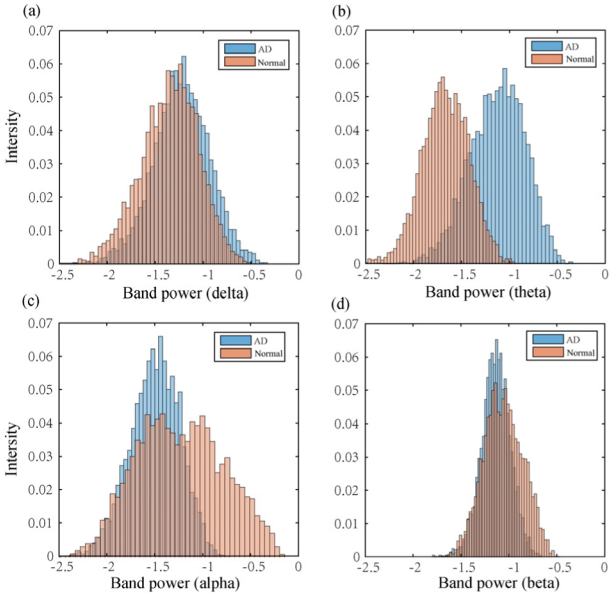

Fig. 6. (a) Delta, (b) Theta, (c) Alpha, and (d) Beta sub-band energy characteristic distribution of latent factors signals. It can be concluded that the distribution of AD and normal groups is significantly different in theta band.

features of TSK fuzzy model, and supervised learning is used to train the EEG feature recognition of AD.

Furthermore, the individual components are divided into four frequency bands. Besides, the energy feature of each frequency band is calculated and continuous energy eigenvalues are obtained. The distribution statistics of energy characteristics of sub-frequency bands of all 10 latent factors are carried out according to frequency bands, which is shown in Fig. 6.

It can be observed that the distribution of AD in delta frequency band is similar to that of normal groups, with the mean value of energy distribution is close. However, it is shown in Fig. 6(b) that AD distribution exists obvious difference with normal group in the theta frequency band. The mean value of the sample in normal group is about -1.75, and the mean value of the energy in the frequency band of AD

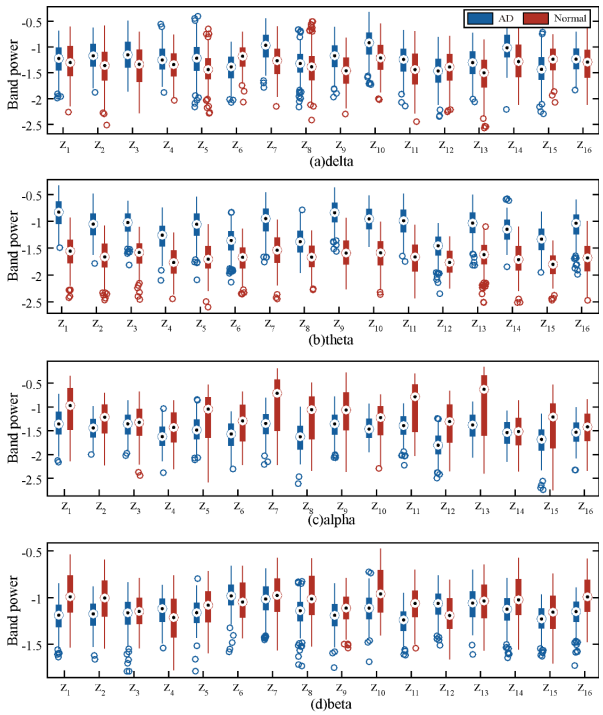

Fig. 7. Box diagram of independent component energy characteristics of sub-bands (a) delta, (b) theta, (c) alpha, and (d) beta in AD and normal latent factor signals. It can be concluded that the distribution of AD and normal groups is significantly different in theta band for subbands.

is about -1. What's more, the degree of separation is more obvious, while indicates that there is a significant difference in the theta band between AD patients and normal group. As can be seen from Fig.6 (c), the distribution range of AD and normal group in the alpha frequency band has much overlap, but the distribution state of two groups is very different. The AD group presents an obvious single peak while the normal group presents a bimodal distribution, which indicates that there are differences in the distribution of different internal latent factors. The average frequency band energy of the AD group is about -1.5, and the average of the two peaks of the normal control group are -1.5 and -0.9, respectively. These results are in consistent with that of Fig. 2: It is found that the relative power of AD in theta and alpha frequency band is different from that of normal groups. This energy comparison further validates the use of latent factors prior to the classification.

