IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

EEG Functional Connection Analysis Based on the Weight Distribution of Convolutional Neural Network

JINGLONG WU[1], PEIWEN HUANG[1], TIANTIAN LIU[1], GO RITSU[1], DUANDUAN CHEN[1], AND TIANYI YAN[1]

[1]Beijing Institute of Technology, Beijing, 100081 China

Corresponding author: TIANYI YAN (e-mail: yantianyi@ bit.edu.cn).

This work was supported by the National Key R&D Program of China (grant number 2018YFC0115400), the National Natural Science Foundation of China (grant numbers 81671776, 61727807, 61633018), the Beijing Municipal Science & Technology Commission (grant numbersZ191100010618004).

ABSTRACT Functional connections are commonly used when exploring the human brain, especially in brain data analysis. However, most of the studies concentrate on traditional statistical analysis. In this paper, we innovatively combined the functional connection with the deep learning algorithms and analysed the matrices after the weight distribution of each layer of the convolutional neural network (CNN) to obtain the connections that play a vital role in the classification. The electroencephalogram (EEG) data used in this paper was acquired through a visual mismatch negativity (MMN) experiment. When dealing with this data, each electrode was regarded as a node in the network, and the phase lag index (PLI) was calculated to construct the functional connection matrices, which were used as inputs for the CNN classification and feature extraction. The matrices after the weight distribution were further analysed by means of graph theory. In this paper, the classification accuracy for deviation and standard stimuli are over 95%, and the theta band achieved the highest accuracy. Through the distributed matrices, we found that there are two regions that obtained larger weights from the convolutional layers, i.e., the temporal lobe and the occipital region. The occipital region is related to our visual experiment, and the temporal lobe region is connected with MMN mechanism. We also considered the strategy of the three-layer CNN according to weight distribution processing.

INDEX TERMS Neural networks, Electroencephalography, Graph theory, Complex networks, Weight distribution, Functional connection.

I. INTRODUCTION

The convolutional neural network (CNN) was proposed by Geoffrey E. Hinton et al. in 2012[1]and has been rapidly developed in computer vision, speech recognition and other fields, and has become a mature algorithm. In medical image data, CNN has been widely used in lesion area identification and disease diagnosis, such as thoraco-abdominal lymph node detection and interstitial lung disease classification of computed tomography (CT) images[2], brain lesion segmentation based on magnetic resonance imaging (MRI) [3], and cognitive and motor developmental outcome score prediction from structural brain networks of infants born preterm based on diffusion tensor imaging (DTI) [2]. As a convenient method for exploring the brain, EEG also has received increasing attention along with deep learning algorithms and has made great progress. These studies include brain-computer interface (BCI) [4-6], brain disease diagnosis [7] and prediction [8-11] and brain state monitoring [12-15]. There are also papers that attempt to establish functional connectivity for electroencephalography (EEG) [16-18], but relatively fewer compared to the former, and using deep learning methods to analyse EEG for brain mechanism exploration is even rarer. One reason is that EEG is a one-dimensional signal which makes it less intuitive than other medical image data such as MRI or DTI; another reason is that the processing of deep learning is often presented in the form of 'black boxes' that make it harder for people to understand. However, compared with other brain image data, the acquisition of EEGs is much more convenient, has lower

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

cost and is more acceptable to ordinary people, so it is more likely to include larger databases, which is very significant to deep learning.

With the development of CNN, there have been increasing studies on the decoding of convolutional networks, attempting to explain the convolutional neural network and to answer why and how convolutional layers can extract features [19-22] In terms of EEG, similar studies have attempted to explain the role of convolutional layers in EEG feature extraction. Schirrmeister RT et al. used CNN to classify brain motor imagery by an end-to-end approach and making some visual attempts [23]. Their research proved that when extracting features, convolutional layers focus on spectral power in the alpha, beta, and high gamma frequencies. However, they artificially limited the extracted features to the power spectral density, treating each electrode as independent, and ignoring the communication between different regions of the brain, and thus, the representation of the EEG was not consummate.

