Multi-Scale Time-Series Kernel-Based Learning Method for Brain Disease Diagnosis

Zehua Zhang, Jiaqi Ding, Junhai Xu, Jijun Tang, and Fei Guo

***Abstract*—The functional magnetic resonance imaging (fMRI) is a noninvasive technique for studying brain activity, such as brain network analysis, neural disease automated diagnosis and so on. However, many existing methods have some drawbacks, such as limitations of graph theory, lack of global topology characteristic, local sensitivity of functional connectivity, and absence of temporal or context information. In addition to many numerical features, fMRI time series data also cover specific contextual knowledge and global fluctuation information. Here, we propose multi-scale time-series kernel-based learning model for brain disease diagnosis, based on Jensen-Shannon divergence. First, we calculate correlation value within and between brain regions over time. In addition, we extract multi-scale synergy expression probability distribution (interactional relation) between brain regions. Also, we produce state transition probability distribution (sequential relation) on single brain regions. Then, we build time-series kernel-based learning model based on Jensen-Shannon divergence to measure similarity of brain functional connectivity. Finally, we provide an efficient system to deal with brain network analysis and neural disease automated diagnosis. On Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, our proposed method achieves accuracy of 0.8994 and AUC of 0.8623. On Major Depressive Disorder (MDD) dataset, our proposed method achieves accuracy of 0.9166 and AUC of 0.9263. Experiments show that our proposed method outperforms other existing excellent neural disease automated diagnosis approaches. It shows that our novel prediction method performs great accurate for identification of brain diseases as well as existing outstanding prediction tools.**

***Index Terms*—Functional magnetic resonance imaging, time-series kernel, disease diagnosis, alzheimeris disease, major depressive disorder, Jensen-Shannon divergence.**

Manuscript received August 27, 2019; revised December 8, 2019, January 30, 2020, and March 15, 2020; accepted March 23, 2020. Date of publication March 30, 2020; date of current version January 5, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 61772362 and 61972280 and in part by National Key R&D Program of China under Grants 2018YFC0910405 and 2017YFC0908400. *(Corresponding author: Fei Guo.)*

Zehua Zhang, Jiaqi Ding, Junhai Xu, and Fei Guo are with the College of Intelligence and Computing, Tianjin University, Tianjin 300350, China (e-mail: zehua_new@yeah.net; 13256683086@163.com; jhxu@tju.edu.cn; guofeieileen@163.com).

Jijun Tang is with the College of Intelligence and Computing, Tianjin University, Tianjin 300350, China, and with the Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin 300350, China, and also with the Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208 USA (e-mail: tangjijun@tju.edu.cn).

Digital Object Identifier 10.1109/JBHI.2020.2983456

I. Introduction

THE functional magnetic resonance imaging (fMRI) can be used as a noninvasive technique for analyzing brain activity, like brain network analysis, neural disease automated diagnosis and so on. The fMRI data quantify neuronal activity by detecting changes associated with cerebral blood flow, measuring intrinsic Blood-Oxygen-Level-Dependent (BOLD) signal fluctuations of distributed brain regions. Correlation among BOLD signals can show Functional Connectivity (FC) relationship of distributed brain regions. Some studies convert FC information to brain network, for analyzing distributed networks corresponding to brain function and mining sensitive properties for psychological disease states [1]–[8].

Based on graph theory, brain network analysis on fMRI data can provide concise quantitative information about connectivity property of distributed brain regions [9], [10]. Temporal correlation in neuronal activity is reflection of linear or nonlinear interaction within different time scales [11], [12]. Effective connection can be estimated from observed synchronous or asynchronous perturbations, in order to indicate direct or indirect influences of distributed brain regions [13], [14]. Graph theory applied to brain network analysis [15], not only provides quantitative measurement for determining connectivity information of local brain activity, but also affords general framework for analyzing heterogeneous graph of different data [16].

However, many existing methods based on graph theory [17]–[19] have some drawback, such as lack of global topology characteristic [20], [21], local sensitivity of functional connectivity [22]–[24], and absence of temporal or context information. In addition to many numerical features, fMRI time series data also cover specific contextual knowledge and global fluctuation information. [25].

In this paper, we propose a statistical analysis method based on multi-scale time-series kernel-based learning model for brain disease diagnosis. First, we calculate correlation value within and between brain regions over time, and extract multi-scale synergy expression probability distribution between brain regions as well as state transition probability distribution on a single brain region. Secondly, we build time-series kernel-based learning model based on Jensen-Shannon divergence to measure similarity of brain functional connectivity. Thirdly, we provide an efficient system to deal with problems of brain network analysis and neural disease automated diagnosis, which can effectively study pathological changes of mental disorders on fMRI data. Experiments show that our proposed method performs great

accurate for identification of brain diseases as well as existing outstanding prediction tools.

II. Related Works

The widely used functional connection model still has some shortcomings. In recent years, there have been some improvements to traditional methods. Among them, novel high-order FC correlations can be extracted to characterize how low-order correlations between different pairs of brain regions interact with each other [3], which captures local changes and uses a sliding window approach. A new switching delayed particle swarm optimization (SDPSO) algorithm is proposed to optimize parameters of Support Vector Machine (SVM) [26]. Different from common approaches that deal with vector-based representation of data through feature engineering, kernel-based methods offer a natural framework to measure similarity between two graphs and further analyze neural disease.

Recently, kernel-based model on some structured objects, such as strings, trees and graphs, has been proposed in many excellent studies [27]–[29] and some different fields [30]–[34]. Haussler [35] firstly defined a principled way for designing kernels on structured objects, called R-convolution kernel model. Most of graph kernels are denoted by comparing small sub-graphs like walks, paths or graphlets. Some existing methods usually obtain better performance based on local detailed features and global topological information, such as shortest-path kernel-based method [36], marginalized kernel-based method [37] and subtree kernel-based method [38].

Furthermore, some researchers use geometric embedding approach to construct graph kernels [39]–[41]. Shrivastava [42] defined an effective kernel via a novel mathematical representation of graphs. Shervashidze [27] proposed a family of efficient kernels via a rapid feature extraction scheme based on Weisfeiler-Lehman isomorphism test. For brain network analysis, Jie [43] constructed a new sub-network kernel to diagnose neural disease with good performance, which considers inherent characteristic and multi-level topological information in brain network.

III. Methods

We propose a novel time-series multi-kernel learning framework on fMRI data for brain disease classification. Our method includes four main steps: (1) image pre-processing and state sequence mapping; (2) calculating multi-scale synergy expression probability distribution between brain regions; (3) calculating state transition probability distribution on single brain regions; (4) time-series kernel modeling based on Jensen-Shannon divergence. The flow chart is shown in Fig. 1.

A. Notations and Definitions

Let $\{A_1, A_2, \ldots, A_k, \ldots, A_K\} \in S$ denotes fMRI data which contains K samples, and $\{y_1, y_2, \ldots, y_k, \ldots, y_K\} \in Y, y_k \in \{0, 1\}$ represents corresponding label vector.