The distribution of the first peak is similar to that of the AD group, while the distribution range of the second peak is significantly different from that of the first peak, indicating that the energy characteristics of latent factor extracted from VAE model are different. As shown in Fig. 6 (d), the beta frequency band energy in the AD group is more concentrated than that in the normal group, and its distribution is similar to that in the delta frequency band. The mean value of the frequency band energy in the AD group is about -1.15, and the distribution range is $[-1.7, -0.6]$. The mean value of the normal control group is about -1, and the distribution range is $[-1.7, -0.5]$.

It can be seen in Fig. 7 that there may exist differences in the latent factors within the same VAE model. Therefore, the 16 latent factors of each frequency band are presented in

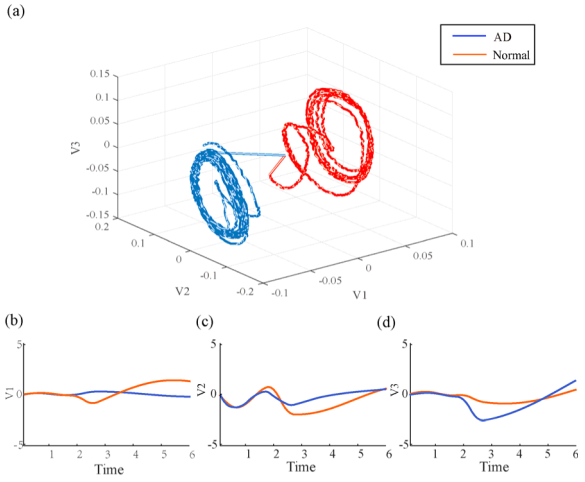

Fig. 8. The low-dimensional characteristic of AD latent factor. (a) the rotation characteristics of AD and normal groups in theta frequency band. (b) The pattern of evolution for each latent factor in the axis of time. The rotation speed and direction are different for AD and normal groups.

the form of box diagram, respectively. In the delta frequency band, most of the latent factors in the AD and normal group have no difference. But mild differences exist in a few variables such as z3, z7, z9, z10, z14. Besides, the power of delta band in AD groups is higher than that in normal groups. In parallel, there is a significant difference in the power of the theta band between AD patients and normal groups (Fig. 7b).

Except for the latent factor Z8, the independent component of Z12 has significant differences in the power of the theta band, which is consistent with the trend of frequency band energy characteristics of the principal component after linear dimension reduction. The theta band energy of AD is higher than that of the normal group. Identically, in alpha band, the mean values of Z7, Z11 and Z13 variables are higher than other variables in normal groups, while the mean values of each variable in the AD group are not significantly different, indicating that different latent factors could represent different implied features. (Fig. 7c). For most of the latent factors, the independent component of beta frequency band power energy does not show significant difference between AD and the normal group (Fig. 7d). For a few latent factors with differences, the frequency band energy of AD group is lower than that of the normal group, which is consistent with the difference of linear dimension-reduction method, but the number of variables with differences is smaller.

