Received: 6 April 2022 | Revised: 19 May 2022 | Accepted: 8 June 2022

DOI: 10.1002/hbm.25994

RESEARCH ARTICLE

WILEY

Spatial–temporal graph convolutional network for Alzheimer classification based on brain functional connectivity imaging of electroencephalogram

Xiaocai Shan[1,2] | Jun Cao[2] | Shoudong Huo[1] | Liangyu Chen[3] | Ptolemaios Georgios Sarrigiannis[4] | Yifan Zhao[2]

[1]Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China

[2]School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, UK

[3]Department of Neurosurgery, Shengjing Hospital of China Medical University, Shenyang, China

[4]Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

Correspondence
Yifan Zhao, School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield MK43 0AL, UK.
Email: yifan.zhao@cranfield.ac.uk

Abstract

Functional connectivity of the human brain, representing statistical dependence of information flow between cortical regions, significantly contributes to the study of the intrinsic brain network and its functional mechanism. To fully explore its potential in the early diagnosis of Alzheimer's disease (AD) using electroencephalogram (EEG) recordings, this article introduces a novel dynamical spatial–temporal graph convolutional neural network (ST-GCN) for better classification performance. Different from existing studies that are based on either topological brain function characteristics or temporal features of EEG, the proposed ST-GCN considers both the adjacency matrix of functional connectivity from multiple EEG channels and corresponding dynamics of signal EEG channel simultaneously. Different from the traditional graph convolutional neural networks, the proposed ST-GCN makes full use of the constrained spatial topology of functional connectivity and the discriminative dynamic temporal information represented by the 1D convolution. We conducted extensive experiments on the clinical EEG data set of AD patients and Healthy Controls. The results demonstrate that the proposed method achieves better classification performance (92.3%) than the state-of-the-art methods. This approach can not only help diagnose AD but also better understand the effect of normal ageing on brain network characteristics before we can accurately diagnose the condition based on resting-state EEG.

KEYWORDS

artificial intelligence, brain association, electroencephalogram, graph convolutional neural network, machine learning

1 | INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia, resulting in the loss of memory and cognitive impairments. Most commonly it occurs among elderly people over the age of 65, but it may also occur earlier (Brookmeyer et al., 1998, 2007). According to the World Health Organization, more than 55 million people are currently diagnosed with some type of dementia (World Health Organization, 2021). Due to increasing global ageing, the technologies to diagnose AD more effectively and accurately are highly demanded. Numerous studies have shown that synaptic dysfunction is an early feature of AD and that decreased synaptic density in the neocortex

and limbic regions could account for AD-associated disturbances in brain function (Masliah et al., 2001; Scheff et al., 2006, 2007; Terry et al., 1991). Being more extensive than the corresponding neuronal loss when analysed in the same brain regions, synaptic dysfunction has been proved to be the best neuropathological correlate of cognitive impairment in AD patients (Davies et al., 1987; DeKosky & Scheff, 1990; Scheff et al., 2007; Terry et al., 1991). The pathological progression of AD leads to cortical disconnections and manifests as functional connectivity alterations (Nobukawa et al., 2020). Electroencephalography (EEG) is a noninvasive diagnostic method for studying the bioelectrical function alterations and degeneration of the brain. Consisting of scalp electric potential differences, EEG is one of the first measurements that directly reflect the functioning of synapses in real-time (Jelic, 2005; Michel et al., 2009). In contrast to functional MRI or PET which detect indirect metabolic signals, EEG offers several additional attractions: noninvasiveness, high time-resolution, wide availability, low cost, and direct access to neuronal signalling (Michel et al., 2009; Smailovic et al., 2018).

For the diagnosis of AD based on EEG, there are two main development areas in recent years. The first area is statistically topological brain function characteristics. Finding statistical biomarkers of AD is based on analysing graph topographic or dynamic patterns. For graph topography, Jalili (2017) constructed functional brain networks and relevant graph theory metrics based on EEG for discriminating AD from Healthy Controls (HC). Similarly, Tylová et al. (2018) proposed the permutation entropy for measuring the chaotic behaviour of EEG and observed a statistically significant decrease in permutation entropy at all channels of AD. Fan et al. (2018) employed multiscale entropy as the biomarker to characterize the nonlinear complexity at multiple temporal scales to capture the topographic pattern of AD. For dynamic patterns, Zhao et al. (2019) proposed a method to measure nonlinear dynamics of functional connectivity for distinguishing between AD and HC, and Tait et al. (2020) developed a biomarker by combining microstate transitioning complexity and the spectral measure. The second area is data-driven machine learning or deep learning methods with various input features for the AD classification. Researchers have proven that SVM (Nobukawa et al., 2020; Song et al., 2018; Tavares et al., 2019; Trambaiolli et al., 2017), fuzzy model (Yu et al., 2019), KNN (Safi & Safi, 2021), bagged trees (Oltu et al., 2021), and artificial neural network (ANN) (Ieracitano et al., 2020; Rodrigues et al., 2021; Triggiani et al., 2017) can help test the validity of the input topographic or dynamic patterns or features, such as multiscale entropy vector (Song et al., 2018), functional connectivity (Nobukawa et al., 2020; Song et al., 2018; Yu et al., 2019), cepstral distances (Rodrigues et al., 2021), Hjorth parameters (Safi & Safi, 2021), coherence (Oltu et al., 2021), bispectrum (Ieracitano et al., 2020), power spectral density (Oltu et al., 2021; Tavares et al., 2019), frequency bands (Trambaiolli et al., 2017; Triggiani et al., 2017), and wavelet transform (Ieracitano et al., 2020; Oltu et al., 2021).

Among these machine learning methods, it has been proven that deep learning has exceptional performance in terms of the accuracy of classification. Convolutional neural networks (CNN), a leading deep learning structure for data on Euclidean space, outperform the above machine learning algorithms in classification accuracy (Craik et al., 2019). Based on the fast Fourier transform for extracting spectral features of EEG for AD diagnosis, Bi and Wang (2019) developed a discriminative convolutional Boltzmann machine; Ieracitano et al. (2019) and Deepthi et al. (2020) proposed a CNN model, respectively. Based on combining latent factors output by the encoder part of variational auto-encoder of EEG, Li, Wang, et al. (2021) extracted characteristics of AD. Based on the time-frequency analysis using CWT, Huggins et al. (2021) proposed an AlexNet model for the classification of AD, mild cognitive impairment subjects, and HC. With a connections matrix from EEG, Alves et al. (2021) presented a CNN for classifying AD and schizophrenia. The above CNN applications on AD classification (Bi & Wang, 2019; Deepthi et al., 2020; Huggins et al., 2021; Ieracitano et al., 2019; Li, Wang, et al., 2021) focus more on learning the locally and continuously changed multiscaled features on the Euclidean space from the EEG signals, neglecting the functional connectivity features. Although Alves et al. (2021) used the connection of EEG channels as the CNN input, it neglected the temporal EEG channels features and the input connections topology feature cannot be modelled effectively due to the arranged order of EEG channels. The key convolutional filters on CNN structures in the above applications cannot fully mine the multiscale topological interactive information of EEG channels.

Considering the complexity of EEG signals in the spatial and temporal domain, how to extract more abstract geometric features for better generalisation using the deep learning methods remains tremendously troubling. The structure–function connectivity network of EEG is non-Euclidean data because the channels are discrete and discontinuous in the spatial domain. Each EEG channel can be considered as a node and there is a cross-channel interaction between nodes. Instead, geometric graph-based deep learning methods would provide a more suitable way to learn the cross-channel topologically associated features of EEG. Building neural networks under the graph theory, graph convolutional neural networks (GCNs) have been developed specifically to handle highly multirelational graph data by jointly leveraging node-specific sequential features and cross-nodes topologically associative features in the graph domain (Gallicchio & Micheli, 2010; Gori et al., 2005; Scarselli et al., 2008; Sperduti & Starita, 1997). In recent 2 years, GCNs have been applied in the diagnoses of various brain disorders, such as children's ASD evaluation (Zhang et al., 2021), detection of epileptic (Zeng et al., 2020; Zhao et al., 2021), seizure prediction (Li, Liu, et al., 2021), and epilepsy classification (Chen et al., 2020). As far as we are concerned, there are no AD diagnostic approaches based on GCN-related models.

To enable this application, an adjacency matrix, representing the topological association between different EEG channels, must be constructed as the key input of GCN. EEG functional networks are widely used in cognitive neuroscience, for example, decision making (Si et al., 2019), emotion recognition (Li, Liu, et al., 2019), and Schizophrenia research (Li, Wang, et al., 2019). Traditional statistical functional correlation measures include Pearson correlation (PC) (Chen et al., 2020; Zhao et al., 2021), Tanh nonlinearity (Li, Liu, et al., 2021),

average correlation coefficients (Zeng et al., 2020), and covariance (Zhang et al., 2021), which cannot fully represent the complex brain connectivity. Many advanced methods to measure functional connectivity (FC) have been developed, such as phase locking values (PLV) and phase lag index (PLI) in the time domain (Franciotti et al., 2019; Mormann et al., 2000; Van Mierlo et al., 2014), magnitude squared coherence (MSC) and imaginary part of coherence (IPC) in the frequency domain (Al-Ezzi et al., 2021; Babiloni et al., 2005; Van Diessen et al., 2015; Wendling et al., 2009), and wavelet coherence (WC) in the time–frequency domain (Franciotti et al., 2019). These values can measure the degree of synchronisation between different brain regions and alterations in complex behaviours produced by the interaction among widespread brain regions (Babiloni et al., 2005; Sakkalis, 2011; Tafreshi et al., 2019; Van Mierlo et al., 2014), which have been proved important for AD classification using the statistical (Jalili, 2017; Zhao et al., 2019), and machine learning methods (Nobukawa et al., 2020; Song et al., 2018; Yu et al., 2019). However, the research on the combination of FC and GCN is limited, especially for AD-related research. Using these efficient FCs to construct the input adjacency matrix of GCN may promisingly provide more insightful information for the brain function interaction and lead to a higher classification accuracy of brain-related disorders.

In this article, a novel spatial–temporal GCN (ST-GCN) is proposed to classify AD from HC, benefiting from the adjacency matrix constructed by a variety of FC measures and the raw EEG recordings. We tested six adjacency matrices based on PC, MSC, IPC, WC, PLV, and PLI using EEG recordings from patients with AD and HCs. ST-GCN can jointly leverage the cross-channel topological connectivity features and channel-specific temporal features. To the best of the authors' knowledge, this is the first attempt for GCN to distinguish between AD and HC based on EEG recordings.