In this article, we focus on two issues, whether functional connectivity can be used as in EEG classification and what the convolutional layers do during the classification. Traditional EEG features include time, frequency and time-frequency domain features [24-27], frequently appearing in the prediction and classification of EEG signals. However, for brain mechanism research, these features neglect the communication of different electrodes or brain regions, which is vital and should be noticed. Information exchange between regions can be assessed by time coupling or time dependence between signals collected in different regions, which is also referred to as "functional connectivity" [28, 29]. The phase lag index (PLI) has been commonly used to evaluate functional connections in recent years. Compared to the component of coherency (IC) and phase coherence (PC), PLI minimizes the effects of common source signals, which is of great significance for EEG signals [18]. Wang Luyao et al. used functional connectivity built by PLI to investigate whether it would affect audio-visual integration during normal ageing [30]. Tóth, Brigitta et al. used PLI to measure large-scale phase synchronization in six frequency bands, and found that early cortical networks in infants were significantly more hierarchical and more cost-effective [31]. Yang Yuan et al. used PLI of different frequency bands to study the dynamic functional connections in facial perception, and found that significant differences between facial and non-facial perception are mainly present in θ-band connectivity [32]. There are also some articles combining functional connectivity with machine learning methods. Wajid Mumtaz used EEG-derived synchronization likelihood (SL) features as inputs for support vector machines (SVM), naive Bayes (NB), and logistic regression (LR) classifiers, to automatically diagnose depression [33]. They also used these classifiers based on functional connectivity to detect alcohol use barriers [34]. Both achieved high accuracy. However, they did not further analyse the functional connections. For deep learning, there is still much left to do.

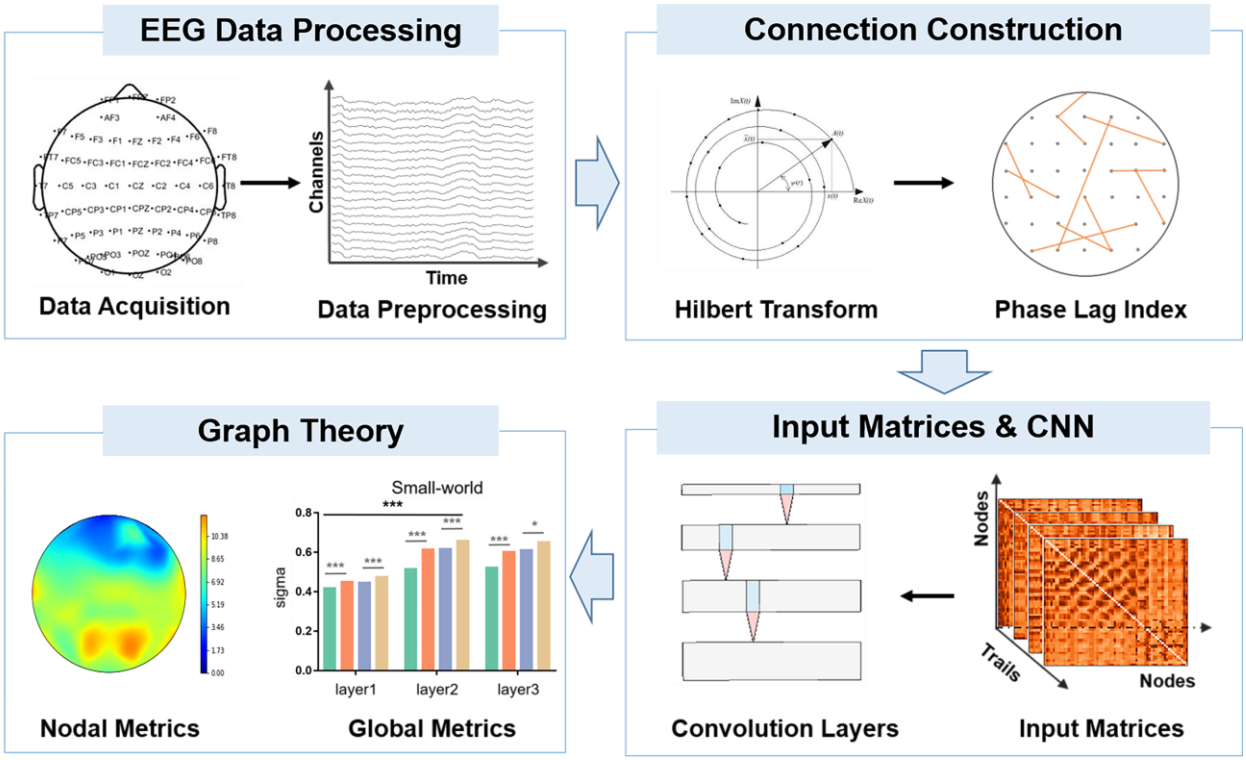


FIGURE 1. **The overview of data processing in this paper.**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