Furthermore, the dynamics of neural network are further extracted by latent factors analysis. It is found that the neural state rotates with time, such as the state of a pendulum rotates in a space which is determined by velocity. In order to verify the dynamic characteristics of the latent factors, the signals are projected onto the three-dimensional state space, and the dynamics are jointly represented by multiple independent components. The results are shown in Fig. 8 for theta band. Because the dimensionality related to the dynamic structure is not found, the mapping method is used to find an information plane in each independent component and capture the strongest rotation trend in each independent component. As can be seen from Fig. 8, there is a transient rotation of the neural state represented by the latent signal, and the rotation trend direction of the neural state of AD patients is different from that of normal people. Meanwhile, the range of the rotational state of AD patients is larger and the rotation is slower. This result is consistent with the power spectrum characteristic of higher dimensional oscillation analysis.

The low-dimensional trajectory expresses the time evolution process of the key EEG channels of cortical rhythm, thereby representing the time evolution law of the cerebral cortex rhythm oscillation activity. The running speed in low-dimensional space has obvious periodicity, which is closely related to the changes in cerebral cortical rhythm oscillations. As can be seen from the orange dotted line, the fluctuations in the running speed of the low-dimensional trajectory of the cortical rhythm are consistent with the changes in rhythm information. The renewal speed of the cerebral cortex may be positively correlated with the rhythm. The amplitude of the rhythm curve evolves smoothly from the high point to the low point, while the lowest point of the amplitude goes downward. The evolution of a cycle is obviously steep, and the low-dimensional motion trajectory intuitively shows this result, which has certain research value for the specific analysis of the rhythmic oscillation activity of the cerebral cortex. In addition, in the AD state, the cortical rhythm activity in the low-dimensional space is significantly different from the normal state. The activity in the normal state moves faster than the AD state. In the component space, the rotation speed is faster. This result is consistent with the result in dynamic analysis of electrical time frequency between AD and normal human brain.

C. Classification of Potential EEG Factors

Primarily, the energy characteristics of the four frequency bands of each latent factor are used as the input of TSK classifier. In each training, the training data is cross-verified by 10 folds. The training set (including data collection and label set) is randomly divided into 10 subsets. Taking turns to choose the 1 in 10 holds as a test to evaluate the use of sample set, the remaining part 9 samples are used as the input in the process of the training sample set. The authentication process will repeat until the traversal 10 times the average shares a set of 10 samples, and will be 10 times take average as the final classification accuracy verification results. Classification results are shown in table II, each band in the highest classification accuracy of latent factors in the form of a bold said out. Differences in the classification results of different latent factors in the same frequency band can be observed. Among them, the latent factor with the highest classification accuracy appears in the theta frequency band. That is, the classification accuracy of independent component 9 can reach 0.9430, but the accuracy of some variables under this frequency band is lower than 0.75, which indicates that different latent factors contain different hidden information and have different explanatory effects on AD (shown in TABLE II).

Multiple parameters can be taken as input vector to TSK classifier for feature classification. In that case, optimal factors under different frequency band are selected. Two bands combination, three bands and four band are chosen in TSK training. The optimal input vector and classification results are shown

TABLE II
CLASSIFICATION ACCURACY OF ENERGY CHARACTERISTICS IN FOUR FREQUENCY BANDS

latent factor	Delta	Theta	Alpha	Beta
z_1	0.5410	0.9335	0.7184	0.7152
z_2	0.6363	0.8846	0.6946	0.7136
z_3	0.6156	0.9256	0.5745	0.5441
z_4	0.5933	0.8465	0.6788	0.6137
z_5	0.6488	0.9020	0.7722	0.6298
z_6	0.6585	0.7911	0.6850	0.6421
z_7	0.7026	0.8735	**0.7848**	0.5692
z_8	0.5650	0.7326	0.7486	0.6695
z_9	**0.7135**	**0.9430**	0.7105	0.5981
z_{10}	0.7041	0.9352	0.6692	0.6976
z_{11}	0.6234	0.9272	0.7817	**0.7435**
z_{12}	0.5522	0.7515	0.7800	0.6691
z_{13}	0.6375	0.9209	0.7727	0.5630
z_{14}	0.7040	0.8590	0.5680	0.6442
z_{15}	0.6408	0.8765	0.7518	0.6251
z_{16}	0.5792	0.8939	0.6060	0.7119

in Table III. It can be observed that for the latent factor of delta frequency band (the accuracy is 0.7135), adding variables of other frequency bands as common inputs can improve the classification results. Besides, TSK can achieve the optimal classification accuracy with the combination of theta frequency band.