2 | METHODS

2.1 | Spatial–temporal graph convolutional network

In 1997, Sperduti and Starita first adopted neural networks to direct acyclic graphs (Sperduti & Starita, 1997), which motivated the early studies on GCNs (Gallicchio & Micheli, 2010; Gori et al., 2005; Scarselli et al., 2008). Currently, there are two basic approaches to generalising convolutions to structure graph data forms: spatial-based and spectral-based GNNs. Spatial-based GNNs define graph convolutions by rearranging vertices into certain grid forms which can be processed by normal convolutional operations (Niepert et al., 2016; Yu et al., 2017). Bruna et al. (2013) presented the first prominent spectral-based GCNs by applying convolutions in spectral domains with graph Fourier transforms. Since then, there have been increasing improvements, approximations, and extensions on spectral-based GNNs (Defferrard et al., 2016; Henaff et al., 2015; Kipf & Welling, 2016; Levie et al., 2018) to reduce the computational complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$ (Defferrard et al., 2016; Kipf & Welling, 2016). Visually, a graph convolution can handle the complexity of graph data by generalising a 2D convolution, motivated by the successful applications of CNNs in Euclidean space (Wu et al., 2020). Being considered as special graph data, each pixel of an image can be taken as a node whose neighbours are determined by a filter and a 2D convolution takes the weighted average of adjacent pixel values of each node. Similarly, graph convolutions can be performed by taking the weighted average of a node's neighbourhood information, which is unordered and variable in size, and different from images.

As shown in Figure 1a, the proposed architecture of ST-GCN is composed of two spatial–temporal convolutional blocks (ST-Conv Blocks), each of which is formed with one spatial graph convolution layer (Spatial Graph-Conv) and two sequential convolution layers (Temporal 1D-Conv). ST-Conv block can be stacked based on the complexity of specific cases. Layer normalisation is utilised within every ST-Conv Block to prevent overfitting. The EEG channels X with the adjacency matrix W are uniformly processed by ST-Conv Blocks to explore spatial and temporal dependence coherently. A flattened layer integrates comprehensive features to generate the final AD/HC classification.

For comparison, we designed a structure of a classical temporal convolutional neural network (T-CNN) shown in Figure 1b, inspired by some related works (Deepthi et al., 2020; Huggins et al., 2021; Li, Wang, et al., 2021). The difference between these two model structures is that T-CNN only has EEG channels X at the input without the adjacency matrix W, and no spatial graph convolution unit is used in the feature layer in the middle of the T-CNN. Other aspects of T-CNN are the same as ST-GCN. For brevity, we illustrate the structural details of each part of ST-GCN in the following.

2.1.1 | Spatial Graph-Conv

Based on the concept of a spectral graph convolution, we introduce the notion of a graph convolution operator "$*\mathcal{G}$," multiplying a signal $x \in \mathrm{R}^n$ in the spatial space with a kernel Θ,

$$\Theta *\mathcal{G} x = \Theta(L)x = \Theta\left(U\Lambda U^T\right)x = U\Theta(\Lambda)U^T x, \tag{1}$$

where the graph Fourier basis $U \in R^{n\times n}$ is a matrix of eigenvectors of the normalised graph Laplacian $L = I_n - D^{-\frac{1}{2}}WD^{-\frac{1}{2}} = U\Lambda U^T \in R^{n\times n}$ (I_n is an identity matrix, $D \in R^{n\times n}$ is the diagonal degree matrix with $D_{ii} = \sum_j W_{ij}$, $\Lambda \in R^{n\times n}$ is the diagonal matrix of eigenvalues of L, and filter $\Theta(\Lambda)$ is a diagonal matrix). A graph signal x is filtered by a kernel Θ with multiplication between Θ and the graph Fourier transform $U^T x$ (Shuman et al., 2013).

We utilise Chebyshev polynomials and first-order approximations (Kipf & Welling, 2016) here to reduce expensive computations of kernel Θ in Equation (1) due to its $\mathcal{O}(n^2)$ complex multiplications. Kernel Θ can be restricted to a polynomial of Λ as $\Theta(\Lambda) = \sum_{k=0}^{K-1}\theta_k\Lambda^k$, where $\theta \in R^K$ is a vector of polynomial coefficients and K is the kernel size determining the maximum radius of the convolution from central nodes. Chebyshev polynomial $T_k(x)$ is traditionally used to approximate kernels as a truncated expansion of order $K-1$ as

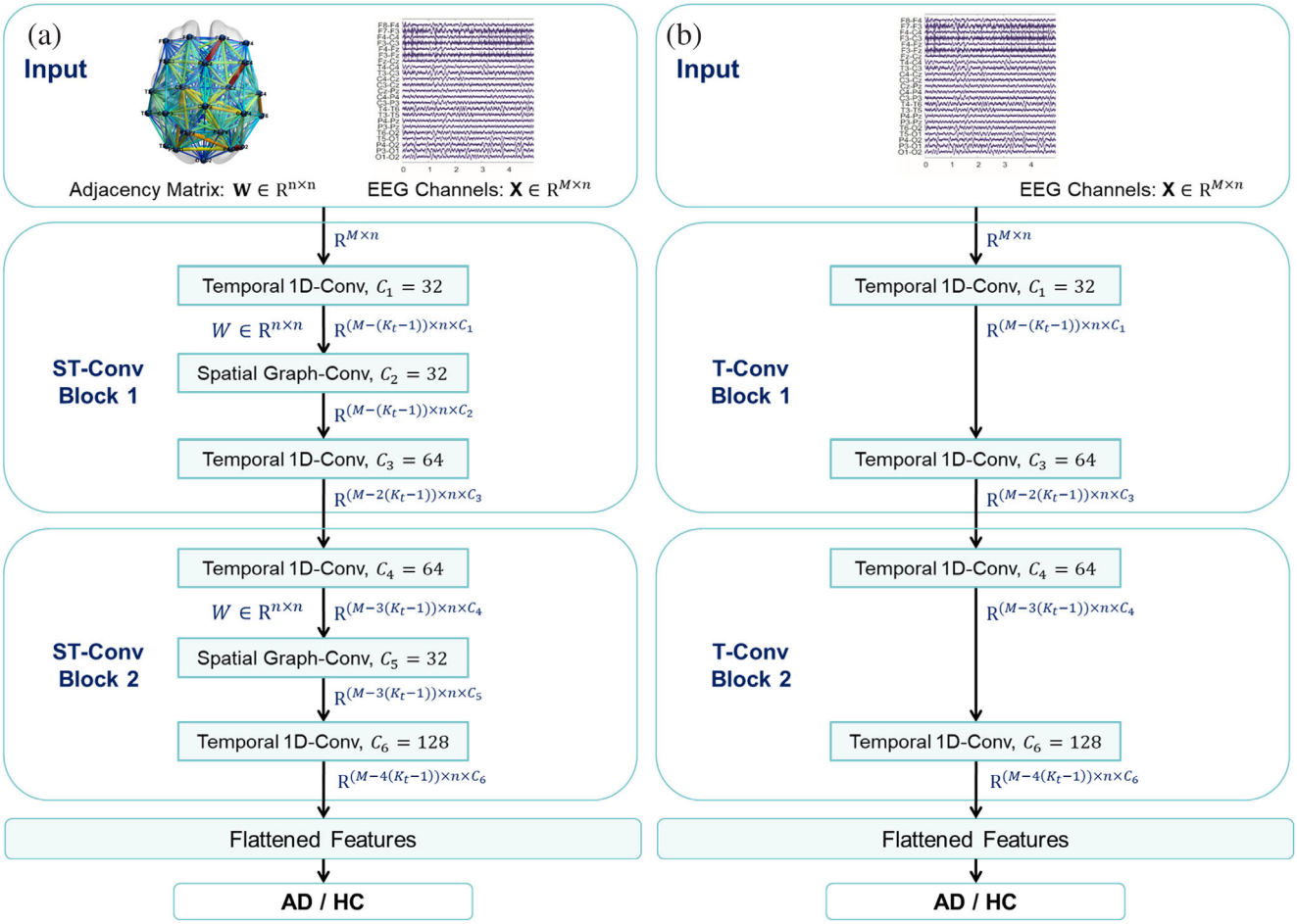

FIGURE 1 The flowchart of (a) the proposed ST-GCN framework and (b) the T-CNN framework for comparison. X is sized M^*N ($M = 25$, representing the length of the mini-epoch channel; $N = 23$, representing the 23 channels) and the input W is sized N^*N. The (i,j)th entry of the adjacency matrix W denotes spatial coupling correlation strength between the ith and jth of all the 23 different channels, the detailed calculation of which is presented in section 2.2. K_t is the size of the temporal 1D-Conv filter, set as 3 here. C_l $(l = 1,2,\ldots,6)$ is the number of filters in each layer. The estimated computational complexity of ST-GCN and T-CNN are 54M and 52M, respectively

$\Theta(\Lambda) \approx \sum_{k=0}^{K-1} \theta_k T_k\left(\tilde{\Lambda}\right)$ by rescaled $\tilde{\Lambda} = 2\Lambda/\lambda_{\max} - I_n$, where $\lambda_{\max}$ denotes the largest eigenvalue of L (Hammond et al., 2011). Then the graph convolution in Equation (1) can be rewritten as

$$\Theta * \mathcal{G}\, x = \Theta(L)\, x \approx \sum_{k=0}^{K-1} \theta_k T_k\left(\tilde{L}\right) x, \tag{2}$$

where $T_k\left(\tilde{L}\right) \in R^{n\times n}$ is the Chebyshev polynomial of order k evaluated at the scaled Laplacian $\tilde{L} = 2L/\lambda_{\max} - I_n$. By recursively computing K-localised convolutions through a polynomial approximation, the computational cost of Equation (1) can be reduced to $\mathcal{O}(K|\varepsilon|)$ as Equation (2). By stacking multiple localised graph convolutional layers with a first-order approximation of graph Laplacian, a layer-wise linear formulation can be defined (Kipf & Welling, 2016). Further assumption $\lambda_{\max} \approx 2$ can be made, due to the scaling and normalisation in neural networks. Thus, Equation (2) can be simplified to

$$\Theta * \mathcal{G}\, x \approx \theta_0 x + \theta_1\left(\frac{2}{\lambda_{\max}} L - I_n\right) x \approx \theta_0 x - \theta_1\left(D^{-\frac{1}{2}} W D^{-\frac{1}{2}}\right) x\ , \tag{3}$$

where θ_0 and θ_1 are two shared parameters of the kernel. θ_0 and θ_1 are replaced by a single parameter θ by letting $\theta_0 = -\theta_1 = \theta$ to constrain parameters and stabilise numerical performances. By renormalising W and D with $\tilde{W} = W + I_n$ and $\tilde{D}_{ii} = \sum_j \tilde{W}_{ij}$ separately, graph convolution can be expressed as

$$\Theta * \mathcal{G}\, x = \theta\left(I_n + D^{-\frac{1}{2}} W D^{-\frac{1}{2}}\right) x = \theta\left(\tilde{D} - \frac{1}{2}\tilde{W}\tilde{D} - \frac{1}{2}\right) x. \tag{4}$$

The graph convolution operator "$*\mathcal{G}$" defined on $x \in R^n$ can be extended to multidimensional tensors. For a signal with C_i channels $X \in R^{n\times C_i}$, the graph convolution can be generalised to

$$y_j = \sum_{i=1}^{C_i} \Theta_{i,j}(L)\, x_i \in R^n, 1 \le j \le C_0, \tag{5}$$

with the $C_i \times C_0$ vectors of Chebyshev coefficients $\Theta_{i,j} \in R^K$ (y_j is the output after graph convolution, C_i and C_0 are the sizes of input and output of the feature maps, respectively). A graph convolution for 2D

variables is denoted as "$\Theta * \mathcal{G} X$" with $\Theta \in R^{K \times C_i \times C_0}$. The input of ST-GCN is composed of M frames of EEG channels graph as shown in Figure 1a. Each frame X_t can be regarded as a matrix whose column i is a C_i-dimensional value of X_t at the ith node in graph $\mathcal{G}_t$, as $X \in R^{n \times C_i}$ (in this case, $C_i = 1$). For each time step t of M, the equal graph convolution operation with the same kernel Θ is imposed on $X_t \in R^{n \times C_i}$ in parallel. Thus, the graph convolution can be generalised to 3D variables, noted as "$\Theta * \mathcal{G} \chi$" with $\chi \in R^{M \times n \times C_i}$.