After building a functional connection network, we analysed the connectivity of the edges by means of graph theory to further understand the characteristics of the constructed network. Graph theory describes mathematical methods applied to representations of networks reduced to their essence: vertices (nodes) and edges (connections) [35]. Studies have shown that information exchange between brains may be limited by optimal processing capabilities when maintaining cost efficiency and self-recovery [36, 37]. In the exploration of the brain mechanism, graph theory was first applied to MRI data [38-40], and then researchers introduced it into EEG analysis; these studies included the exchange of information through the frontal to the occipital lobe during sleep [41], the development of neonatal brain network connections [31], and the differences between the functional connections of people with and without schizophrenia.

In summary, in terms of brain analysis based on EEG, the functional connection is a widely used method because of the representation of information exchange between different regions in the brain. However, most of these studies are statistical analysis, and only a small part of them attempt traditional machine learning methods. The application of deep learning to EEGs has been widely used in EEG classification and prediction in brain analysis, and a few studies have attempted to explain convolutional layers, but information exchange between electrodes has been ignored. In this paper, we calculate the PLI as an indicator to build a functional connection, and analyse the output of the convolutional layers to explore the role of convolutional layers in feature extraction. The output matrices are further analysed through graph theory. From all these analyses, we found two important regions. The processing flow is shown in Fig. 1.

II. MATERIALS AND METHODS

A. DATA ACQUISITION

1) SUBJECTS AND STIMULATION

Fifteen students from the Beijing Institute of Technology, including six women and nine men, aged between 20 and 23, participated in the experiment. All subjects were in good physical and mental health and right-handed, with normal or corrected-to-normal vision. Prior written notification of the study was given before the experiment began.

As shown in Fig. 2(a), pictures were placed 70 cm from the subject, with the viewing angle of 3.68°×3.42°, and display time of 100 milliseconds. The arrows in the picture appear on both sides of the cross at the centre of the screen, with a 500-millisecond interval between each stimulus. Standard stimuli (Std) and deviation stimuli (Dev) were placed symmetrically from the target area to reduce the impact on the subject. The standard stimuli in the experiment were vertical upward arrows, accounting for 80% of the stimulation sequence, and the deviation stimuli were horizontal left and horizontal right arrows, each accounting for 10% of the stimulation sequence. In the course of the experiment, under the premise of ensuring at least two standard stimuli between every two deviation stimuli, the order of stimuli was pseudo-random. To minimize the influence of other environmental factors on the experiment, the cross in the middle of the screen randomly became larger or smaller during the stimulation process, and the subject was required to press the left or right button as soon as the cross changed.

2) EEG RECORDING

EEG signals were continuously recorded using the NeuroLab's digital amplifier and electrode cap with 64 Ag/AgCl electrodes placed according to the extended International 10-20 system (Fig. 2(b)), and the tip of the nose was used as a reference. The signal was filtered between 1-100 HZ with a sampling rate of 1,000 Hz. The electrodes