Here, we define one of fMRI data as a multivariate time series data A_k, as follows:

$$A_k = \{T_1^k, T_2^k, \ldots, T_n^k, \ldots, T_N^k\} \tag{1}$$

where T_n^k represents n-th time series of k-th fMRI data.

The univariate time series data T_n^k can be defined as follows:

$$T_n^k = \{t_{n,1}^k, t_{n,2}^k, \ldots, t_{n,m}^k, \ldots, t_{n,M}^k\} \tag{2}$$

where $t_{n,m}^k$ represents the value of m-th time point in n-th time series of k-th fMRI data.

In addition, $I = \{I_1, I_2, \ldots, I_t, \ldots, I_T\}$ represents a collection of multiple intervals, where $I_t = [r_t, s_t]$ denotes a positive integer interval. And also, $U = \{u_1, u_2, \ldots, u_e, \ldots, u_E\}$ represents a state space, where u_e denotes a state.

B. Image Pre-Processing and Numerical Sequence Mapping

First, we perform standardized preprocessing on all collected data [44]. Then, we use anatomical templates to convert original voxel-based image into ROI-based form [45], [46]. Thirdly, we apply a statistical method on multi-variable time series in order to map numerical sequence into state sequence [47].

We perform image pre-processing for fMRI data by using a standard pipeline, carried out via Statistical Parametric Mapping (SPM12, www.fil.ion.ucl.ac.uk/spm/software/spm12) software package on Matlab. The data pre-processing procedure includes slice timing, realign, segment, normalization and band-pass filtered. First five volumes of scanned data are discarded to allow magnetization to approach dynamic equilibrium in each participant. Each slice is corrected in slice timing by resampling slices to eliminate time difference. Subsequently, a realignment analysis is performed with middle image of testing sequence as a reference; the data of each participant with a translation exceeding 3 millimetre (mm) and rotation exceeding 3 degree are removed. Individual structural images are linearly co-registered to mean functional image, and then transformed structural images are segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Following this, all functional imaging data is normalized to Montreal Neurological Institutes (MNI) space and re-sampled to $3 \times 3 \times 3\,\text{mm}^3$. Data is detrended and band-pass filtered ($0.01\ \text{Hz} < f < 0.08\ \text{Hz}$) and sources of spurious variance, such as signals from WM, CSF and movement parameters, which are extracted from realignment process, are removed by a linear regression to remove artifacts and reduce physiological noise in CONN toolbox [48].

We adopt an empirical rule to indicate dynamic threshold, which called three-sigma method [49]. It converts a numerical sequence into a state sequence, and represents dynamic threshold as follows:

$$th(T_n^k, \eta) = \mu(T_n^k) + \eta \cdot \sigma(T_n^k) \tag{3}$$

where

$$\mu(T_n^k) = \frac{\sum_{m=1}^{M} t_{n,m}^k}{|T_n^k|} \tag{4}$$

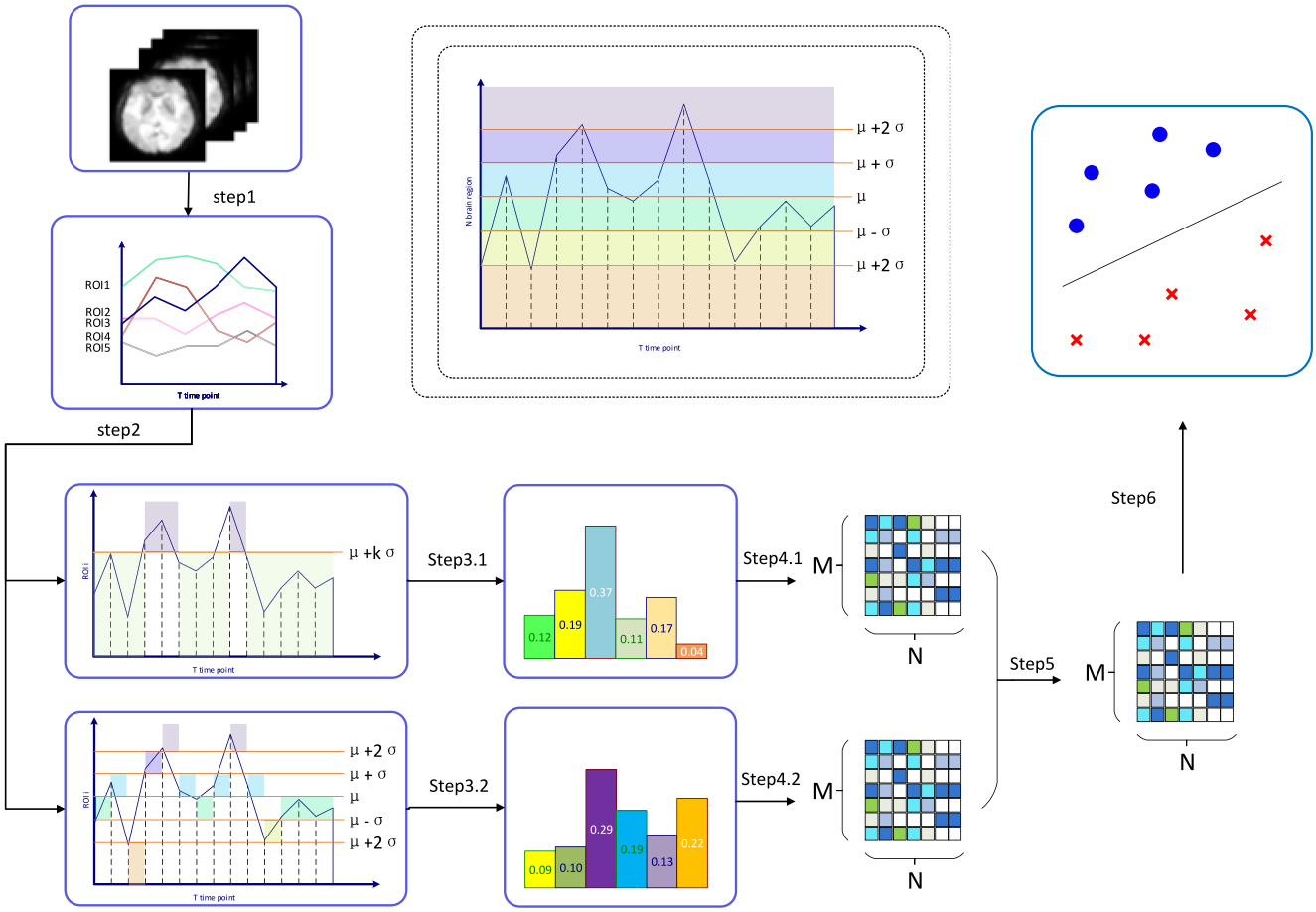

Fig. 1. The flowchart of our proposed time-series kernel-based learning method: (1) image pre-processing and state sequence mapping; (2) calculating correlation value within and between brain regions over time; (3) calculating multi-scale synergy expression probability distribution between brain regions; (4) calculating state transition probability distribution on single brain regions; (5) time-series kernel modeling based on Jensen-Shannon divergence; (6) dealing with brain network analysis and neural disease automated diagnosis.