2.1.2 | Temporal 1D-Conv

Inspired by Gehring et al. (2017) that CNNs have the superiority of fast training in sequential-series analysis, we employ an entire convolutional structure on a temporal axis to capture sequential dynamic behaviours of EEG recordings. As shown in Figure 1a, a sequential convolutional layer contains a 1D convolution with a width K_t kernel followed by ReLu (a rectified linear unit function) as a nonlinearity. For each node in graph $\mathcal{G}$, its corresponding sequential convolution explores K_t neighbours of input elements, leading to shortening the length of sequences by $K_t - 1$ each time. Thus, an input of a sequential convolution for each node can be regarded as a length-M sequence with C_i channels as $Y \in R^{M \times C_i}$. The convolution kernel $\Gamma \in R^{K_t \times C_i \times C_0}$ is designed to map the input Y to a single output $P \in R^{(M - K_t + 1) \times (C_0)}$. Similarly, the temporal convolution can be generalised to 3D variables by employing the same convolution kernel Γ to every channel node in $\mathcal{G}$ equally, noted as "$\Gamma * \mathcal{T} \mathcal{Y}$" with $\mathcal{Y} \in R^{M \times n \times C_i}$.

The input and output of ST-Conv Blocks are all 3D tensors. For input $x^l \in R^{M \times n \times C_l}$ of block l, the output $x^{l+1} \in R^{(M - 2(K_t - 1)) \times n \times C^{l+1}}$ is computed by

$$x^{l+1} = \Gamma_1^l * \mathcal{T}\, \mathrm{ReLU}\left(\Theta^l * \mathcal{G}\left(\Gamma_0^l * \mathcal{T}\, x^l\right)\right), \tag{6}$$

where Γ_0^l and Γ_1^l are the upper and lower temporal kernels within block l, respectively; Θ^l is the spectral kernel of a graph convolution; $\mathrm{ReLU}(\cdot)$ denotes a rectified linear unit function. After stacking three ST-Conv Blocks, the output features are fused as a flattened layer (Figure 1a). We can obtain a final output $Z \in R^{n \times c}$ from the fully connected layer and calculate the classification result by applying a sigmoid transformation as $\widehat{Cl} = \mathrm{sig}(Z\omega + b)$, where $\omega \in R^c$ is a weight vector and b is a bias. We use the *binary cross-entropy* loss to measure the classification performance.

All models were trained on the CPU of DELL DESKTOP-D3UM3P9 with the Tensorflow platform in Microsoft Windows 10, and the optimiser used here is the Adam optimisation.

2.2 | Adjacency matrix

The spatial information carried by EEG signals plays an important role in AD/HC classification (Babiloni et al., 2005; Sakkalis, 2011; Tafreshi et al., 2019; Van Mierlo et al., 2014). The adjacency matrix $W \in R^{n \times n}$ of each mini-epoch, representing the spatial correlation along with the channel signals $X \in R^{n \times M}$ is one of the inputs of the ST-GCN model shown in Figure 1a. In some studies on brain disorders based on GCN (Chen et al., 2020; Li, Liu, et al., 2021; Zeng et al., 2020; Zhang et al., 2021; Zhao et al., 2021), the relationship of different channels of EEG is short of effective prior guidance and the adjacency matrix cannot ensure the utilisation of the coupling information between each channel. To address these issues, we first apply functional connectivity, which has been proven to be useful in AD classification, to construct the adaptive adjacency matrix to extract spatial coupling features.

The raw EEG signals of each channel and the association of channels are modelled by a graph. The nodes of the graph denote the feature vector of EEG signals, which are the raw mini-epoch EEG data. The adjacency matrix is constructed by Pearson correlation analysis and five functional connectivities of the mini-epoch EEG channels without any preset threshold to avoid the potential risk caused by manual selection, such as in Chen et al. (2020), Zhang et al. (2021), and Zhao et al. (2021). Specifically, the (i,j)th entry of the adjacency matrix W denotes spatial coupling strength between the ith and jth channels. Thus, W indicates that all channels are interconnected with different weights. Pair-wise correlation analysis among these 23 different channels is conducted using five functional connectivities and Pearson correlation for comparison, the calculation of which is presented below.

2.2.1 | Pearson correlation

The most well-known functional connectivity measure is the correlation, also called the Pearson correlation coefficient. It calculates the instantaneous linear interdependency between two signals based on the amplitudes of the signals in the time domain and it ranges from -1 to 1. The Pearson correlation coefficient between signal X and Y can be defined as follows:

$$\rho_{xy} = \frac{E\left[(x - \mu_x)(y - \mu_y)\right]}{\sigma_x \sigma_y}, \tag{7}$$

where E is the expected value, μ_x and μ_y are the mean values and σ_x and σ_y are the standard deviations of X and Y time series.

2.2.2 | Magnitude squared coherence

MSC is a linear method to estimate the interconnections between the PSD (power spectral density) of two signals in the frequency domain. The MSC of signals X and Y can be written as

$$\mathrm{MSC}_{xy}(f) = C_{xy}^2 = \frac{S_{xy}(f)^2}{|S_{xx}(f)| \times |S_{yy}(f)|}, \tag{8}$$

where $S_{xx}(f)$ and $S_{yy}(f)$ are the PSD of signal X and Y, respectively, and $S_{xy}(f)$ is the cross PSD at frequency f.

2.2.3 | Imaginary part of coherence

To avoid the volume conduction effects, instead of looking at the magnitude squared coherency, the imaginary part of the coherency is calculated by

$$IC_{xy} = RC_{xy} + IC_{xy}. \tag{9}$$

2.2.4 | Wavelet coherence

WC is generally acknowledged as a qualitative estimator that can represent the dynamic relations in the time–frequency domain between signals (Tafreshi et al., 2019). The wavelet transforms is defined as the convolution of the input x with a Wavelet family $\theta(u)$,

$$W_x(t,f) = \int_{-\infty}^{+\infty} x(u) \cdot \theta^*_{t,f}(u) du. \tag{10}$$

Given input signals x and y, wavelet cross-spectrum around time t, and frequency f can be derived by the Wavelet transforms of x and y,

$$CW_{xy}(t,f) = \int_{t-\delta/2}^{t+\delta/2} W_x(\tau,f) \cdot W^*_y(\tau,f), \tag{11}$$

where * defines the complex conjugate and δ is assumed as a frequency-depending time scalar. WC at the time t and frequency f is derived as

$$WC_{xy}(t,f) = \frac{|CW_{xy}(t,f)|}{|CW_{xx}(t,f) \times CW_{yy}(t,f)|^{1/2}}. \tag{12}$$

2.2.5 | Phase locking value

Phase synchronization assumes that two oscillation signals without amplitude synchronization can have phase synchronization. The PLV is high-frequently utilised to obtain the strength of phase synchronisation (Van Mierlo et al., 2014). The instantaneous phase of a signal x is given by

$$\emptyset_x(t) = \arctan \frac{\tilde{x}(t)}{x(t)}, \tag{13}$$

where $\tilde{x}(t)$ is the Hilbert transform of $x(t)$, defined as

$$\tilde{x}(t) = \frac{1}{\pi} PV \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau} d\tau, \tag{14}$$

where PV refers to the Cauchy principal value. The PLV for two signals is then defined as

$$PLV = \left| \frac{1}{N} \sum_{j=0}^{N-1} e^{j(\emptyset_x(j\Delta t) - \emptyset_y(j\Delta t))} \right|, \tag{15}$$

where Δt defines the sampling period and N indicates the sample number of each signal. The range of PLV is between 0 and 1, where 0 shows a lack of synchronization and 1 indicates strict phase synchronization.

2.2.6 | Phase lag index

Similarly to the calculation of PLV, PLI captures the asymmetry of the distribution of phase differences between two signals and is calculated based on the relative phase difference between the two signals

$$PLI = |E[\text{sign}(\emptyset_x(j\Delta t) - \emptyset_y(j\Delta t))]|, \tag{16}$$

where $\emptyset_x(j\Delta t) - \emptyset_y(j\Delta t)$ is the phase difference between two signals, sign stands for signum function, E is the expected value, and || indicates the absolute value. PLI values range between 0 and 1, where 0 can indicate possibly no coupling and 1 refers to perfect phase locking.

2.3 | Data set and preprocessing

All patients and healthy controls included in this work were recruited in the Sheffield Teaching Hospital memory clinic and all provided written informed consent. The EEG study underwent ethics approval by the Yorkshire and The Humber (Leeds West) Research Ethics Committee (reference number 14/YH/1070). AD patients had their diagnosis confirmed between 1 month and up to 2 years before recording their EEG while they had mild to moderate cognitive deficits, according to their Mini-mental state examination. All AD subjects had brain MRI scans to eliminate other alternative causes of dementia. For the aged and gender-matched HC cohort, normal MRI brain scans and cognitive assessments were required before their EEG recordings. Further details about recruitment, diagnostic criteria, and study design can be found in the previously published work (Blackburn et al., 2018).