The same situation can also be observed in the latent factors of theta band, and the combination of theta band and beta band can obtain the optimal classification result. However, as for the latent factors of alpha frequency band, when the alpha frequency band and delta are used as input features for model training, the test classification accuracy decrease compared with that of single input, which indicates that the parameters of delta frequency band may not be able to produce information complementarity with alpha frequency band, or the model may be overfitted. The best results of classification accuracy appear at 0.9810 when taking delta, theta, alpha and beta frequency band as common input. Meanwhile, for the original EEG data, the accuracy achieves 0.7526 when taking power from four frequency bands as common input. This accuracy is better than taking single frequency band signal as input. Besides, the delta frequency band is proved to be effective in enhancing the accuracy, while the alpha frequency band takes poor effect in enhancing the accuracy. In this way, it is crucial to choose an appropriate input features combination via machine learning.

In addition, TSK algorithm is compared with the other four traditional Linear methods commonly used in machine identification, including SVM, Linear Discriminant Analysis

TABLE III
OPTIMAL INPUT VECTOR AND CLASSIFICATION RESULTS BASED ON TSK CLASSIFIERS

Input parameter	accuracy	sensitivity	specificity	original data
Delta & Theta	0.9574	0.9799	0.9328	0.7214
Delta & Alpha	0.7641	0.9259	0.6113	0.6127
Delta & Beta	0.8290	0.8375	0.8179	0.6874
Theta & Alpha	0.9557	0.9495	0.9614	0.7175
Theta & Beta	0.9622	0.9516	0.9692	0.7324
Alpha & Beta	0.9145	0.9227	0.9016	0.7120
Delta & Theta & Alpha	0.9699	0.9862	0.9516	0.7128
Delta & Theta & Beta	0.9731	0.9766	0.9689	0.7327
Delta & Alpha & Beta	0.9098	0.9125	0.9123	0.6981
Theta & Alpha & Beta	0.9795	0.9862	0.9760	0.7132
Delta & Theta & Alpha & Beta	0.9810	0.9802	0.9806	0.7526

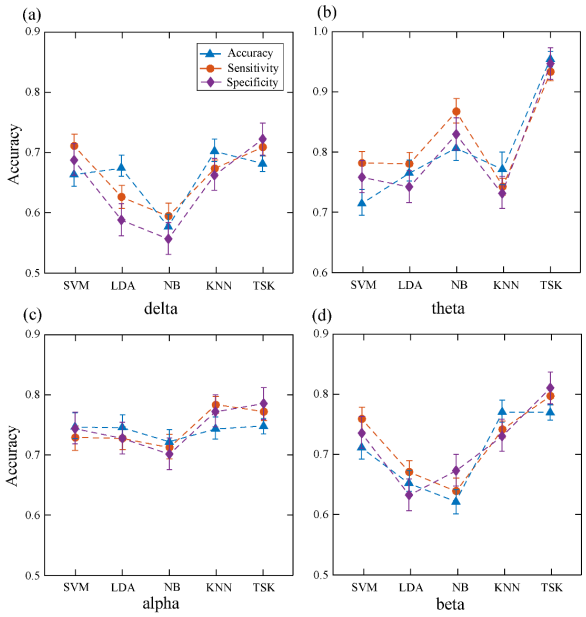

Fig. 9. Comparison of (a) Delta, (b) Theta, (c) Alpha, and (d) Beta sub-bands of energy as single feature inputs for different classifiers. For one frequency band signal input, TSK perform better than other method. In addition, the performance of classification gets well in theta frequency band.