and

$$\sigma(T_n^k) = \frac{\sum_{m=1}^{M}(t_{n,m}^k - \mu(T_n^k))^2}{|T_n^k| - 1} \tag{5}$$

In multivariate time series A_k, we calculate a corresponding dynamic threshold $th(T_n^k)$ for univariate time series data T_n^k. Therefore, we convert a numerical sequence into a state sequence according to mapping function $f(\cdot)$, as follows:

$$f(t_{n,m}^k, \eta) = \begin{cases} \text{State 0}, & t_{n,m}^k < th(T_n^k, \eta_1) \\ \text{State 1}, & th(T_n^k, \eta_1) \le t_{n,m}^k < th(T_n^k, \eta_2) \\ \cdots & \\ \text{State s}, & th(T_n^k, \eta_s) \le t_{n,m}^k < th(T_n^k, \eta_{s+1}) \\ \cdots & \\ \text{State S}, & th(T_n^k, \eta_S) \le t_{n,m}^k \end{cases} \tag{6}$$

C. Probability Distribution of Multi-Scale Synergy Expression

We extract discrete probability distribution of multi-scale synergy expression between two time series of brain regions. Here, we calculate interactional relation between two brain regions, based on correlation value between brain regions over time.

Firstly, we evaluate temporal dynamic property between two time series data as follows:

$$\phi(t_{n_1,m_1}^{k_1}, t_{n_2,m_2}^{k_2}) = \psi(f(t_{n_1,m_1}^{k_1}, \eta^*), f(t_{n_2,m_2}^{k_2}, \eta^*)) \tag{7}$$

where $f(\cdot)$ represents mapping function and η^* represents mapping parameters. Here, we convert original sequence into two-state sequence, where $\eta = \{0, 1\}$.

In multivariate time series data A_k, correlation value between T_i^k and T_j^k in interval $I_t = [r_t, s_t]$ is defined as follows:

$$C_{\phi(\cdot)}^k(i, j, I_t) = \sum_{m=1}^{M}\sum_{l=r_t}^{s_t} \phi(t_{i,m}^k, t_{j,m+l}^k) \tag{8}$$

where $C_{\phi(\cdot)}^k \in \mathcal{R}^{N\times N\times T}$ and $C_{\phi(\cdot)}^k(i, j, I_t) \neq C_{\phi(\cdot)}^k(j, i, I_t)$.

Then, we propose a discrete probability distribution $P_{\phi(\cdot)}^k$ in multi-scale time series data, as follows:

$$P_{\phi(\cdot)}^k = \{p_{\phi(\cdot)}^k(i, j, I_t) | i, j \in [1, N], I_t \in I\} \tag{9}$$

where $p_{\phi(\cdot)}^k(i, j, I_t)$ represents the proportion of correlation value between i-th time series data and j-th time series data

in interval I_t, as follows:

$$p^k_{\phi(\cdot)}(i,j,I_t) = \frac{C^k_{\phi(\cdot)}(i,j,I_t)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} C^k_{\phi(\cdot)}(i,j,I_t)} \tag{10}$$

D. Probability Distribution of State Transition on Single Brain Region

We extract discrete probability distribution of state transition on a single brain region. Here, we calculate sequential relation on one single brain region, based on correlation value within brain region over time.

Firstly, we mainly calculate probability of one-step state transition of a single brain region, mapping original sequence to a multi-state sequence. The state space is $U = \{u_1, u_2, \ldots, u_e, \ldots, u_E\}$, and the state sequence Π is defined as follows:

$$\Pi^k_i = f(T^k_i, \eta^*) = \{\pi^k_{i,1}, \ldots, \pi^k_{i,m}, \ldots, \pi^k_{i,M}\}, \pi^k_{i,m} \in U \tag{11}$$

where Π^k_i represents the state sequence after mapping of i-th time series in k-th sample, and $\pi^k_{i,m}$ is the state at m-th time point. Thus, we convert original sequence into multi-state sequence, where $\eta = \{-2, -1, 1, 2\}$.

Then, we calculate one-step transition probability P^k_t, as follows:

$$P^k_t = \{p^k_t(e,f,i) | u_e, u_f \in U, i \in [1,N]\} \tag{12}$$

where $p^k_t(e,f,i)$ is the probability of one-step state transition from state u_f to state u_e in i-th sequence, as follows:

$$p^k_t(e,f,i) = \frac{\sum_{m=1}^{M-1}(\pi^k_{i,m} == u_f \ \&\ \pi^k_{i,m+1} == u_e)}{M-1} \tag{13}$$

E. Time-Series Kernel on Jensen-Shannon Divergence

We design a time-series kernel based on Jensen-Shannon divergence to measure similarity of multivariate time series data. For analyzing discrete distribution, we calculate similarity between two probability distributions $P^{k_1}_{\phi(\cdot)}$ and $P^{k_2}_{\phi(\cdot)}$ to measure similarity of two multivariate time series data A_{k_1} and A_{k_2}.

Generally, Kullback-Leibler divergence is a common method to measure how one distribution is different from another distribution. For two discrete probability distributions P and Q, Kullback-Leibler divergence (KLD) from P to Q is defined as follows:

$$D_{KL}(P \parallel Q) = \sum_o P(o) \log \frac{P(o)}{Q(o)} \tag{14}$$

However, Kullback-Leibler divergence is asymmetric and unboundedness. Different from Kullback-Leibler divergence, Jensen-Shannon divergence (JSD) [50], [51] is a symmetrized and smoothed method to measure difference between two discrete probability distributions P and Q, as follows:

$$D_{JS}(P \parallel Q) = \frac{1}{2} D_{KL}(P \parallel M) + \frac{1}{2} D_{KL}(Q \parallel M) \tag{15}$$

where $M = (P+Q)/2$ and $0 \le D_{JS}(P \parallel Q) \le 1$.

For probability distribution set $\{P_1, \ldots, P_D\}$, Kullback-Leibler divergence is generally defined as follows:

$$D_{JS}(P_1, \ldots, P_D) = H\left(\sum_{d=1}^{D} \omega_d P_d\right) - \sum_{d=1}^{D} \omega_d H(P_d) \tag{16}$$

where $\omega_1, \ldots, \omega_D$ are weights for $P_1, \ldots, P_D$, $H(\cdot)$ is Shannon entropy, and $0 \le D_{JS}(P_1, \ldots, P_D) \le \log_2(D)$.