All participants were younger than 70 years old, including 19 AD patients and 20 HC participants. EEG recordings were undertaken with an XLTEK 128-channel headbox (Optima Medical Ltd.) and Ag/AgCL electrodes at a sampling frequency of 2 kHz by implementing a modified 10–10 overlapping a 10–20 international system of electrode placement, with a referential montage (linked earlobe reference). Thirty-minute resting state (task-free – participants were instructed to rest and refrain from thinking anything specific) EEG recordings were obtained from each participant including sustained periods of keeping their eyes closed (EC) alternating with periods during which they kept their eyes open (EO). The recordings obtained were subsequently reviewed by a neurophysiologist—on an XLTEK

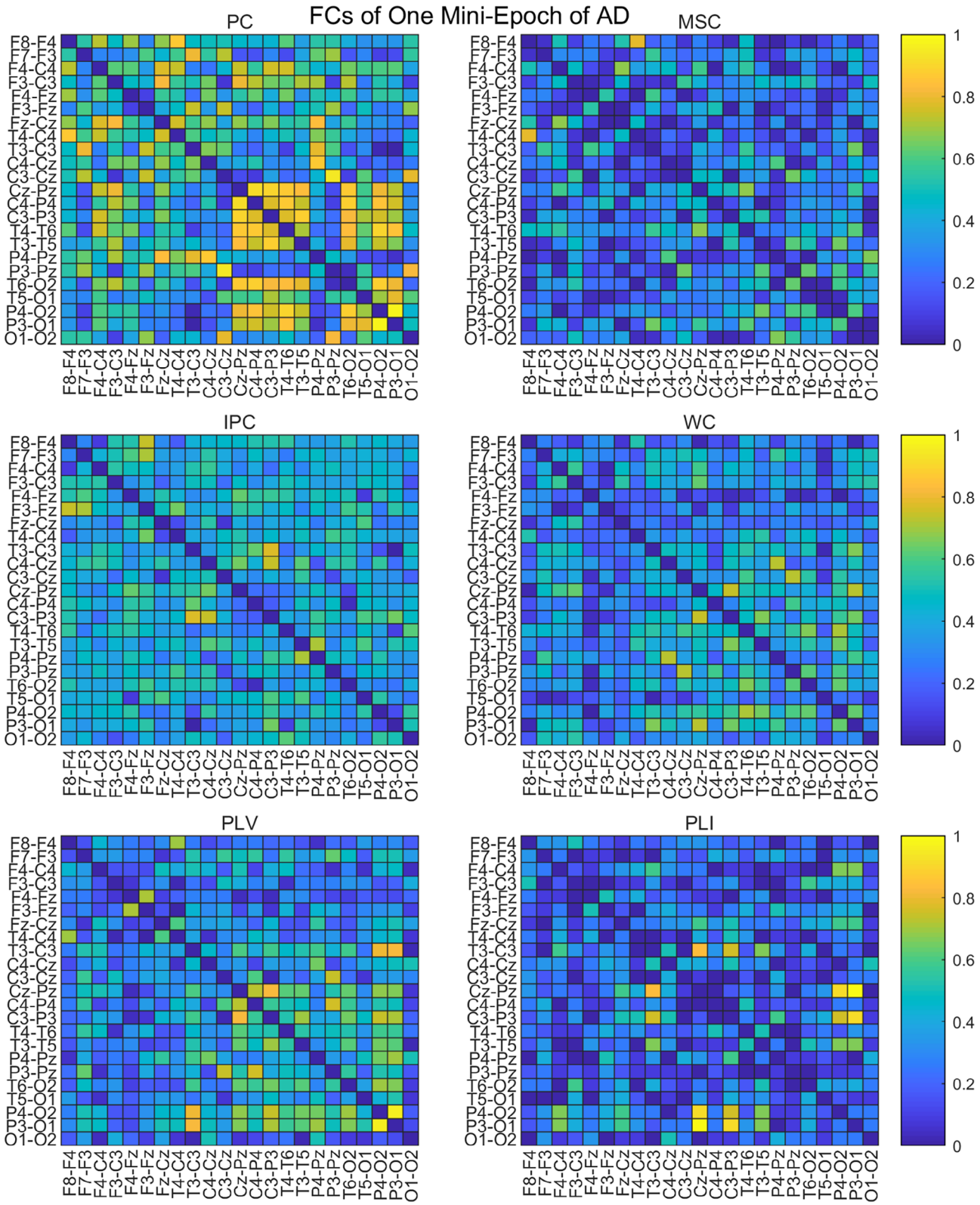

FIGURE 2 The adjacency matrices of one mini-epoch of AD from six FC methods

review station. For each participant, three 12-s artefact-free epochs of EC and EO were selected. To avoid volume conduction effects related to the common reference electrode, 23 bipolar derivations were created: F8–F4, F7–F3, F4–C4, F3–C3, F4–FZ, FZ–CZ, F3–FZ, T4–C4, T3–C3, C4–CZ, C3–CZ, CZ–PZ, C4–P4, C3–P3, T4–T6, T3–T5, P4–PZ, P3–PZ, T6–O2, T5–O1, P4–O2, P3–O1, and O1–O2.

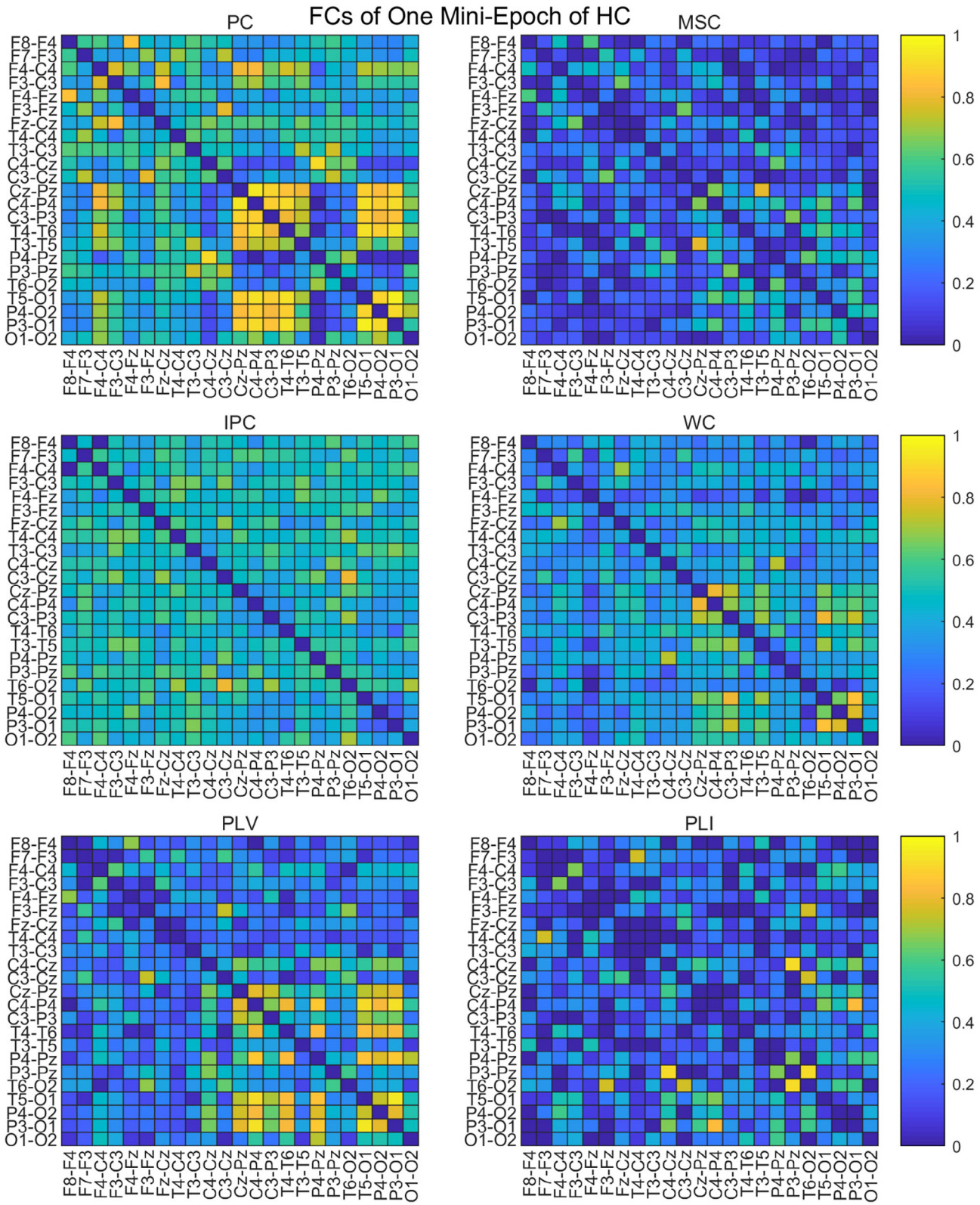

FIGURE 3 The adjacency matrices of one mini-epoch of HC from six FC methods

For both EO and EC of 19 ADs and 20 HCs, there are three artefact-free epochs, each of which lasts 12 s. We first applied the Butterworth filter for every epoch to subsample the EEG signals at 100 Hz and obtained the corresponding six bands (Delta, Theta, Alpha, Beta, Gamma, and Full band of 0–48 Hz). Then, data segmentation is utilised to obtain 5472 mini-epochs for ADs and 5760 mini-

Eye states	FC	Delta	Theta	Alpha	Beta	Gamma	Full
EC	PC	72.1	78.2	78.3	82.5	81.2	90.3
	MSC	72.2	78.1	78.9	81.9	80.8	90.2
	IPC	73.8	78.4	80.2	82.3	82.1	90.7
	WC	73.9	78.7	81.0	83.2	82.5	92.3
	PLV	73.0	78.6	79.8	82	82.2	91.1
	PLI	73.9	78.2	79.6	82.3	81.1	92.1
EO	PC	70.9	75.6	76.8	77.4	75.5	88.1
	MSC	71.3	75.7	75.4	77.8	75.5	88.3
	IPC	71.0	76.0	77.4	78.1	75.6	88.7
	WC	71.2	76.8	78.2	78.5	76	89.4
	PLV	71.0	75.8	76.0	78.1	75.9	87.2
	PLI	70.3	75.0	75.6	76.5	75.8	87.9

TABLE 1 Classification accuracy for HC and AD in EC and EO states within six bands by ST-GCN with different adjacency matrices

Abbreviations: IPC, imaginary part of coherence; MSC, magnitude squared coherence; PC, Pearson correlation; PLI, phase lag index; PLV, phase locking value; WC, wavelet coherence.

Eye states	FC	Delta	Theta	Alpha	Beta	Gamma	Full
EC	PC	69.0	76.7	77.3	80.3	75.9	88.1
	MSC	70.4	76.1	74.0	80.7	74.8	86.8
	IPC	69.4	77.8	77.1	78.2	77.5	88.8
	WC	71.0	78.0	76.7	80.3	80.5	89.0
	PLV	71.4	78.2	76.9	77.9	77.4	88.1
	PLI	68.5	76.6	78.7	80.6	75.5	87.4
EO	PC	67.9	75.3	77.2	71.3	74.3	85.6
	MSC	67.0	74.8	76.9	75.0	69.8	85.9
	IPC	65.0	76.1	76.5	72.9	74.1	86.4
	WC	70.1	75.9	77.9	76.2	73.7	86.3
	PLV	65.8	74.5	75.9	71.5	73.5	85.0
	PLI	67.3	74.9	76.9	71.6	71.9	85.7

TABLE 2 Classification accuracy for HC and AD in EC and EO states within six bands by T-CCN with different adjacency matrices

Abbreviations: IPC, imaginary part of coherence; MSC, magnitude squared coherence; PC, Pearson correlation; PLI, phase lag index; PLV, phase locking value; WC, wavelet coherence.