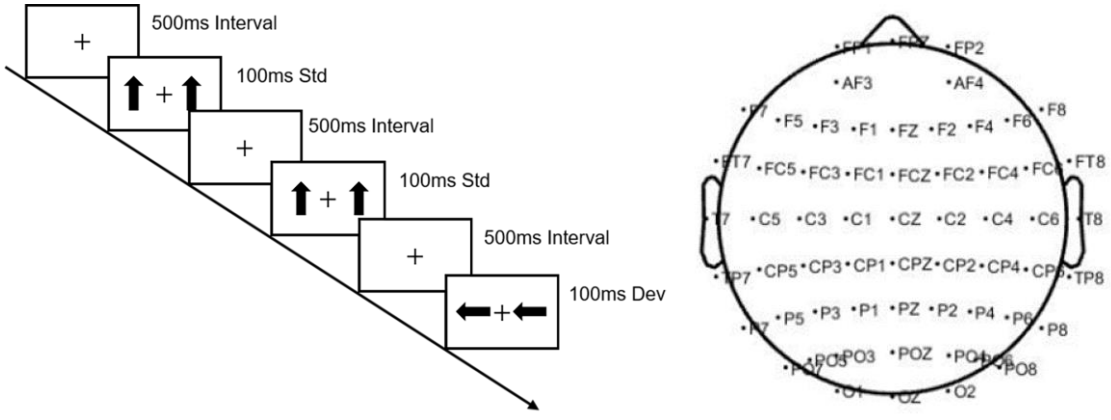


FIGURE 2. Experimental paradigm and electrode positions. (a) The intervals were 500 milliseconds, and each picture was shown for 100 milliseconds. The stimuli were random, and there were at least two standard stimuli between deviation stimuli. (b) Electrodes were placed according to the 10-20 system.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

recording vertical electro-oculography (VEOG) and horizontal electro-oculography (HEOG) were placed on the upper and lower sides of the right eye and 10 mm from the lateral canthus.

However, we found that the EEG of two subjects were abnormal and unable to use due to artificial problems during the experiment. Therefore, we used the other 13 subject's data in our analysis.

3) EEG PREPROCESSING

In the existing research, the analysis of visual MMN is mainly in low-frequency (under 30HZ) EEG, and we adopted such filtering frequency in our previous and current processing [46]. Therefore, we used a traditional 30 Hz low-pass filter. The independent component analysis (ICA) in Matlab R2013a with the open-source toolbox EEGLAB was effectively used for artifacts removal in the EEG. In addition, we manually deleted the artifact components such as electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG) after ICA.

We also divided the data into different frequency bands according to the EEG rhythm (theta: 4-7 HZ; alpha: 8-13 HZ; beta: 14-30 HZ).

B. FUNCTIONAL CONNECTION CONSTRUCTION

The PLI was first proposed by Stam, Cornelis J et al. [18]; they believed that the PLI is more suitable for measuring the strength of the connection between EEG electrodes due to the volume conduction effect in the EEG electrodes and the presence of the active reference electrodes compared with IC and PC, which means the PLI can obtain a stable phase estimate even in the presence of common source signals. In our previous practice, it has also proved that PLI is stable and effective in building functional connection. Therefore, in this article, we also used the PLI as an indicator to build a functional connection.

1) PHASE LAG INDEX

We used the Hilbert Huang transform (HHT) to obtain the instantaneous phase of the signal and to divide the signal into theta, alpha and beta bands according to the EEG rhythm.

The central idea in calculating the PLI is to discard the phase difference centred around 0 and π, i.e., when the distribution is centred around the zero phase difference, the asymmetry index of the phase difference distribution can be defined. The phase difference $\Delta\phi$ has different probabilities in the interval $[-\pi, 0]$ and the interval $[0, \pi]$. If there is no coupling between two signals, the distribution is called 'flat'. When the distribution is flat, or the median phase difference is equal to or centred around 0 or π, the asymmetry of the phase difference distribution can be obtained by (1).

$$PLI = |< sign[\Delta\phi(t_k)] >| \quad (1)$$

The range of the PLI is between $[0,1]$. When the PLI is 0, there is no coupling between the signals, and the phase difference is centred around 0 or π. The stronger the phase lock between signals, the closer the value of the PLI is to 1. When the value of the PLI is 1, it means the two signals are perfectly phase-locked.

2) INTERREGIONAL PHASE SYNCHRONIZATION

There are usually two methods for calculating the PLI of the EEG, as shown in Fig. 3. One method is connectivity over time, which is usually more sensitive in resting EEG or longer-lasting EEG data; however, its temporal precision is poor. The other method is connectivity across trials, which has higher temporal precision and can better represent the transients in the connection [47].