Here, if distribution P is the same as distribution Q, $D_{JS}(P \parallel Q) = 0$; if P is not similar to Q, $D_{JS}(P \parallel Q)$ has a high value and the upper limit is 1. We adopt $1 - D_{JS}(P \parallel Q)$ as kernel to evaluate similarity between P and Q.

Then, we calculate similarity between two discrete probability distributions $P^{k_1}_{\phi(\cdot)}$ and $P^{k_2}_{\phi(\cdot)}$ by JSD, as follows:

$$\begin{aligned}
&D_{JS}(P^{k_1}_{\phi(\cdot)} \parallel P^{k_2}_{\phi(\cdot)}) \\
&= \frac{1}{2} D\left(P^{k_1}_{\phi(\cdot)} \parallel \frac{P^{k_1}_{\phi(\cdot)} + P^{k_2}_{\phi(\cdot)}}{2}\right) \\
&\quad + \frac{1}{2} D\left(P^{(k_2)}_{\phi(\cdot)} \parallel \frac{P^{k_1}_{\phi(\cdot)} + P^{k_2}_{\phi(\cdot)}}{2}\right) \\
&= \frac{1}{2} \sum p^{k_1}_{\phi(\cdot)} \left(\log p^{k_1}_{\phi(\cdot)} - \log \frac{p^{k_1}_{\phi(\cdot)} + p^{k_2}_{\phi(\cdot)}}{2}\right) \\
&\quad + \frac{1}{2} \sum p^{k_2}_{\phi(\cdot)} \left(\log p^{k_2}_{\phi(\cdot)} - \log \frac{p^{k_1}_{\phi(\cdot)} + p^{k_2}_{\phi(\cdot)}}{2}\right) \\
&= \frac{1}{2} \left(\sum p^{k_1}_{\phi(\cdot)} \log p^{k_1}_{\phi(\cdot)} + \sum p^{k_2}_{\phi(\cdot)} \log p^{k_2}_{\phi(\cdot)}\right) \\
&\quad - \sum \frac{p^{k_1}_{\phi(\cdot)} + p^{k_2}_{\phi(\cdot)}}{2} \log \frac{p^{k_1}_{\phi(\cdot)} + p^{k_2}_{\phi(\cdot)}}{2}
\end{aligned} \tag{17}$$

Finally, we measure time-series kernel on multivariate time series data A_{k_1} and A_{k_2}, as follows:

$$\mathcal{K}_{\phi(\cdot)}(A_{k_1}, A_{k_2}) = 1 - D_{JS}(P^{k_1}_{\phi(\cdot)} \parallel P^{k_2}_{\phi(\cdot)}) \tag{18}$$

$$\mathcal{K}_t(A_{k_1}, A_{k_2}) = 1 - D_{JS}(P^{k_1}_t \parallel P^{k_2}_t) \tag{19}$$

$$\mathcal{K} = \alpha \cdot \mathcal{K}_{\phi(\cdot)} + (1-\alpha) \cdot \mathcal{K}_t \tag{20}$$

where α is the harmonic coefficient.

F. Kernel-Based Learning

Based on above feature extraction methods, we construct corresponding customized kernels, respectively. We adopt Support Vector Machine (SVM) [52] for binary classification. The decision function is shown as follows:

$$\gamma(A_k) = \text{sign}\left\{\sum_{i=1}^{K} \alpha_i y_i \cdot \mathcal{K}(A_k, A_i) + b\right\} \tag{21}$$

Algorithm 1: Multi-Scale Time-Series Kernel.

Input: Multivariate time series $A_g = \{T_1^g, \ldots, T_n^g, \ldots, T_N^g\}$ and $A_h = \{T_1^h, \ldots, T_n^h, \ldots, T_N^h\}$, where $T_n^g = \{t_{n,1}^g, \ldots, t_{n,m}^g, \ldots, t_{n,M}^g\}$ and $T_n^h = \{t_{n,1}^h, \ldots, t_{n,m}^h, \ldots, t_{n,M}^h\}$; Interval collection $I = \{I_1, \ldots, I_t, \ldots, I_T\}$, where $I_t = [r_t, s_t]$;

Output: Time-series kernel $\mathcal{K}(A_g, A_h)$.

function Multi-Scale Time-Series KernelA_g, A_h
  for $i = 1 : N$ **do**
    for $j = 1 : N$ **do**
      for $t = 1 : T$ **do**
        Calculate correlation value $C_{\phi(\cdot)}(i, j, I_t)$ between T_i and T_j in interval I_t:
        $C_{\phi(\cdot)}^g(i, j, I_t) = \sum_{m=1}^{M} \sum_{l=r_t}^{s_t} \phi(t_{i,m}^g, t_{j,m+l}^g)$,
        $C_{\phi(\cdot)}^h(i, j, I_t) = \sum_{m=1}^{M} \sum_{l=r_t}^{s_t} \phi(t_{i,m}^h, t_{j,m+l}^h)$;
      end for
    end for
    for $t = 1 : T$ **do**
      Calculate discrete probability distribution P_t of one-step transition from u_f to u_e in i-th sequence:
      $p_t^g(e, f, i) = \frac{\sum_{m=1}^{M-1}(\pi_{i,m}^g = u_f \ \&\ \pi_{i,m+1}^g = u_e)}{M-1}$,
      $p_t^h(e, f, i) = \frac{\sum_{m=1}^{M-1}(\pi_{i,m}^h = u_f \ \&\ \pi_{i,m+1}^h = u_e)}{M-1}$;
    end for
  end for
  Calculate discrete probability distribution $P_{\phi(\cdot)}$ between i-th and j-th time series data in interval I_t:
  $p_{\phi(\cdot)}^g(i, j, I_t) = \frac{C_{\phi(\cdot)}^g(i,j,I_t)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} C_{\phi(\cdot)}^g(i,j,I_t)}$,
  $p_{\phi(\cdot)}^h(i, j, I_t) = \frac{C_{\phi(\cdot)}^h(i,j,I_t)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} C_{\phi(\cdot)}^h(i,j,I_t)}$;
  Calculate similarity between two discrete probability distributions:
  $D_{JS}(P_{\phi(\cdot)}^g \parallel P_{\phi(\cdot)}^h) = \frac{1}{2}\sum p_{\phi(\cdot)}^g \log p_{\phi(\cdot)}^g + \frac{1}{2}\sum p_{\phi(\cdot)}^h \log p_{\phi(\cdot)}^h - \sum \frac{p_{\phi(\cdot)}^g + p_{\phi(\cdot)}^h}{2} \log \frac{p_{\phi(\cdot)}^g + p_{\phi(\cdot)}^h}{2}$
  $D_{JS}(P_t^g \parallel P_t^h) = \frac{1}{2}\sum p_t^g \log p_t^g + \frac{1}{2}\sum p_t^h \log p_t^h - \sum \frac{p_t^g + p_t^h}{2} \log \frac{p_t^g + p_t^h}{2}$
  Obtain time-series kernel on multivariate time series data:
  $\mathcal{K}_{\phi(\cdot)}(A_g, A_h) = 1 - D_{JS}(P_{\phi(\cdot)}^g \parallel P_{\phi(\cdot)}^h)$
  $\mathcal{K}_t(A_g, A_h) = 1 - D_{JS}(P_t^g \parallel P_t^h)$
  $\mathcal{K}(A_g, A_h) = \alpha \cdot \mathcal{K}_{\phi(\cdot)} + (1 - \alpha) \cdot \mathcal{K}_t$
end Function