(LDA), Naive Bayesian (NB) and K-Nearest Neighbor (KNN) algorithm. The Delta, Theta, Alpha and Beta frequency band energy features are selected as input of various classifiers in the above paper, and the recognition results are shown in Fig.9. It can be observed from the figure that for the theta band and beta band, the TSK classifier can show better classification effect than the linear classifier, but this advantage is not obviously reflected in the delta band and alpha band. Furthermore, the delta, theta, alpha, beta band four parameters as features combination input into multiple classifiers, can get the classification results as shown in table IV. Among the methods, linear classifier in the NB method is the optimal classification, identification accuracy can reach 94.95%, and TSK

TABLE IV
CLASSIFICATION RESULTS BASED ON DIFFERENT CLASSIFIERS

Input parameter	accuracy	sensitivity	specificity	original
SVM	0.9361	0.9021	0.9155	0.7251
LDA	0.8871	0.7995	0.8904	0.6902
NB	0.9495	0.8862	0.9360	0.7310
KNN	0.9160	0.9334	0.8896	0.6872
TSK	**0.9810**	**0.9802**	**0.9806**	**0.7526**

method in identifying AD latent factor characteristics better than the traditional linear methods. Besides, the decoding of latent factors is better than that of original EEG signals. The results are shown in Tabel IV.

According to the classification results of energy feature classification of latent factor quantum frequency band based on multiple classifiers, it can be seen that the theta frequency band in the latent factor space can provide more effective information for AD feature recognition. Theta oscillation is usually closely related to working memory and personality characteristics. Studies have shown that memory tasks can induce significant Theta oscillation, and the size of working memory load can affect the intensity of Theta oscillation in EEG. Therefore, after coding by VAE method, the characteristics of higher brain activities such as memory and language in EEG can be retained and reflected in latent factors. The analysis of latent factors and feature extraction for AD based on VAE method has important significance for the establishment of recognition model.

IV. CONCLUSION AND DISCUSSIONS

In this paper, we have proposed a novel machine learning method for AD identification with EEG signal. To improve model interpretability and identification accuracy, VAE and TSK fuzzy system models are employed. Latent factors are constructed to investigate the default brain dynamics across subjects in AD processes. Taken the energy features of latent factors as independent inputs, a fuzzy rule-based TSK model is used to classify AD and normal EEG signals. Compared with linear classifier, TSK fuzzy classifier has better performance in classification of energy feature from sub-frequency bands of latent factors. The accuracy of identification could meet 98.10% when the combination of energy features of four frequency bands is used as model input.

The latent factors can recover the dynamics of EEG when using VAE method. VAE realizes information compression and reconstruction in the form of unsupervised learning, and the encoder is used to describe the probability distribution of each potential attribute, which can also realize the extraction of latent factors [34]. The latent factors can be further applied to identify brain functional networks or locate the sources of intracranial current without observing all of their EEG signals [45]. Each factor captures a pattern of co-activation across EEG signals. When the number of latent factors is over 3, the degree of similarity begins to approach that of original data. Besides, the reconstructed data is basically the same as the original data when the number is over 10. AD distribution exists obvious difference with normal group in the theta frequency band, and the rotation trend direction of the neural state for AD is different from that of normal group. This result is consistent with the power spectrum characteristic of higher dimensional oscillation. The accuracy of identification could up to 98.10% when the combination of energy features of four frequency bands is used as model input. As a comparison, the recognition accuracy of 89.02% could be achieved by using the energy characteristics of low-dimensional signal of theta band via PCA. Experimental results demonstrate that autoencoder-like neural networks are suitable for unsupervised EEG modeling.