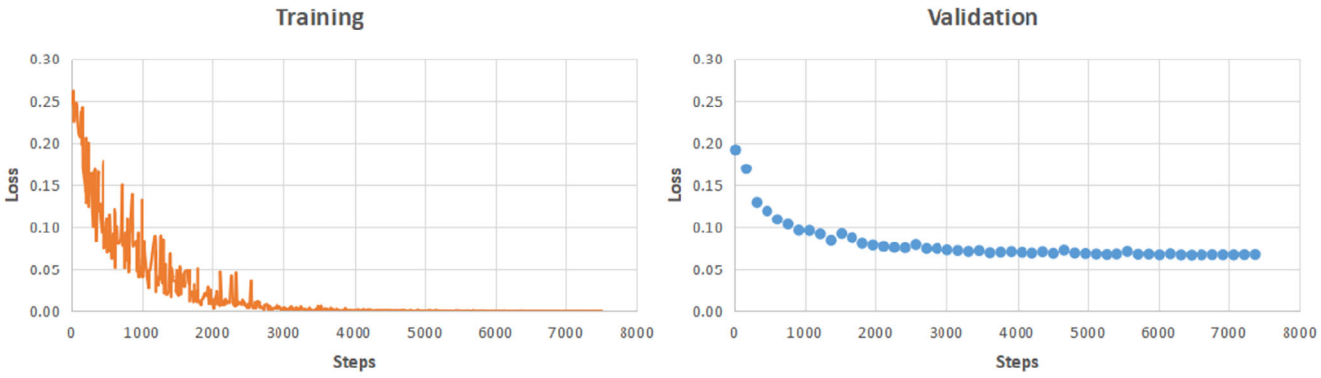

FIGURE 4 Training and convergence process for Full band data with WC connectivity by the ST-GCN model

epochs for HCs with a window size of 25 data points without overlapping. For each mini-epoch signal, six functional connectivities (PC, MSC, IPC, WC, PLV, and PLI) were calculated to obtain the corresponding adjacency matrices. Finally, for each measure of EC or EO, 3744 samples (1824 for ADs and 1920 for HCs) were selected randomly covering two-thirds of each 12 s epoch and split for

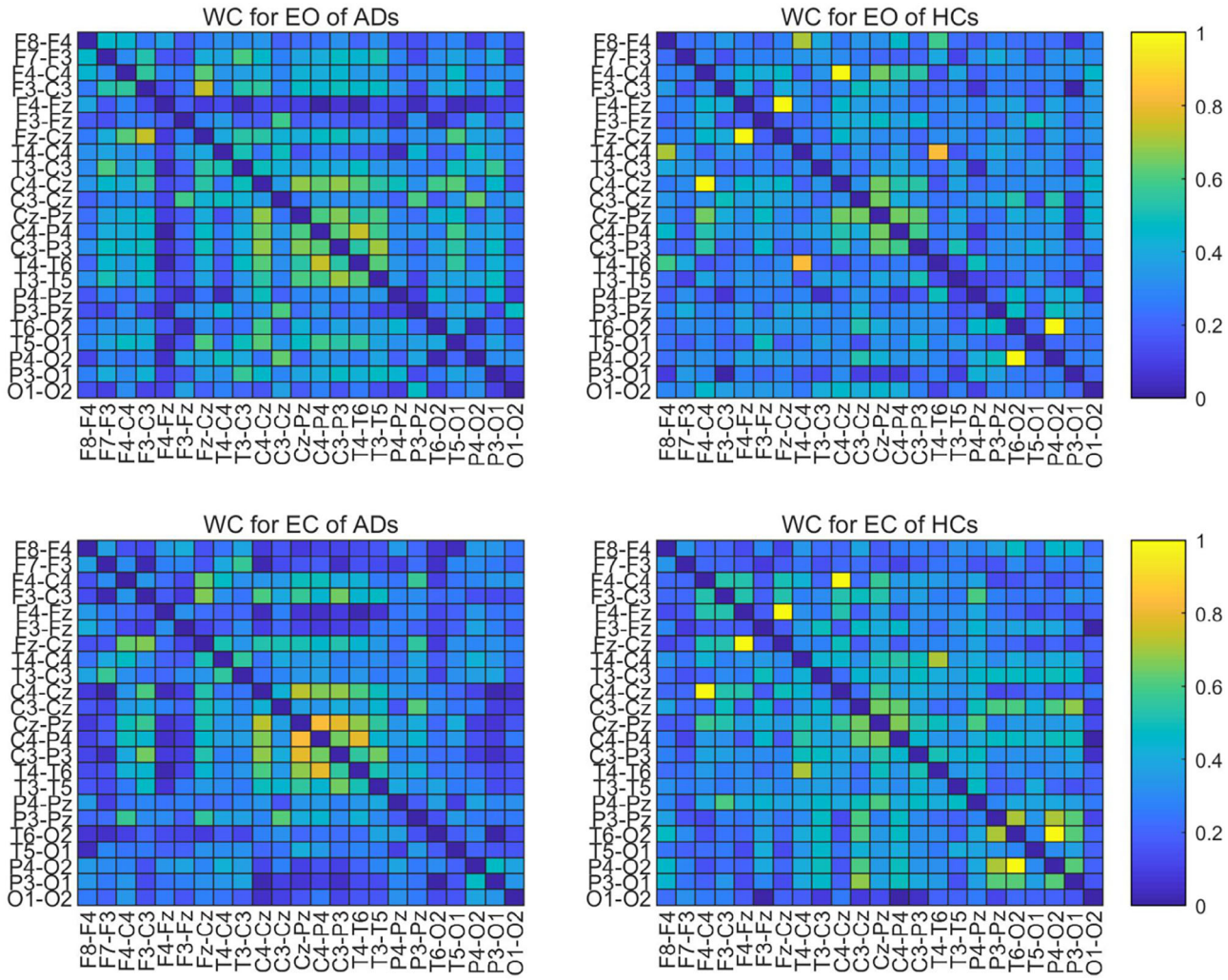

FIGURE 5 The averaged WC adjacency matrices of ADs and HCs

10-folder cross-validation to generate the training data set and validation data set. The remaining 1872 samples (912 for ADs and 960 for HCs) were considered as the testing data set. The purpose of the validation data set during training is to obtain the model loss on the validation data set after the optimization of each epoch. If the loss of the validation data set is decreasing, we save the optimised model until the end of all training epochs, which can prevent the model from overfitting.

3 | RESULTS

3.1 | Adjacency matrix from different functional connectivities

To compare the adjacency matrices calculated by six functional connectivities, we take the Full band of one mini-epoch of AD (Figure 2) and HC (Figure 3) as an example to see their difference. For the mini-epoch signal of both AD and HC, PC generates higher coupling between channels than the five functional connectivity methods. The adjacency matrix calculated by MSC is relatively lower than the others and WC and PLV show similar coupling distribution with PC but with overall lower strength. IPC produces the most uniform connection distribution. PLI produces the sparsest connection distribution, and there are some strong connection points close to 1. Therefore, adjacency matrices from the six calculation methods have differences in the coupling analysis between EEG channels, which may have an impact on the classification performance of the ST-GCN method.

3.2 | Overall classification performance

We conducted a direct comparison of classifying AD and HC participants using the extracted six adjacency matrices by training the ST-GCN and T-CNN methods on the EO and EC states and six frequency bands independently. Tables 1 and 2 show the corresponding testing accuracies from ST-GCN and T-CNN models, respectively. For both methods, the classification accuracy of all frequency bands and adjacency matrix calculation methods in the EC state is higher than the corresponding EO state, which is consistent with previous research results (Barry et al., 2007; Wan et al., 2019). For different frequency bands, when the 1–48 Hz full-

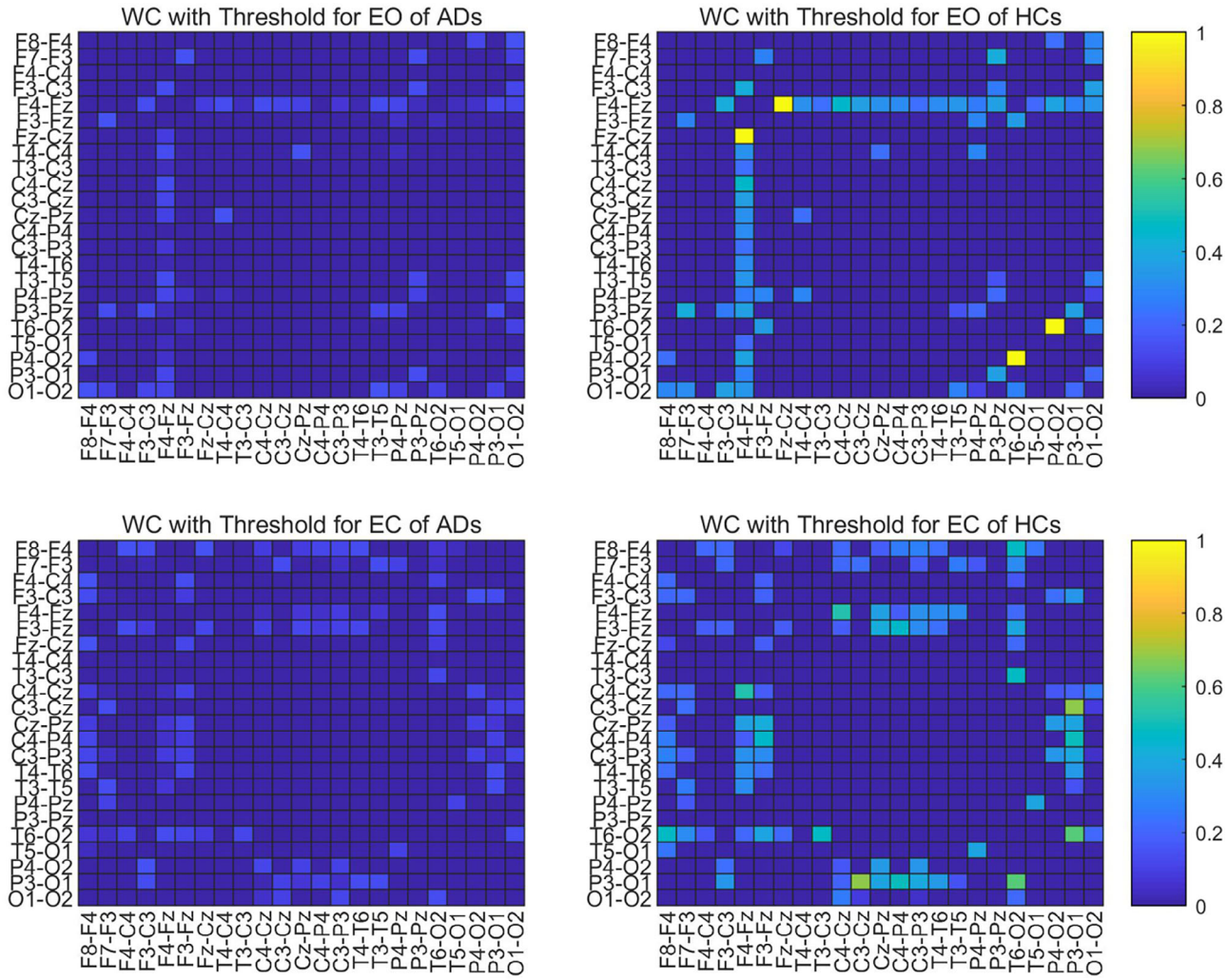

FIGURE 6 Threshold of averaged WC adjacency matrices of ADs and HCs

band signal and the corresponding adjacency matrix are used as the input, the mean classification accuracies are about 91.1% by ST-GCN and 88.0% by T-CNN, respectively, which are higher than all sub-bands data. For ST-GCN, the mean classification accuracy of Beta band data is the highest (about 82.3% for EC and 77.7% for EO), and the mean classification accuracy of Delta band data is the lowest (about 73.1% for EC and 71.0% for EO). For T-CCN, the mean classification accuracy of Alpha band data is the highest (about 76.8% for EC and EO), and the mean classification accuracy of Delta band data is the lowest (about 68.6% for EC and EO). Overall, ST-GCN outperforms T-CNN on both eye states and almost all sub-band data, which indicates that spatial topology constraints can indeed mine EEG features, resulting in improved classification accuracy.