where $\mathcal{K}(A_k, A_i)$ represents time-series kernel function, and α_i is calculated as follows:

$$\begin{aligned} \max \quad & \sum_{i=1}^{K} \alpha_i - \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_i \alpha_j \cdot y_i y_j \cdot \mathcal{K}(A_i, A_j) \\ \text{s.t.} \quad & 0 \le \alpha_i \le C \\ & \sum_{i=1}^{K} \alpha_i \gamma_i = 0 \end{aligned} \tag{22}$$

For clarification, pseudo-code for Multi-Scale Time-Series Kernel is summarized in Algorithm 1.

IV. Experiments

A. Dataset

Our experiment is applied on two datasets. One is a public Alzheimer's Disease Neuroimaging Initiative database [53], and another is a volunteer experiment of Major Depressive Disorder [54]. In the process of data pre-processing, we deal with raw data by a widely used software package (SPM12), and then divide whole-brain into multiple brain regions based on anatomical template for analysis.

1) ADNI: Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for early detection and tracking of Alzheimer's disease (AD). In Alzheimer's Disease Neuroimaging Initiative database, we apply a total of 169 subjects, including 87 Alzheimer's patients (49 females and 38 males) and 82 normal controls (46 females and 36 males). We download ADNI data from website http://adni.loni.usc.edu/.

2) MDD: In volunteer experiment, we use a total of 60 subjects, including 31 volunteers with Major Depressive Disorder (MDD) (22 females and 9 males, aged 50.5 ± 11.2 years, range $25 - 65$ years) and 29 healthy volunteers (18 females and 11 males, aged 50.1 ± 10.6 years, range $25 - 65$ years). Among those major depressive disorder subjects without comorbidity, shortest duration of illness is more than three months.

B. Brain Anatomical Template

We use anatomical templates to divide whole-brain into Regions Of Interest (ROI). In our work, we use three different templates for comparison, including Automated Anatomical Labeling (AAL) template, Harvard-Oxford template and Brainnetome template. When we use templates, for each regions of interest, mean time series is calculated by averaging Blood-Oxygen-Level-Dependent (BOLD) signals among all voxels within specifically ROI.

1) AAL: Automated Anatomical Labeling (AAL) template is a widely used anatomical template, which divides whole brain into 78 cortical regions, 26 cerebellar regions and 12 subcortical regions according to anatomy [45].

2) Harvard-Oxford Template: Harvard-Oxford atlas covering 48 cortical and 21 subcortical structural areas [55]–[57].

3) Brainnetome Template: Brainnetome template contains more fine-grained functional brain subregions and gives more detailed anatomical information compared with AAL, because it is generated with both functional connectivity and anatomical information [46].

C. Evaluation Criterion

For evaluation of prediction performance, we use threshold-dependent parameters such as Accuracy, Sensitivity and Specificity. There are calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{23}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{24}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{25}$$

where TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively.

What's more, the Area Under ROC curve (AUC) is employed to evaluate our predictive model.

D. Interval Collection I

One of main tasks of this article is to discuss asynchronous functional relationships between brain regions in the human brain. Interval can represent imaging unit from imaging device. We use interval to extract discrete probability distribution of multi-scale synergy expression between two time series of brain regions. Here, "interval" is a flexible concept that we will discuss and analyze it as a parameter.

In our study, $I = \{I_1, I_2, \ldots, I_t, \ldots, I_T\}$ represents a collection of multiple intervals, where $I_t = [r_t, s_t]$ denotes a positive integer interval. The parameter settings must be conform to research hypothesis; for a interval $I_t \in I$, if I_t is close to zero, it means that we extract short-distance asynchronous information; if I_t is far from zero, it indicates that we extract long-distance asynchronous information. In our experiments, we set interval collection I as $\{[0, 0], [1, 1], [2, 2], [3, 12]\}$. Here, [0, 0] represents information for synchronization; [1, 1] and [2, 2] represent short-distance correlation for asynchronization; [3, 12] represents long-distance correlation for asynchronization. Preferably, it is a reasonable way to set parameters in proportion to the length of interval.

TABLE I
DIFFERENT PERFORMANCE OF VARIOUS P-VALUE PARAMETERS ON ALZHEIMER'S DISEASES VIA AAL TEMPLATE

P-value	No. of Feature	Accuracy	Sensitivity	Specificity	AUC
0.05	1608	0.7502	0.7214	0.7815	0.7149
0.01	549	0.7924	0.7628	0.8245	0.7601
0.005	274	**0.8166**	**0.7931**	**0.8415**	**0.7894**
0.001	84	0.6657	0.6254	0.7115	0.6411

*Our novel functional connectivity is extracted as state distribution within (ST: state transition) and between (SE: synergy expression) brain regions.

V. RESULTS AND DISCUSSION

To verify effectiveness of our method, we conduct comparative experiments on three different brain region templates for analyzing two brain diseases. Furthermore, we discuss characteristics of brain regions from perspective of disease diagnosis.

A. Analysis of Feature Extraction

In traditional brain network model, Pearson's Correlation Coefficient (PCC) can be used to calculate brain region correlation and functional connectivity strength. Traditional PCC is order-independent in time series, but lacks contextual and sequential information, which reduces diagnostic capability of brain disease.

For processing fMRI image data, we calculate state distribution within (ST: state transition) and between (SE: synergy expression) brain regions over time, which can explore pathological principles of mental diseases through difference of state distribution. We analyze different performance of various P-value parameters on Alzheimer's Diseases via AAL template, as shown in Table I. The number of features decreases rapidly as P-value going down. When P-value being 0.005, the number of features ranges in a reasonable scope, and more accurate results can be also obtained.

Here, our model is compared with PCC as feature extraction method for analyzing functional connectivity. In our experiments, we test on two data sets with three different brain templates, as shown in Table II. We select features via t-test with confidence level p-value less than 0.005, and evaluate performance via leave-one-out cross-validation (LOOCV). On ADNI and MDD data, the performance of our model (SE+ST) in different brain region templates are better than that of traditional PCC model. It indicates that our model can extract more effective information of fMRI data for disease diagnosis and clinical application.