Comparing with traditional classification methods, we present a novel diagnosis model by integrating fuzzy rules with machine learning, which builds up the reliability and interpretability of classification results. In this paper, N-TSK provides an effective tool in identifying AD from latent factors. It is found that TSK method has more excellent performance in identifying AD characteristics than other four traditional Linear methods, such as SVM, NB, LDA and KNN. The results are shown in Tabel IV. It may because TSK could transform the decision output into a value between 0 and 1 through fuzzy rules, which is closer to human thinking than other dichotomies of either 0 or 1 [46]. These findings indicate that our methods show light on identifying AD from the view of latent factors and fuzzy systems. Furthermore, machine learning methods has been applied in modeling and prediction of gene-expression patterns, diagnosis of epileptic seizures [47]. Another feasible application of fuzzy algorithms is predicted-warning of AD. This method helps to reduce its dependence on the acquisition of data labels [48]. At the same time, the results of low-dimensional dynamic changes indicate that the slower renewal speed of the cerebral cortex and the weakening of information transmission capacity may be related to the appearance of neurofibrillary tangles and amyloid. These factors cause abnormal rhythms, which lead to the occurrence of diseases.

Collectively, this work provides a feasible tool for the identification of neurological dysfunction from the view of latent factors, especially contributing to the diagnosis of AD. For limitations, the dataset is small and doesn't include MCI subject, which limits the effect of training and classification. Besides, the analysis of EEG depends on the sensor level, which restricts the interpretation of the findings. Future works may focus on designing more accurate algorithms for the identification of AD EEG, by combing complex network methods and machine learning techniques.

REFERENCES

[1] C. Wattmo, E. Londos, and L. Minthon, "Response to cholinesterase inhibitors affects lifespan in Alzheimer's disease," *BMC Neurol.*, vol. 14, no. 1, p. 173, 2014.
[2] W. H. De *et al.*, "EEG abnormalities in early and late onset Alzheimer's disease: Understanding heterogeneity," *J. Neurol., Neurosurg. Psychiatry*, vol. 82, no. 1, pp. 67–71, 2011.

[3] L. Cai *et al.*, "Functional integration and segregation in multiplex brain networks for Alzheimer's disease," *Frontiers Neurosci.*, vol. 14, pp. 14–51, Feb. 2020.

[4] G. Small and R. Bullock, "Defining optimal treatment with cholinesterase inhibitors in Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 2, pp. 177–184, Mar. 2011.

[5] J. L. Cummings, "Biomarkers in Alzheimer's disease drug development," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 1218–1222, 2011.

[6] G.-F. Chen *et al.*, "Amyloid beta: Structure, biology and structure-based therapeutic development," *Acta Pharmacologica Sinica*, vol. 38, no. 9, pp. 1205–1235, Sep. 2017.

[7] G. B. Frisoni *et al.*, "Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers," *Lancet Neurol.*, vol. 16, pp. 661–676, Aug. 2017.

[8] R. Wang, J. Wang, S. Li, H. Yu, B. Deng, and X. Wei, "Multiple feature extraction and classification of electroencephalograph signal for Alzheimers' with spectrum and bispectrum," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 25, no. 1, Jan. 2015, Art. no. 013110.

[9] D. Guo, M. Perc, T. Liu, and D. Yao, "Functional importance of noise in neuronal information processing," 2018, *arXiv:1812.09897*. [Online]. Available: https://arxiv.org/abs/1812.09897

[10] Z. Gao *et al.*, "Complex networks and deep learning for EEG signal analysis," *Cognit. Neurodyn.*, vol. 15, pp. 369–388, Aug. 2021.

[11] C. Babiloni, "Cortical neural synchronization mechanisms in Alzheimer's disease revealed by resting state EEG rhythms: A window on cholinergic function," *Alzheimer's Dementia*, vol. 9, no. 4, p. 515, 2013.

[12] N. Mammone *et al.*, "Brain network analysis of compressive sensed high-density EEG signals in AD and MCI subjects," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 527–536, Jan. 2019.

[13] N. Mammone *et al.*, "Permutation disalignment index as an indirect, EEG-based, measure of brain connectivity in MCI and AD patients," *Int. J. Neural Syst.*, vol. 27, no. 5, Aug. 2017, Art. no. 1750020.

[14] A. Horvath, "EEG and ERP biomarkers of Alzheimer's disease a critical review," *Frontiers Biosci.*, vol. 23, no. 1, pp. 183–220, 2018.