It is worth noting that for different adjacency matrix calculation methods, the classification accuracy of the WC method is better than all other methods, which proves the effectiveness of time–frequency analysis for extracting coupling features. For ST-GCN, the overall accuracy of the five functional connection methods is higher than that of the PC method, indicating that the usage of functional connectivity as the adjacency matrix can improve the performance of the ST-GCN model to a certain extent.

The training and convergence process for Full band data with WC connectivity by the proposed ST-GCN model is shown in Figure 4.

3.3 | Wavelet coherence as adjacency matrix

To further explore the effectiveness of WC which has the best classification performance as the adjacency matrix, we conduct a statistical analysis of the WCs of all AD and HC. By averaging the WC adjacency matrices of the full band of all EO and EC mini-epochs for AD and HC, respectively, as shown in Figure 5, we can analyse their statistical characteristics. For ADs, the EC state has a slightly higher coupling strength than EO in the temporo-occipital area (the middle part of the adjacency matrix), while for HCs, the EC state has a bit higher overall coupling strength than EO. For both EO and EC, the strongest inter-channel correlation of AD is less than 0.8 (the middle part of the left two images of Figure 5 – temporo-occipital area), while HC has some

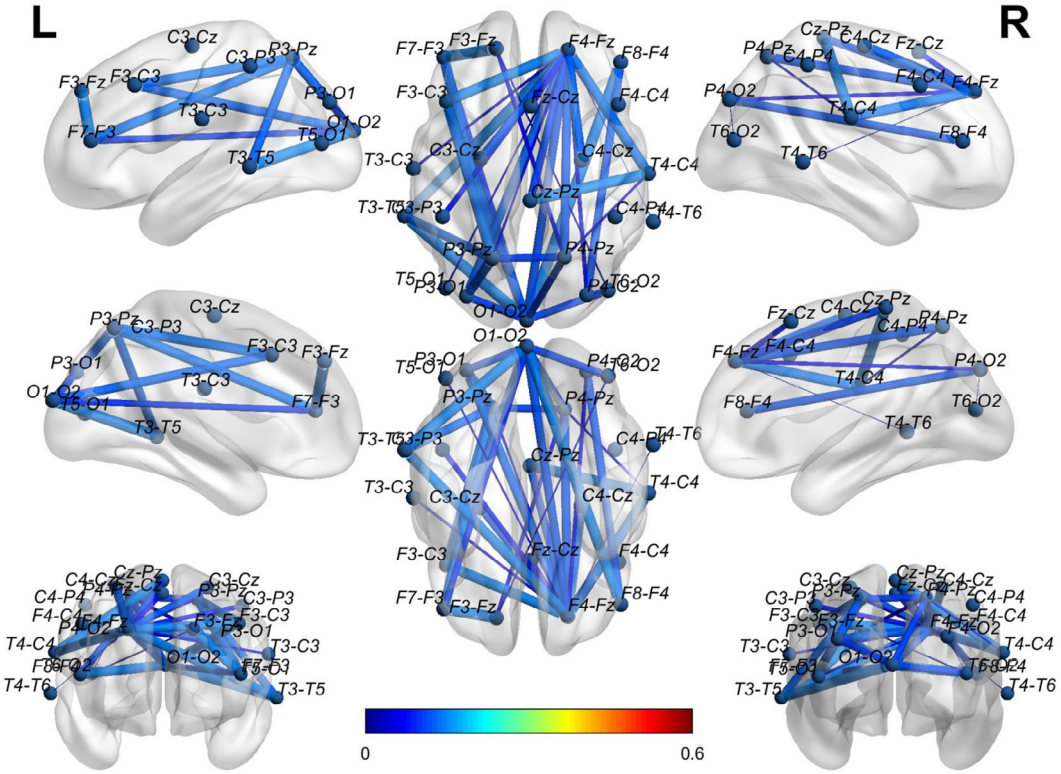

FIGURE 7 The 3D brain mapping of averaged WC adjacency matrices for EO state of ADs

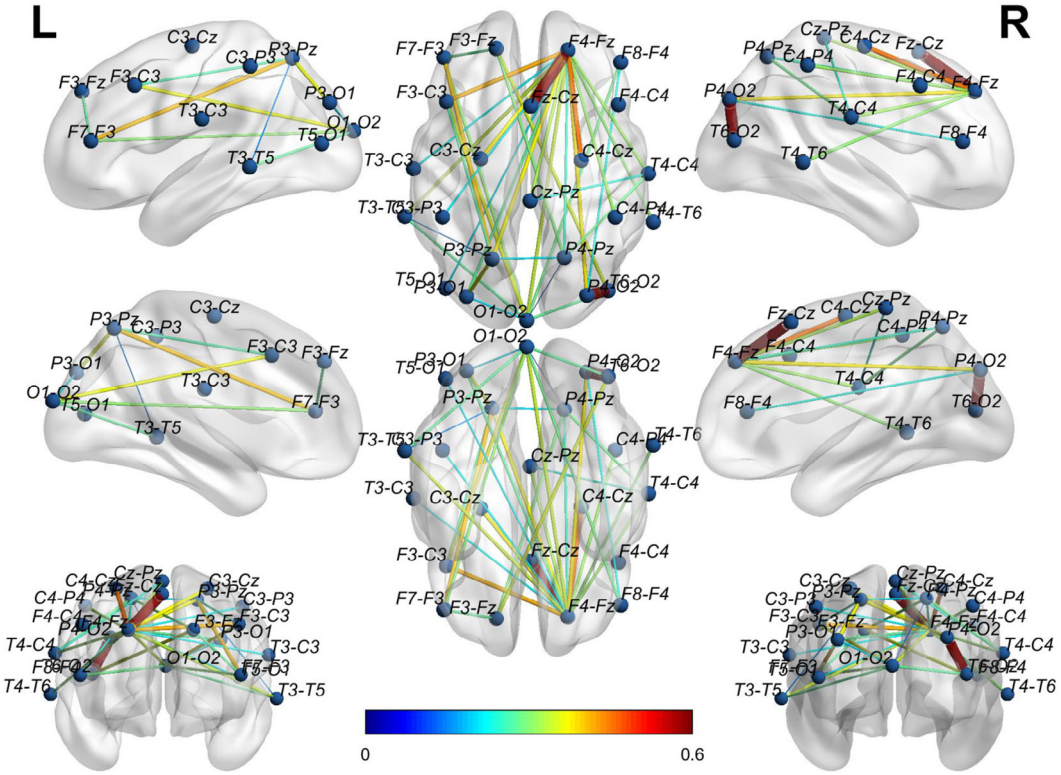

FIGURE 8 The 3D brain mapping of averaged WC adjacency matrices for EO state of HCs

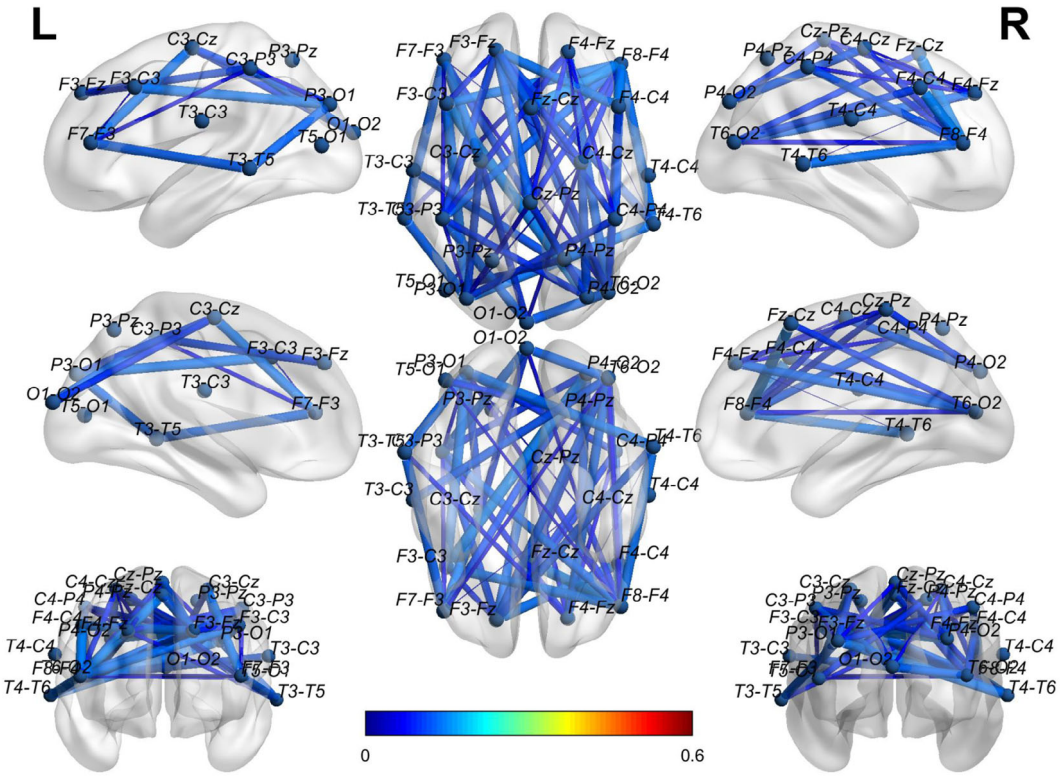

FIGURE 9 The 3D brain mapping of averaged WC adjacency matrices for EC state of ADs

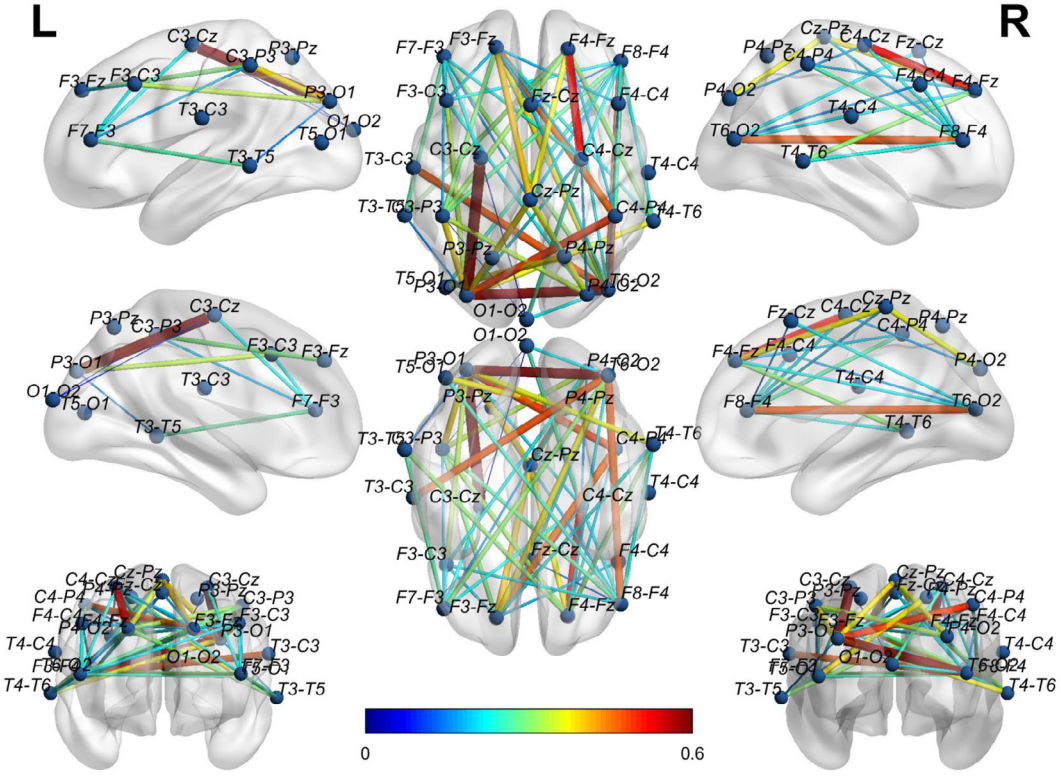

FIGURE 10 The 3D brain mapping of averaged WC adjacency matrices for EC state of HCs

interchannel correlations close to 1 (the upper-left frontocentral and bottom-right posterior area of the right two images of Figure 5). For both EO and EC, the connectivity between ADs channels is lower than that of HCs in the bottom-right corner (posterior area).