B. Compared With Kernel Models

We adopt kernel model based on Jensen-Shannon Divergence (JSD) to construct new feature space. Here, we apply JSD time series kernel on synergy expression distribution between brain regions (K_SE) and state transition probability distribution

TABLE II
COMPARISON WITH PEARSON'S CORRELATION COEFFICIENT (PCC) AS FEATURE EXTRACTION MODEL

Disease	Brain Template	Model	Accuracy	Sensitivity	Specificity	AUC
AD	AAL	PCC	0.6449	0.6092	0.6829	0.6098
		SE+ST	**0.8166**	**0.7931**	**0.8415**	**0.7894**
	Brainnetome	PCC	0.6746	0.6552	0.6951	0.6401
		SE+ST	**0.8402**	**0.8391**	**0.8415**	**0.8027**
	Harvard-Oxford	PCC	0.6508	0.5862	0.7195	0.5982
		SE+ST	**0.7811**	**0.7356**	**0.8293**	**0.7561**
MDD	AAL	PCC	0.6167	0.6129	0.6207	0.5935
		SE+ST	**0.8000**	**0.7742**	**0.8275**	**0.8082**
	Brainnetome	PCC	0.5500	0.5806	0.5172	0.5667
		SE+ST	**0.7500**	**0.7419**	**0.7586**	**0.7751**
	Harvard-Oxford	PCC	0.5500	0.5161	0.5862	0.5611
		SE+ST	**0.7333**	**0.6774**	**0.7931**	**0.7401**

*Our novel functional connectivity is extracted as state distribution within (ST: state transition) and between (SE: synergy expression) brain regions.

TABLE III
COMPARISON WITH DIFFERENT KERNEL MODELS AND DYNAMIC FUNCTIONAL CORRELATION METHODS

Disease	Brain Template	Method	Accuracy	Sensitivity	Specificity	AUC
AD	AAL	SE+ST	0.8166	0.7931	0.8415	0.7894
		K_ST	0.7456	0.7586	0.7317	0.7114
		K_SE	0.8698	0.8390	0.9024	0.8267
		K_SE+ST	**0.8876**	**0.8506**	**0.9268**	**0.8562**
	Brainnetome	SE+ST	0.8402	0.8391	0.8415	0.8027
		K_ST	0.8343	0.8161	0.8537	0.8114
		K_SE	0.8639	0.8621	0.8659	0.8485
		K_SE+ST	**0.8935**	**0.8851**	**0.9024**	**0.8802**
	Harvard-Oxford	SE+ST	0.7811	0.7356	0.8293	0.7561
		K_ST	0.6982	0.7471	0.6463	0.6485
		K_SE	0.8047	0.7931	0.8171	0.7714
		K_SE+ST	**0.8402**	**0.8046**	**0.8780**	**0.7824**
MDD	AAL	SE+ST	0.8000	0.7742	0.8275	0.8082
		K_ST	0.5833	0.6452	0.5172	0.5914
		K_SE	0.8667	0.8065	**0.9310**	0.8851
		K_SE+ST	**0.9000**	**0.8710**	**0.9310**	**0.9295**
	Brainnetome	SE+ST	0.7500	0.7419	0.7586	0.7751
		K_ST	0.6833	0.6452	0.7241	0.7168
		K_SE	0.7667	**0.8387**	0.6897	0.7614
		K_SE+ST	**0.8500**	**0.8387**	**0.8621**	**0.8647**
	Harvard-Oxford	SE+ST	0.7333	0.6774	0.7931	0.7401
		K_ST	0.6500	0.7097	0.5862	0.6628
		K_SE	**0.8500**	0.8387	**0.8621**	**0.8705**
		K_SE+ST	0.8333	**0.8709**	0.7931	0.8587

*Our novel functional connectivity is extracted as state distribution within (ST: state transition) or/and between (SE: synergy expression) brain regions.

within single brain regions (K_ST), respectively. Also, we apply JSD time series kernel on combined features (K_SE+ST).

In our experiments, we compare three kernel models with above feature-based model (SE+ST), as shown in Table III. On ADNI and MDD data, the performance of combined kernel (K_SE+ST) in different brain region templates are better than that of feature-based model (SE+ST). Accuracy of synergy expression kernel is higher than that of state transition kernel. It indicates that state transition information can be complementary to synergy expression information in whole brain regions.

C. Compared With Existing Methods

We compare our multi-scale time-series kernel-based learning method with many existing methods, such as traditional graph feature method (Baseline), Weisfeiler-Lehman graph kernel framework (WL-edge, WL-subtree and WL-shortestpath) [27], shortest-path (Shortest-path) [36], sliding window method (FON) [3], sub-network kernel method (SKL) [43], and method of Xu *et al.* [54].

TABLE IV
COMPARISON OF OUR METHOD AND SEVEN EXISTING METHODS ON ADNI

Method	Accuracy	Sensitivity	Specificity	AUC
Baseline	0.5858	0.5747	0.5976	0.5612
WL-edge	0.6272	0.6437	0.6098	0.6084
WL-subtree	0.7811	0.7816	0.7805	0.7645
WL-Shortestpath	0.6095	0.5977	0.6220	0.5735
Shortest-path	0.7396	0.8161	0.6585	0.6938
FON	0.8580	0.8161	0.9024	0.8195
SKL	0.8462	0.8046	0.8902	0.8166
Our Method	**0.8876**	**0.8506**	**0.9268**	**0.8562**

TABLE V
COMPARISON OF OUR METHOD AND THREE EXISTING METHODS ON MDD

Method	Accuracy	Sensitivity	Specificity	AUC
Baseline	0.6167	0.6129	0.6207	0.6514
Shortest-path	0.7833	0.8065	0.7586	0.8135
Xu et al.	0.8667	**0.8710**	0.8621	0.9103
Our Method	**0.9000**	**0.8710**	**0.9310**	**0.9295**

TABLE VI
IMPORTANT ROIS SELECTED BY OUR METHOD ON AAL TEMPLATE

Region 1	Region 2	State	Distance
Pallidum_L	Frontal_Inf_Tri_R	syn	—
Frontal_Inf_Tri_R	Pallidum_L	syn	—
Frontal_Mid_Orb_L	Calcarine_R	syn	—
Frontal_Mid_Orb_L	Heschl_L	asy	short
Frontal_Mid_Orb_R	Thalamus_L	asy	short
Frontal_Sup_L	Hippocampus_R	asy	short
Amygdala_L	Frontal_Inf_Oper_R	asy	short
Frontal_Mid_Orb_R	Thalamus_L	asy	short
Rectus_L	Frontal_Sup_Medial_L	asy	short
Rectus_L	Lingual_R	asy	short
Paracentral_Lobule_L	Cingulum_Post_R	asy	long
Frontal_Sup_Medial_L	Hippocampus_R	asy	long
Cingulum_Post_L	Pallidum_R	asy	long
Frontal_Mid_Orb_L	Heschl_L	asy	long
Caudate_L	Hippocampus_L	asy	long
Rectus_L	Lingual_L	asy	long
Thalamus_R	Amygdala_L	asy	long
Heschl_L	Frontal_Mid_Orb_L	asy	long

*State means synchronization (syn) or asynchronization (asy); Distance means interval with short-distance or long-distance.