We calculated both the PLI of connectivity over time and connectivity across trials and then used the t-distributed stochastic neighbour embedding (t-SNE) algorithm to reduce the data to two dimensions [48]. The results of one subject are shown in Fig. 4. We find that data obtained by the method of connectivity over trials is separable, whereas data calculated by connectivity over time is not due to the instability of EEG in different trials. In addition, because of the nature of MMN experiments, there is always sample imbalance in the data (standard stimulus and bias stimulus ratio is 5:1 in ours), and connectivity over trials can solve this imbalance and make each category of stimuli have the same number of samples. In this paper, because there are two kinds of standard stimulus (left arrows and right arrows) and only one kind of deviation stimulus (upward arrow), the ratio of the final samples of the standard stimulus to the deviation stimulus is 2 to 1. In addition, connectivity over trials can reduce the differences between individuals because of the accumulation of multiple trials, which is more conducive to our further analysis of the EEG mechanism related to MMN. The resulting connection matrices are shown in Fig. 5. We considered the graphs of theta, beta and alpha bands as the three channels of the RGB maps and used them as inputs for the convolutional network.

C. CONVOLUTIONAL NEURAL NETWORK FOR EEG

According to the method mentioned in the previous section, for each subject, we obtain two types of deviation stimuli and

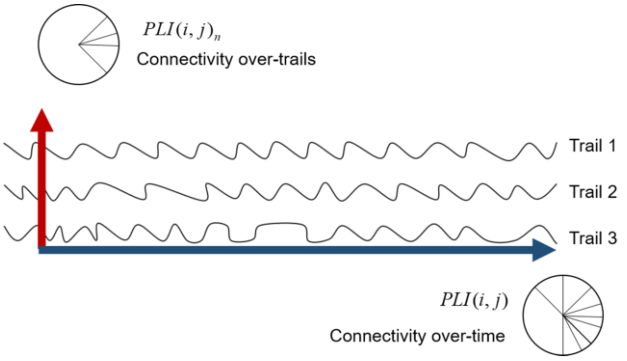


FIGURE 3. Two different methods for calculating PLI. Connectivity over-trials simultaneously calculates phase difference between two adjacent trials to obtain the average value after accumulating. Connectivity over time calculates the difference between every two nodes in each trial.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

one type of standard stimuli, with 800 samples for each type, so there are 31,200 samples in the total of 13 people. Since the calculation method of PLI reduced the difference between the subjects, we did not need to separately construct a model for each subject but built a model for all the subjects. In our cross-validation, data from one subject was completely separated to ensure test sample independently during training. We thought that it is defective to randomly select test samples crossing all subjects due to high similarity of the EEG sequence from same subject. In addition, considering small number of subjects, we only used one subject's data as testing sample to complete cross-validation each test. In addition, the other data were left for training in random order to complete the cross-validation.

A simple structure of a neural network includes an input layer, an implicit layer and an output layer. Each layer of the network has multiple neurons, which are mapped to the next layer through an activation function between each neuron. There is a corresponding weight between each neuron, and the output is the classification category. In this paper, we built a convolutional network using a basic structure, as shown in Fig. 6. Each block consisted of a convolutional layer with 3*3 filters, a linear rectification function, and a pooling layer. We used three convolutional blocks and two fully connected layers to complete the classification of EEG signals. To maintain the correspondence between the output pattern and the electrode position, we set the step size to 1 with padding in each of the convolutional layers and the first two pooling layers with padding. In addition, we chose max pooling in every pooling layer. Considering the data, number of parameters and the computation time, we did not build deeper networks.

In the training process, we set the batch size to 50, i.e., 50 samples per training step. In addition, these samples and the average weight distribution map of the convolutional layer outputs were randomly stored at the beginning, the middle and the end of the model training. Since the output matrix does not have a figurative meaning and only represents the weight distribution in this paper, we also averaged the different convolutional kernels of each layer and obtained the matrix of the weight distribution for each output layer. Although keeping the image in the same shape increases the number of calculations, it ensures sufficient position information, so we can carry out further analysis and comparison. All calculations were performed on NVIDIA chips using TensorFlow1.4 as the framework.