To visually show the connectivity between the channels in the bottom-left corner for both EO and EC states, we thresholded the left two AD adjacency matrices in Figure 5 to screen out the channels with connectivity below 0.15. Figure 6 reports the corresponding cross-channel indexes for HCs. Through the software BrainNet, the 3D connectivity distributions in the brain of ADs and HCs are visually displayed for EO (Figures 7 and 8) and EC (Figures 9 and 10), respectively. Comparing Figures 7 and 8 for EO, the connectivities of HCs between the channels at the front and back parts of the brain are much higher than those of ADs. P4-O2 to T6-O2 and F4-Fz to Fz-Cz have the highest coupling strengths (close to 0.6) for channel pairs of HCs, and the corresponding coupling strengths of ADs are close to 0. Comparing Figures 9 and 10 for EC, the connectivity of HCs between the channels at the back of the brain is much higher than that of ADs. C3-Cz to P3-O1 and P3-O1 to T6-O2 have the highest coupling strengths (close to 0.6) for channel pairs of HCs, and the corresponding coupling strengths of ADs are as low as 0.1.

4 | CONCLUSIONS

We proposed a Spatial–temporal Graph Convolutional Network (ST-GCN) for classifying Alzheimer's disease and Healthy Controls groups by jointly leveraging cross-channel topological association features and channel-specific temporal features of EEG recordings. Different from the currently leading GCN applications for diagnosing brain disorders, this method utilises brain functional connectivity methods for exploring the complex interactive information between EEG channels as well as the single-channel-based dynamic information. The main goal of this work was to determine whether the cross-channel topologically associated features constrained by the functional connectivity can reveal more hidden information in data and extend the applicability of the GCN-based algorithm. For the clinical AD and HC EEG recordings, ST-GCN has exhibited superior performance in achieving the highest classification accuracy with wavelet coherence as the adjacency matrix. For the tested data set, the overall classification accuracy of ST-GCN is higher than the classical T-CNN method on both eye states and different frequency bands, which suggests that spatial topology constraints can indeed mine brainwave features and thereby improve the classification accuracy. Different from existing studies for AD diagnosis that are based on either topological brain function characteristics of EEG (Fan et al., 2018; Jalili, 2017; Tait et al., 2020; Tylová et al., 2018; Zhao et al., 2019) or temporal dynamic features (Bi & Wang, 2019; Deepthi et al., 2020; Huggins et al., 2021; Ieracitano et al., 2019; Li, Wang, et al., 2021), the proposed ST-GCN considers both the adjacency matrix of functional connectivity from multiple EEG channels and corresponding dynamics of signal EEG channel simultaneously. It has the potential to pick up the anomaly of AD not only in the frequency response of local areas but also in the functional connectivity across different regions. Furthermore, the visualisation of wavelet coherence adjacency matrices increases the transparency of this solution by providing evidence of brain anomaly in terms of functional connectivity. This investigation is important as it will increase the trust in the developed AI-based solution. This algorithm lays a potentially effective strategy for the applications of other brain disorders.

In the present study, due to the limited number of subjects in the data set, the Leave-One-Subject-Out or cross-subject validation is not discussed to avoid biased conclusions caused by data insufficiency. The accuracy of a hand-out cross-subject validation by ST-GCN and T-CNN can be found in Tables S1 and S2, Supporting Information. In this case, 65% of the subjects were used for training and the remaining 35% were used for testing. Although the overall accuracy is dropped significantly for both methods, the proposed ST-GCN still outperforms T-CNN.

To reduce volume conduction effects from a common reference, bipolar derivations were used to assess the degree of differences between various pairs of electrodes for two different cohorts of subjects. With this approach—the use of bipolar pairs of electrodes—the effects of volume conduction are reduced but not eliminated. We recognise that this work is based on a sensor level scalp EEG analysis, and we do not claim to be able to precisely localise the spatial characteristics underpinning the EEG sensor findings.

ACKNOWLEDGMENT

This research is based on data provided by the National Institute for Health Research (NIHR) Sheffield Biomedical Research Centre (Translational Neuroscience)/NIHR Sheffield Clinical Research Facility.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Xiaocai Shan https://orcid.org/0000-0003-0817-0357
Jun Cao https://orcid.org/0000-0003-2121-7631
Shoudong Huo https://orcid.org/0000-0002-8145-317X
Liangyu Chen https://orcid.org/0000-0002-6817-8644
Ptolemaios Georgios Sarrigiannis https://orcid.org/0000-0002-8380-8755
Yifan Zhao https://orcid.org/0000-0003-2383-5724

REFERENCES

Al-Ezzi, A., Kamel, N., Faye, I., & Gunaseli, E. (2021). Analysis of default mode network in social anxiety disorder: Eeg resting-state effective connectivity study. *Sensors*, *21*, 4098. https://doi.org/10.3390/s21124098

Alves, C. L., Pineda, A. M., Roster, K., Thielemann, C., & Rodrigues, F. A. (2021). EEG functional connectivity and deep learning for automatic diagnosis of brain disorders: Alzheimer's disease and schizophrenia. arXiv preprint: arXiv2110.06140.

Babiloni, F., Cincotti, F., Babiloni, C., Carducci, F., Mattia, D., Astolfi, L., Basilisco, A., Rossini, P. M., Ding, L., Ni, Y., Cheng, J., Christine, K., Sweeney, J., & He, B. (2005). Estimation of the cortical functional

connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. *Neuroimage*, *24*, 118–131. https://doi.org/10.1016/j.neuroimage.2004.09.036

Barry, R. J., Clarke, A. R., Johnstone, S. J., Magee, C. A., & Rushby, J. A. (2007). EEG differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, *118*, 2765–2773.

Bi, X., & Wang, H. (2019). Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning. *Neural Networks*, *114*, 119–135.

Blackburn, D. J., Sarrigiannis, P. G., De Marco, M., Zhao, Y., Venneri, A., Lawrence, S., Unwin, Z. C., Blyth, M., Angel, J., Baster, K., Farrow, T. F. D., Wilkinson, I. D., Billings, S. A., Venneri, A., & Sarrigiannis, P. G. (2018). A pilot study investigating a novel non-linear measure of eyes open versus eyes closed EEG synchronization in people with Alzheimer's disease and healthy controls. *Brain Sciences*, *8*, 1–19. https://doi.org/10.3390/brainsci8070134

Brookmeyer, R., Gray, S., & Kawas, C. (1998). Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *American Journal of Public Health*, *88*, 1337–1342.

Brookmeyer, R., Johnson, E., Ziegler-Graham, K., & Arrighi, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimer's Dementia*, *3*, 186–191.

Bruna, J., Zaremba, W., Szlam, A., & LeCun, Y. (2013) Spectral networks and locally connected networks on graphs. arXiv preprint: arXiv1312.6203.

Chen, X., Zheng, Y., Niu, Y., & Li, C. (2020). Epilepsy classification for mining deeper relationships between EEG channels based on GCN. In *2020 international conference on computer vision, image and deep learning (CVIDL)* (pp. 701–706). IEEE.

Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: A review. *Journal of Neural Engineering*, *16*, 31001.

Davies, C. A., Mann, D. M. A., Sumpter, P. Q., & Yates, P. O. (1987). A quantitative morphometric analysis of the neuronal and synaptic content of the frontal and temporal cortex in patients with Alzheimer's disease. *Journal of the Neurological Sciences*, *78*, 151–164.

Deepthi, L. D., Shanthi, D., & Buvana, M. (2020). An intelligent Alzheimer's disease prediction using convolutional neural network (CNN). *International Journal of Advanced Research in Science, Engineering and Technology*, *11*, 12–22.

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in Neural Information Processing Systems*, *29*, 3844–3852.

DeKosky, S. T., & Scheff, S. W. (1990). Synapse loss in frontal cortex biopsies in Alzheimer's disease: Correlation with cognitive severity. *Annals of Neurology*, *27*, 457–464.

Fan, M., Yang, A. C., Fuh, J.-L., & Chou, C.-A. (2018). Topological pattern recognition of severe Alzheimer's disease via regularized supervised learning of EEG complexity. *Frontiers in Neuroscience*, *12*, 685.

Franciotti, R., Falasca, N. W., Arnaldi, D., Famà, F., Babiloni, C., Onofrj, M., Nobili, F. M., & Bonanni, L. (2019). Cortical network topology in prodromal and mild dementia due to Alzheimer's disease: Graph theory applied to resting state EEG. *Brain Topography*, *32*, 127–141. https://doi.org/10.1007/s10548-018-0674-3

Gallicchio, C., & Micheli, A. (2010). Graph echo state networks. In *The 2010 international joint conference on neural networks (IJCNN)* (pp. 1–8). IEEE.

Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017). Convolutional sequence to sequence learning. *Proceedings of the 34th International Conference on Machine Learning, PMLR*, 70, 1243–1252.

Gori, M., Monfardini, G., & Scarselli, F. (2005). A new model for learning in graph domains. In *Proceedings of 2005 IEEE international joint conference on neural networks* (pp. 729–734). IEEE.

Hammond, D. K., Vandergheynst, P., & Gribonval, R. (2011). Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, *30*, 129–150.

Henaff, M., Bruna, J., & LeCun, Y. (2015). Deep convolutional networks on graph-structured data. arXiv preprint: arXiv1506.05163.

Huggins, C. J., Escudero, J., Parra, M. A., Scally, B., Anghinah, R., de Araújo, A. V. L., Basile, L. F., & Abasolo, D. (2021). Deep learning of resting-state electroencephalogram signals for 3-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing. *Journal of Neural Engineering*, *18*(4).