1) Performance on ADNI: On ADNI dataset, our method is compared to seven existing methods, as shown in Table IV. Our method obtains best accuracy of 0.8876 and best AUC of 0.8562. It achieves accuracy improvement of 0.0294 and AUC improvement of 0.0358. It indicates that our method is far superior to some traditional graph feature methods, and a little superior than many outstanding graph kernel methods

2) Performance on MDD: On MMD dataset, our method is compared with three existing methods, as shown in Table V. Our method obtains best accuracy of 0.9000 and best AUC of 0.9295. It achieves accuracy improvement of 0.0333 and AUC improvement of 0.0192. It indicates that our method is much better than some traditional graph theory methods, and slightly better than many current state-of-the-art methods.

D. Important Brain Regions Associated With AD

In this section, we investigate importance of ROIs (brain regions) associated with AD. Here, we analyze important ROIs on AAL template that is a widely used anatomical template.

We apply a statistical analysis of multi-scale state probability distribution between two brain regions. Furthermore, we can select some important ROIs according to their classification performance on above extracted feature set, as shown in Table VI. Important ROIs have been selected by our statistical method on AAL template, including hippocampus [58], [59], cingulate [59], amygdala [60] and heschl gyrus [61], [62].

VI. Conclusion

In this paper, our proposed method makes following several contributions. Firstly, we propose a statistical analysis model based on multi-scale state probability distribution between two brain regions and state transition probability distribution of a single brain region. This model can effectively and accurately analyze differences and pathological changes of mental disorders in fMRI imaging. Secondly, we apply above two probability distributions as feature vector on classical classification methods, in order to obtain efficient performance. In addition, as a similarity measure applied on the brain network model, edge weight calculation technique on JSD has a better performance than traditional PCC since we consider context information of fMRI data. Finally, we build a neural disease diagnosis framework, which has been verified on ADNI and MDD data and shows excellent diagnostic ability.

References

[1] E. Damaraju *et al.*, "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *Neuroimage Clin.*, vol. 5, no. C, pp. 298–308, 2014.

[2] C. Y. Wee, S. Yang, P. T. Yap, and D. Shen, "Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification," *Brain Imag. Behav.*, vol. 10, no. 2, pp. 342–356, 2016.

[3] X. Chen, H. Zhang, Y. Gao, C. Y. Wee, G. Li, and D. Shen, "High-order resting-state functional connectivity network for MCI classification," *Human Brain Mapping*, vol. 37, no. 9, pp. 3282–3296, 2016.

[4] Z. Nianyin, W. Zidong, L. Yurong, D. Min, C. Jie, and L. Xiaohui, "Time series modeling of nano-gold immunochromatographic assay via expectation maximization algorithm," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3418–3424, Dec. 2013.

[5] Z. Nianyin *et al.*, "Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach," *IEEE Trans. Med. Imag.*, vol. 33, no. 5, pp. 1129–1136, May 2014.

[6] F. Li, L. Tran, K. H. Thung, S. Ji, D. Shen, and J. Li, "A robust deep model for improved classification of AD/MCI patients," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 5, pp. 1610–1616, Sep. 2015.

[7] D. Štern, C. Payer, N. Giuliani, and M. Urschler, "Automatic age estimation and majority age classification from multi-factorial MRI data," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1392–1403, Jul. 2019.

[8] M. Liu, J. Zhang, D. Nie, P. T. Yap, and D. Shen, "Anatomical landmark based deep feature representation for MR images in brain disease diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1476–1485, Sep. 2018.

[9] K. J. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," *Brain Connect*, vol. 1, no. 1, pp. 13–36, 2011.

[10] T. Zhou, K. H. Thung, M. Liu, and D. Shen, "Brain-wide genome-wide association study for alzheimer's disease via joint projection learning and sparse regression model," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 1, pp. 165–175, Jan. 2019.

[11] D. Zhou, W. K. Thompson, and G. Siegle, "Matlab toolbox for functional connectivity," *Neuroimage*, vol. 47, no. 4, pp. 1590–1607, 2009.

[12] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 758–769, Apr. 2018.

[13] K. J. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *Neuroimage*, vol. 19, no. 4, pp. 1273–1302, 2003.

[14] S. Sun *et al.*, "Graph theory analysis of functional connectivity in major depression disorder with high-density resting state EEG data," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 429–439, Mar. 2019.

[15] B. M. Tijms *et al.*, "Alzheimer's disease: Connecting findings from graph theoretical studies of brain networks," *Neurobio. Aging*, vol. 34, no. 8, pp. 2023–2036, 2013.

[16] P. Gonzalez-Navarro, Y. M. Marghi, B. Azari, M. Akcakaya, and D. Erdogmus, "An event-driven ar-process model for EEG-based BCIs with rapid trial sequences," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 798–804, May 2019.

[17] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[18] G. Chen *et al.*, "Classification of alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging," *Int. J. Med. Radiol.*, vol. 259, no. 1, pp. 213–21, 2011.

[19] M. Liu, Y. Gao, P. T. Yap, and D. Shen, "Multi-hypergraph learning for incomplete multi-modality data," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1197–1208, Jul. 2018.

[20] E. J. Sanz-Arigita *et al.*, "Loss of 'small-world' networks in alzheimer's disease: Graph analysis of FMRI resting-state functional connectivity," *PlOs One*, vol. 5, no. 11, 2010, Art. no. e13788.

[21] C. Lin, B.-H. Wang, N. Jiang, R. Xu, N. Mrachacz-Kersting, and D. Farina, "Discriminative manifold learning based detection of movement-related cortical potentials," *IEEE Trans. Neural Syst.Rehabil. Eng.*, vol. 24, no. 9, pp. 921–927, Sep. 2016.

[22] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral Cortex*, vol. 24, no. 3, pp. 663–676, 2014.

[23] R. M. Hutchison *et al.*, "Dynamic functional connectivity: Promise, issues, and interpretations," *Neuroimage*, vol. 80, no. 1, pp. 360–378, 2013.

[24] N. Leonardi *et al.*, "Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest," *Neuroimage*, vol. 83, no. 4, pp. 937–950, 2013.

[25] W. Li, L. Qiao, L. Zhang, Z. Wang, and D. Shen, "Functional brain network estimation with time series self-scrubbing," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 6, pp. 2494–2504, Nov. 2019.

[26] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, and Y. Li, "A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of alzheimer's disease," *Neurocomputing*, vol. 320, pp. 195–202, 2018.

[27] N. Shervashidze, P. Schweitzer, E. J. V. Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-Lehman graph kernels," *J. Mach. Learn. Res.*, vol. 12, no. 3, pp. 2539–2561, 2011.