[15] C. T. Briels, D. N. Schoonhoven, C. J. Stam, H. de Waal, P. Scheltens, and A. A. Gouw, "Reproducibility of EEG functional connectivity in Alzheimer's disease," *Alzheimer's Res. Therapy*, vol. 12, no. 1, pp. 1–14, Dec. 2020.

[16] J. J. van der Zande, A. A. Gouw, I. van Steenoven, P. Scheltens, C. J. Stam, and A. W. Lemstra, "EEG characteristics of dementia with Lewy bodies, Alzheimer's disease and mixed pathology," *Frontiers Aging Neurosci.*, vol. 10, p. 190, Jul. 2018.

[17] H. Yu, X. Wu, L. Cai, B. Deng, and J. Wang, "Modulation of spectral power and functional connectivity in human brain by acupuncture stimulation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 5, pp. 977–986, May 2018.

[18] A. A. Gouw *et al.*, "Heterogeneity of small vessel disease: A systematic review of MRI and histopathology correlations," *J. Neurol. Neurosurg. Psychiatry*, vol. 82, pp. 126–135, Feb. 2010.

[19] R. Wang, J. Wang, H. Yu, X. Wei, C. Yang, and B. Deng, "Decreased coherence and functional connectivity of electroencephalograph in Alzheimer's disease," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 24, no. 3, Sep. 2014, Art. no. 033136.

[20] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, Dec. 2010.

[21] C. Pandarinath *et al.*, "Inferring single-trial neural population dynamics using sequential auto-encoders," *Nature Methods*, vol. 15, no. 10, pp. 805–815, Oct. 2018.

[22] O. Kouropteva, O. Okun, and M. Pietikäinen, "Incremental locally linear embedding," *Pattern Recognit.*, vol. 38, no. 10, pp. 1764–1767, Oct. 2005.

[23] C. M. Wang, P. R. Zhang, Y. Zhang, L. Zhang, X. Chi, and M. Zhang, "Research on correlation between complicated acute lung injury/acute respiratory distress syndrome and plasminogen activator inhibitor-1 in patients with acute cerebral infarction," *J. Clin. Med. Pract.*, vol. 16, p. 23, May 2012.

[24] R. Gao *et al.*, "Zero-VAE-GAN: Generating unseen features for generalized and transductive zero-shot learning," *IEEE Trans. Image Process.*, vol. 29, pp. 3665–3680, Jan. 2020.

[25] X. Li *et al.*, "Latent factor decoding of multi-channel EEG for emotion recognition through autoencoder-like neural networks," *Frontiers Neurosci.*, vol. 14, p. 87, Mar. 2020.

[26] Z. Rošťáková, R. Rosipal, S. Seifpour, and L. J. Trejo, "A comparison of non-negative tucker decomposition and parallel factor analysis for identification and measurement of human EEG rhythms," *Meas. Sci. Rev.*, vol. 20, no. 3, pp. 126–138, Jun. 2020.

[27] L. L. Hawley, N. A. Rector, A. DaSilva, J. M. Laposa, and M. A. Richter, "Technology supported mindfulness for obsessive compulsive disorder: Self-reported mindfulness and EEG correlates of mind wandering," *Behav. Res. Therapy*, vol. 136, Jan. 2021, Art. no. 103757.

[28] Y. Tang, D. Chen, L. Wang, A. Y. Zomaya, J. Chen, and H. Liu, "Bayesian tensor factorization for multi-way analysis of multi-dimensional EEG," *Neurocomputing*, vol. 318, pp. 162–174, Nov. 2018.

[29] E. Kinney-Lang, L. Spyrou, A. Ebied, R. F. M. Chin, and J. Escudero, "Tensor-driven extraction of developmental features from varying paediatric EEG datasets," *J. Neural Eng.*, vol. 15, no. 4, Aug. 2018, Art. no. 046024.

[30] D. Kobak *et al.*, "Demixed principal component analysis of neural population data," *Elife*, vol. 5, Apr. 2016, Art. no. e10989.