Ieracitano, C., Mammone, N., Bramanti, A., Hussain, A., & Morabito, F. C. (2019). A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings. *Neurocomputing*, *323*, 96–107.

Ieracitano, C., Mammone, N., Hussain, A., & Morabito, F. C. (2020). A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Networks*, *123*, 176–190.

Jalili, M. (2017). Graph theoretical analysis of Alzheimer's disease: Discrimination of AD patients from healthy subjects. *Information Sciences*, *384*, 145–156.

Jelic, V. (2005). EEG in dementia, review of the past, view into the future. In S.N. Sarbadhikari (Ed.), *Depression and Dementia: Progress in Brain Research, Clinical Functioning and Future Trends* (pp. 245–304). Nova Publisher.

Kipf, T. N. and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint:. arXiv1609.02907.

Levie, R., Monti, F., Bresson, X., & Bronstein, M. M. (2018). Cayleynets: Graph convolutional neural networks with complex rational spectral filters. *IEEE Transactions on Signal Processing*, *67*, 97–109.

Li, F., Wang, J., Liao, Y., Yi, C., Jiang, Y., Si, Y., Peng, W., Yao, D., Zhang, Y., Dong, W., & Xu, P. (2019). Differentiation of schizophrenia by combining the spatial EEG brain network patterns of rest and task P300. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*, 594–602.

Li, K., Wang, J., Li, S., Yu, H., Zhu, L., Liu, J., & Wu, L. (2021). Feature extraction and identification of Alzheimer's disease based on latent factor of multi-channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 1557–1567.

Li, P., Liu, H., Si, Y., Li, C., Li, F., Zhu, X., Huang, X., Zeng, Y., Yao, D., Zhang, Y., & Xu, P. (2019). EEG based emotion recognition by combining functional connectivity network and local activations. *IEEE Transactions on Biomedical Engineering*, *66*, 2869–2881.

Li, Y., Liu, Y., Guo, Y.-Z., Liao, X.-F., Hu, B., & Yu, T. (2021). Spatio-temporal-spectral hierarchical graph convolutional network with semisupervised active learning for patient-specific seizure prediction. *IEEE Transactions on Cybernetics*. https://doi.org/10.1109/TCYB.2021.3071860

Masliah, E., Mallory, M., Alford, M., DeTeresa, R., Hansen, L. A., McKeel, D. W., & Morris, J. C. (2001). Altered expression of synaptic proteins occurs early during progression of Alzheimer's disease. *Neurology*, *56*, 127–129.

Michel, C. M., Koenig, T., Brandeis, D., Gianotti, L. R. R., & Wackermann, J. (2009). *Electrical neuroimaging*. Cambridge University Press.

Mormann, F., Lehnertz, K., David, P., & Elger, C. E. (2000). Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D: Nonlinear Phenomena*, *144*, 358–369.

Niepert, M., Ahmed, M., & Kutzkov, K. (2016). Learning convolutional neural networks for graphs. In *Proceedings of The 33rd International Conference on Machine Learning, PMLR*, 48, 2014–2023.

Nobukawa, S., Yamanishi, T., Kasakawa, S., Nishimura, H., Kikuchi, M., & Takahashi, T. (2020). Classification methods based on complexity and synchronization of electroencephalography signals in Alzheimer's disease. *Frontiers in Psychiatry*, *11*, 255.

Oltu, B., Akcsahin, M. F., & Kibarouglu, S. (2021). A novel electroencephalography based approach for Alzheimer's disease and mild cognitive impairment detection. *Biomedical Signal Processing and Control*, *63*, 102223.

Rodrigues, P. M., Bispo, B. C., Garrett, C., Alves, D., Teixeira, J. P., & Freitas, D. (2021). Lacsogram: A new EEG tool to diagnose Alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, *25*(9), 3384–3395.

Safi, M. S., & Safi, S. M. M. (2021). Early detection of Alzheimer's disease from EEG signals using Hjorth parameters. *Biomedical Signal Processing and Control*, *65*, 102338.

Sakkalis, V. (2011). Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Computers in Biology and Medicine*, *41*, 1110–1117. https://doi.org/10.1016/j.compbiomed.2011.06.020

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. *IEEE Transactions on Neural Networks*, *20*, 61–80.

Scheff, S. W., Price, D. A., Schmitt, F. A., DeKosky, S. T., & Mufson, E. J. (2007). Synaptic alterations in CA1 in mild Alzheimer disease and mild cognitive impairment. *Neurology*, *68*, 1501–1508.

Scheff, S. W., Price, D. A., Schmitt, F. A., & Mufson, E. J. (2006). Hippocampal synaptic loss in early Alzheimer's disease and mild cognitive impairment. *Neurobiology of Aging*, *27*, 1372–1384.

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, *30*, 83–98.

Si, Y., Wu, X., Li, F., Zhang, L., Duan, K., Li, P., Song, L., Jiang, Y., Zhang, T., Zhang, Y., Chen, J., Gao, S., Biswal, B., Yao, D., & Xu, P. (2019). Different decision-making responses occupy different brain networks for information processing: A study based on EEG and TMS. *Cerebral Cortex*, *29*, 4119–4129.

Smailovic, U., Koenig, T., Kåreholt, I., Andersson, T., Kramberger, M. G., Winblad, B., & Jelic, V. (2018). Quantitative EEG power and synchronization correlate with Alzheimer's disease CSF biomarkers. *Neurobiology of Aging*, *63*, 88–95.

Song, Z., Deng, B., Wang, J., & Wang, R. (2018). Biomarkers for Alzheimer's disease defined by a novel brain functional network measure. *IEEE Transactions on Biomedical Engineering*, *66*, 41–49.

Sperduti, A., & Starita, A. (1997). Supervised neural networks for the classification of structures. *IEEE Transactions on Neural Networks*, *8*, 714–735.

Tafreshi, T. F., Daliri, M. R., & Ghodousi, M. (2019). Functional and effective connectivity based features of EEG signals for object recognition. *Cognitive Neurodynamics*, *13*, 555–566. https://doi.org/10.1007/s11571-019-09556-7

Tait, L., Tamagnini, F., Stothart, G., Barvas, E., Monaldini, C., Frusciante, R., Volpini, M., Guttmann, S., Coulthard, E., Brown, J. T., Kazanina, N., & Goodfellow, M. (2020). EEG microstate complexity for aiding early diagnosis of Alzheimer's disease. *Scientific Reports*, *10*, 1–10.

Tavares, G., San-Martin, R., Ianof, J. N., Anghinah, R., & Fraga, F. J. (2019). Improvement in the automatic classification of Alzheimer's disease using EEG after feature selection. In *2019 IEEE international conference on systems, man and cybernetics (SMC)* (pp. 1264–1269). IEEE.

Terry, R. D., Masliah, E., Salmon, D. P., Butters, N., DeTeresa, R., Hill, R., Hansen, L. A., & Katzman, R. (1991). Physical basis of cognitive alterations in Alzheimer's disease: Synapse loss is the major correlate of cognitive impairment. *Annals of Neurology*, *30*, 572–580.

Trambaiolli, L. R., Spolaôr, N., Lorena, A. C., Anghinah, R., & Sato, J. R. (2017). Feature selection before EEG classification supports the diagnosis of Alzheimer's disease. *Clinical Neurophysiology*, *128*, 2058–2067.

Triggiani, A. I., Bevilacqua, V., Brunetti, A., Lizio, R., Tattoli, G., Cassano, F., Soricelli, A., Ferri, R., Nobili, F., Gesualdo, L., Barulli, M. R., Tortelli, R., Cardinali, V., Giannini, A., Spagnolo, P., Armenise, S., Stocchi, F., Buenza, G., Scianatico, G., … Babiloni, C. (2017). Classification of healthy subjects and Alzheimer's disease patients with Dementia from cortical sources of resting state EEG rhythms: A study using artificial neural networks. *Frontiers in Neuroscience*, *10*, 604.

Tylová, L., Kukal, J., Hubata-Vacek, V., & Vyšata, O. (2018). Unbiased estimation of permutation entropy in eeg analysis for Alzheimer's disease classification. *Biomedical Signal Processing and Control*, *39*, 424–430.

Van Diessen, E., Numan, T., Van Dellen, E., Van Der Kooi, A. W., Boersma, M., Hofman, D., van Lutterveld, R., van Dijk, B. W., van Straaten, E. C. W., Hillebrand, A., & Stam, C. J. (2015). Opportunities and methodological challenges in EEG and MEG resting state functional brain network research. *Clinical Neurophysiology*, *126*, 1468–1481.

Van Mierlo, P., Papadopoulou, M., Carrette, E., Boon, P., Vandenberghe, S., Vonck, K., & Marinazzo, D. (2014). Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization. *Progress in Neurobiology*, *121*, 19–35.

Wan, L., Huang, H., Schwab, N., Tanner, J., Rajan, A., Lam, N. B., Zaborszky, L., Li, C. S. R., Price, C. C., & Ding, M. (2019). From eyes-closed to eyes-open: Role of cholinergic projections in EC-to-EO alpha reactivity revealed by combining EEG and MRI. *Human Brain Mapping*, *40*, 566–577.

Wendling, F., Ansari-Asl, K., Bartolomei, F., & Senhadji, L. (2009). From EEG signals to brain connectivity: A model-based evaluation of interdependence measures. *Journal of Neuroscience Methods*, *183*, 9–18.

World Health Organization. (2021). *Dementia*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/dementia.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *32*, 4–24.

Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint: arXiv1709.04875.

Yu, H., Lei, X., Song, Z., Liu, C., & Wang, J. (2019). Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification. *IEEE Transactions on Fuzzy Systems*, *28*, 60–71.

Zeng, D., Huang, K., Xu, C., Shen, H., & Chen, Z. (2020). Hierarchy graph convolution network and tree classification for epileptic detection on electroencephalography signals. *IEEE Transactions on Neural Networks and Learning Systems*, *13*, 955–968.

Zhang, S., Chen, D., Tang, Y., and Zhang, L. (2021). Children ASD evaluation through joint analysis of EEG and eye-tracking recordings with graph convolution network. 15 651349

Zhao, Y., Dong, C., Zhang, G., Wang, Y., Chen, X., Jia, W., Yuan, Q., Xu, F., & Zheng, Y. (2021). EEG-based seizure detection using linear graph convolution network with focal loss. *Computer Methods and Programs in Biomedicine*, *208*, 106277.

Zhao, Y., Zhao, Y., Durongbhan, P., Chen, L., Liu, J., Billings, S. A., Zis, P., Unwin, Z. C., De Marco, M., Venneri, A., Blackburn, D. J., & Sarrigiannis, P. G. (2019). Imaging of nonlinear and dynamic functional brain connectivity based on EEG recordings with the application on the diagnosis of Alzheimer's disease. *IEEE Transactions on Medical Imaging*, *39*, 1571–1581.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Shan, X., Cao, J., Huo, S., Chen, L., Sarrigiannis, P. G., & Zhao, Y. (2022). Spatial–temporal graph convolutional network for Alzheimer classification based on brain functional connectivity imaging of electroencephalogram. *Human Brain Mapping*, *43*(17), 5194–5209. https://doi.org/10.1002/hbm.25994