[28] R. I. Kondor and J. D. Lafferty, "Diffusion kernels on graphs and other discrete input spaces," in *Proc. Int. Conf. Mach. Learn.*, 2002, pp. 315–322.

[29] T. G. T. Horvath and S. Wrobel, "Cyclic pattern kernels for predictive graph mining," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 158–167.

[30] Y. Ding, J. Tang, and F. Guo, "Identification of drug-side effect association via multiple information integration with centered kernel alignment," *Neurocomput.*, vol. 325, pp. 211–224, 2019.

[31] Y. Shen, J. Tang, and F. Guo, "Identification of protein subcellular localization via integrating evolutionary and physicochemical information into chou's general pseaac," *J. Theor. Biol.*, vol. 462, pp. 230–239, 2019.

[32] L. Jiang, Y. Xiao, Y. Ding, J. Tang, and F. Guo, "FKL-Spa-LapRLS: An accurate method for identifying human microrna-disease association," *BMC Genomics*, vol. 19, no. 10, 2018, Art. no. 911.

[33] C. Shen, Y. Ding, J. Tang, L. Jiang, and F. Guo, "LPI-KTASLP: Prediction of lncrna-protein interaction by semi-supervised link learning with multivariate information," *IEEE Access*, vol. 7, pp. 13 486–13 496, 2019.

[34] L. Jiang, Y. Ding, J. Tang, and F. Guo, "MDA-SKF: Similarity kernel fusion for accurately discovering mirna-disease association," *Front. Genet.*, vol. 9, 2018, doi: 10.3389/fgene.2018.00618.

[35] D. Haussler, "Convolution kernels on discrete structures," *Tech Rep*, vol. 7, pp. 95–114, 1999.

[36] K. M. Borgwardt and H. P. Kriegel, "Shortest-path kernels on graphs," in *Proc. IEEE Int. Conf. Data Mining*, 2006, pp. 74–81.

[37] H. Kashima, "Marginalized kernels between labeled graphs," in *Proc. Int. Conf. Mach. Learn.*, 2003, pp. 321–328.

[38] J. Ricard and G. Noat, "Expressivity versus efficiency of graph kernels," in *Proc. 1st Int. Workshop Mining Graphs Trees Sequences*, 2003, vol. 109, pp. 65–74.

[39] F. D. Johansson, V. Jethava, D. Dubhashi, and C. Bhattacharyya, "Global graph kernels using geometric embeddings," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 694–702.
[40] L. Xiang, Y. Chen, W. Chang, Y. Zhan, and D. Shen, "Deep leaning based multi-modal fusion for fast MR reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 2105–2114, 2019.
[41] T. M. Thomas, D. N. Candrea, M. S. Fifer, D. P. Mcmullen, and N. E. Crone, "Decoding native cortical representations for flexion and extension at upper limb joints using electrocorticography," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 2, pp. 293–303, Feb. 2019.
[42] A. Shrivastava and P. Li, "A new space for comparing graphs," in *Proc. IEEE/ACM Int. Conf. Advances Soc. Netw. Anal. Mining*, 2014, pp. 62–71.
[43] B. Jie, M. Liu, D. Zhang, and D. Shen, "Sub-network kernels for measuring similarity of brain connectivity networks in disease diagnosis," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2340–2353, May 2018.
[44] K. J. Worsley, S. Marrett, P. Neelin, and A. C. Evans, "Chapter 64 a unified statistical approach for determining significant signals in location and scale space images of cerebral activation," *Human Brain Mapping*, vol. 4, no. 1, pp. 58–73, 2015.
[45] N. Tzouriomazoyer *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.
[46] L. Fan *et al.*, "The human brainnetome atlas: A new brain atlas based on connectional architecture," *Cerebral Cortex*, vol. 26, no. 8, pp. 3508–3526, 2016.
[47] Z. Zhang, J. Xu, J. Tang, Q. Zou, and F. Guo, "Diagnosis of brain diseases via multi-scale time-series model," *Front. Neurosci.*, vol. 13, 2019, doi: 10.3389/fnins.2019.00197.
[48] S. Whitfield-Gabrieli and A. Nieto-Castanon, "*Conn*: A functional connectivity toolbox for correlated and anticorrelated brain networks," *Brain Connect.*, vol. 2, no. 3, pp. 125–141, 2012.
[49] WalterA, *Statistical Method From the Viewpoint of Quality Control*. New York, NY, USA: Dover, 1986.
[50] B. Fuglede and F. Topsoe, "Jensen-shannon divergence and hilbert space embedding," in *Proc. Inf. Theory Int. Symp.*, 2005, doi: 10.1109/ISIT.2004.1365067.
[51] D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1858–1860, Jul. 2003.
[52] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[53] C. R. Jack *et al.*, "The alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Resonance Imag.*, vol. 27, no. 4, pp. 685–691, 2010.
[54] X. Geng, J. Xu, B. Liu, and Y. Shi, "Multivariate classification of major depressive disorder using the effective connectivity and functional connectivity," *Front. Neurosci.*, vol. 12, 2018, doi: 10.3389/fnins.2018.00038.
[55] N. Makris *et al.*, "Decreased volume of left and total anterior insular lobule in schizophrenia," *Schizophrenia Res.*, vol. 83, no. 2, pp. 155–171, 2006.
[56] M. W. Woolrich *et al.*, "Bayesian analysis of neuroimaging data in FSL," *Neuroimage*, vol. 45, pp. 0–0, 2009.
[57] S. M. Smith *et al.*, "Advances in functional and structural MR image analysis and implementation as FSL," *Neuroimage*, vol. 23, pp. S208–S219, 2004.
[58] F. Bai *et al.*, "Abnormal functional connectivity of hippocampus during episodic memory retrieval processing network in amnestic mild cognitive impairment," *Biol. Psychiatry*, vol. 65, no. 11, pp. 951–958, 2009.
[59] Y. Zhou, J. H. Dougherty Jr, K. F. Hubner, B. Bai, R. L. Cannon, and R. K. Hutson, "Abnormal connectivity in the posterior cingulate and hippocampus in early alzheimer's disease and mild cognitive impairment," *Alzheimer's Dementia*, vol. 4, no. 4, pp. 265–270, 2008.
[60] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiology Aging*, vol. 32, no. 12, pp. e19–e27, 2011.
[61] J. Soldner, T. Meindl, W. Koch, A. L. W. Bokde, and S. J. Teipel, "Structural and functional neuronal connectivity in alzheimer's disease: A combined DTI and FMRI study," *Nervenarzt*, vol. 83, no. 7, pp. 878–887, 2011.
[62] Y. Han *et al.*, "Frequency-dependent changes in the amplitude of low-frequency fluctuations in amnestic mild cognitive impairment: A resting-state FMRI study," *NeuroImage*, vol. 55, no. 1, pp. 287–295, 2011.