[31] M. T. Kaufman, M. M. Churchland, S. I. Ryu, and K. V. Shenoy, "Cortical activity in the null space: Permitting preparation without movement," *Nature Neurosci.*, vol. 17, no. 3, pp. 440–448, Mar. 2014.

[32] M. M. Churchland *et al.*, "Neural population dynamics during reaching," *Nature*, vol. 487, no. 7405, pp. 51–56, Jul. 2012.

[33] M. Aghagolzadeh and W. Truccolo, "Latent state-space models for neural decoding," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 3033–3036.

[34] Y. Zhao and I. M. Park, "Variational latent Gaussian process for recovering single-trial dynamics from population spike trains," *Neural Comput.*, vol. 29, no. 5, pp. 1293–1316, May 2017.

[35] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel EEG data through convolutional recurrent neural network," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2017, pp. 352–359.

[36] D. Chen, Y. Tang, H. Zhang, L. Wang, and X. Li, "Incremental factorization of big time series data with blind factor approximation," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 2, pp. 569–584, Feb. 2021.

[37] X. Tian *et al.*, "Deep multi-view feature learning for EEG-based epileptic seizure detection," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 1962–1972, Oct. 2019.

[38] H. Yu, X. Li, X. Lei, and J. Wang, "Modulation effect of acupuncture on functional brain networks and classification of its manipulation with EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 1973–1984, Oct. 2019.

[39] Y. Zhang, Z. Zhou, H. Bai, W. Liu, and L. Wang, "Seizure classification from EEG signals using an online selective transfer TSK fuzzy classifier with joint distribution adaption and manifold regularization," *Frontiers Neurosci.*, vol. 14, pp. 2–8, Jun. 2020.

[40] T. Ni, X. Gu, and C. Zhang, "An intelligence EEG signal recognition method via noise insensitive TSK fuzzy system based on interclass competitive learning," *Frontiers Neurosci.*, vol. 14, p. 837, Sep. 2020.

[41] Z. Deng, P. Xu, L. Xie, K.-S. Choi, and S. Wang, "Transductive joint-knowledge-transfer TSK FS for recognition of epileptic EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 8, pp. 1481–1494, Jun. 2018.

[42] Y. Jiang *et al.*, "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2270–2284, Dec. 2017.

[43] Z. Cao, W. Ding, Y.-K. Wang, F. K. Hussain, A. Al-Jumaily, and C.-T. Lin, "Effects of repetitive SSVEPs on EEG complexity using multiscale inherent fuzzy entropy," *Neurocomputing*, vol. 389, pp. 198–206, May 2020.

[44] H. Abbasi, A. J. Gunn, L. Bennet, and C. P. Unsworth, "Latent phase identification of high-frequency micro-scale gamma spike transients in the hypoxic ischemic EEG of preterm fetal sheep using spectral analysis and fuzzy classifiers," *Sensors*, vol. 20, no. 5, p. 1424, Mar. 2020.

[45] T. H. Grandy, M. Werkle-Bergner, C. Chicherio, M. Lövdén, F. Schmiedek, and U. Lindenberger, "Individual alpha peak frequency is related to latent factors of general cognitive abilities," *NeuroImage*, vol. 79, pp. 10–18, Oct. 2013.

[46] H. Yu, X. Lei, Z. Song, C. Liu, and J. Wang, "Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 1, pp. 60–71, Jan. 2020.

[47] M. Esmaeilpour and A. R. A. Mohammadi, "Analyzing the EEG signals in order to estimate the depth of anesthesia using wavelet and fuzzy neural networks," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 4, pp. 12–15, Dec. 2016.

[48] A. Jafarifarmand, M. A. Badamchizadeh, S. Khanmohammadi, M. A. Nazari, and B. M. Tazehkand, "A new self-regulated neuro-fuzzy framework for classification of EEG signals in motor imagery BCI," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1485–1497, Jun. 2018.