D. GRAPH THEOTRTICAL ANALYSIS

In this article, a connection matrix was built by regarding each electrode as a node, so the PLI value can be seen as the connection strength between two nodes. Therefore, all the nodes constituted a graph. Therefore, we applied graph theory for further analysis. We took the network connection matrix as a no direction and no weight graph and limited the threshold of the connection strength to obtain binarized matrices. A value of 1 means that there is a connected edge between the two nodes, and vice versa. We set a total of ten thresholds from the 5% to the top 50% connection. However, to eliminate the influence of threshold selection, we chose the area under the curve (AUC) as the result of the final statistical analysis. In this article, the graph parameters in the global and node perspectives were calculated. All calculations were performed in the MATLAB toolbox GRETNA.

1) SMALL-WORLD METRICS

The clustering coefficient (Cp) measures the degree of grouping for a network. It is a basic parameter in graph theory that indicates the possibility that adjacent nodes of a node are connected to each other. The clustering coefficient of one node is the ratio of the number of edges connected to the neighbour of the node to the number of possible maximum connected edges. As shown in (2), e_i represents actually existing edges, $k_i(k_i-1)$ represents the maximum number of edges that may exist in adjacent nodes. The average of the clustering coefficients of all nodes in the network is the clustering coefficient of the network. Characteristic path length (Lp) is the length of the path with the least number of edges between two nodes. Information could be transmitted faster through the shortest path. As shown in (3), $L_{p,i}$ represents the characteristic path length of each node and it is average value of the shortest path length between this node and the others in the network. In addition, the average value of all nodes is Lp for the whole network.

$$C_{p,i}=\frac{1}{n}\sum_{i\in N}\frac{2e_i}{k_i(k_i-1)} \tag{2}$$

$$L_p=\frac{1}{n}\sum_{i\in N}L_{p,i}=\frac{1}{n}\sum_{i\in N}\frac{\sum_{j\in N,j\neq i}d_{i,j}}{n-1} \tag{3}$$

In general, regular networks have long path lengths and high clustering coefficients, random networks have short path lengths and low clustering coefficients, and small-world networks have higher clustering coefficients than random networks and shorter path lengths than the regular network [49]. Therefore, small-world networks can achieve efficient information delivery at a relatively low cost. Visually, due to the characteristics of high-network clustering of short paths in small-world networks, clusters (a small group of nodes connected to each other) and nodes will be connected by only a few groups. Usually, we use the small-world indicator sigma to measure whether a network is a small-world network. Generally, if sigma is greater than 1, the network is a small-world network, as shown in (4).

$$\sigma=\frac{\frac{C_p}{C_{p,rand}}}{\frac{L_p}{L_{p,rand}}} \tag{4}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

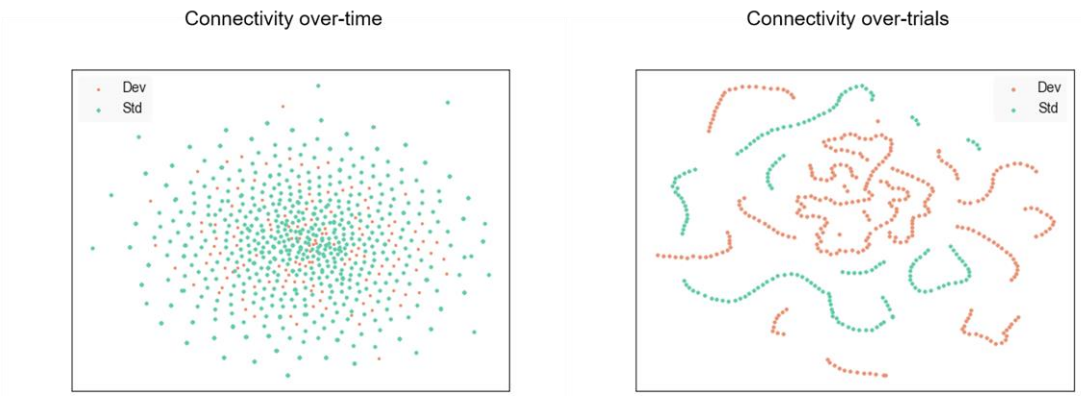


FIGURE 4. **The results of dimensionality reduction using the t-SNE algorithm for one subject. (a) Result obtained by connectivity over time. (b) Result obtained by connectivity over-trials.**

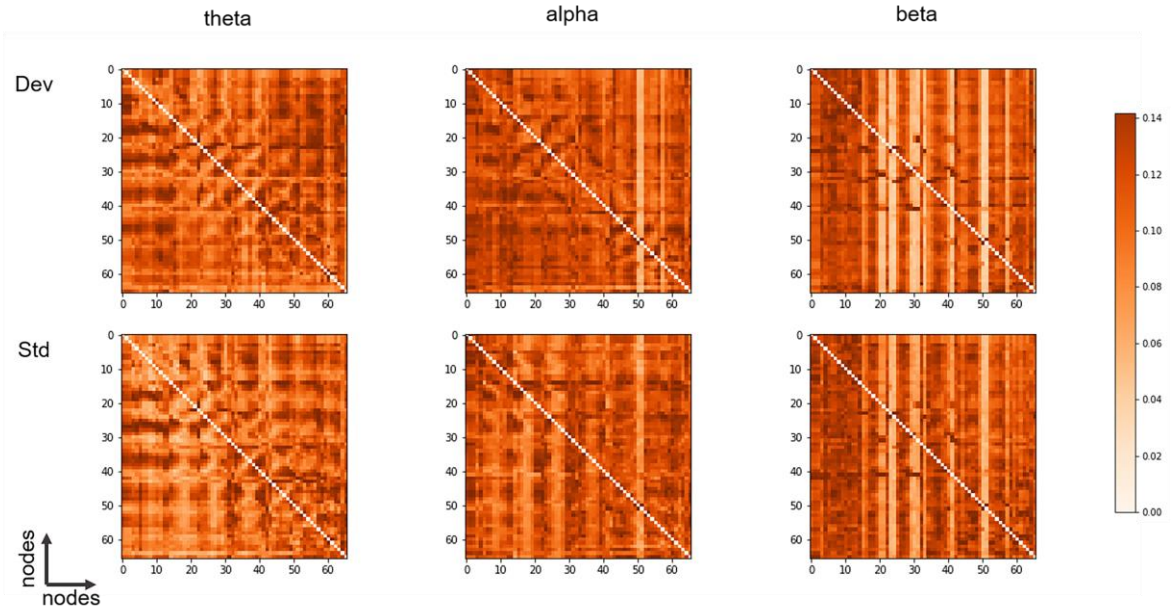


FIGURE 5. **Sample of inputs. Three channels of input data are composed of connectivity matrices of theta, alpha and beta bands. The data was divided into standard stimuli and deviation stimuli.**

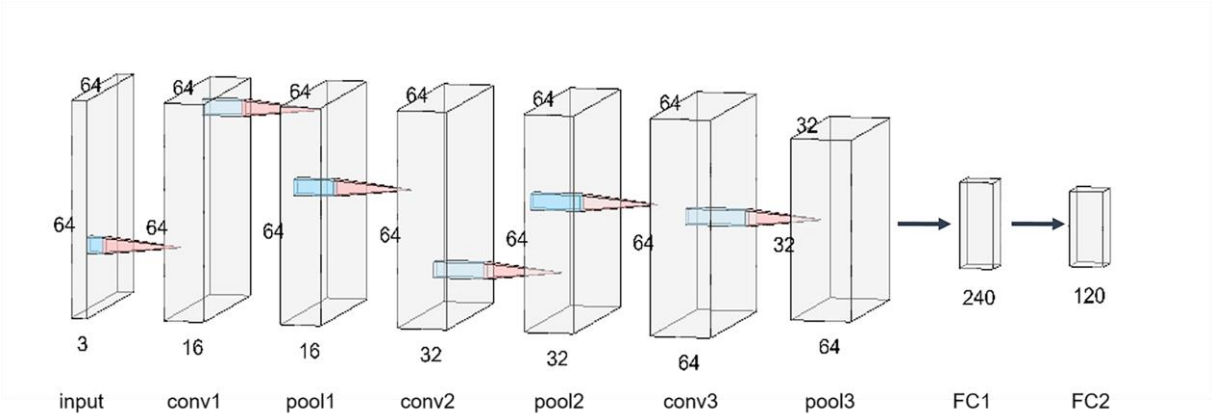


FIGURE 6. **Convolutional network structure. The network consists of three convolutional layers and two fully connected layers. Convolutional layers include convolutional (including ReLU activation function) and pooling layers. Numbers indicates the size of each layer.**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

TABLE I
RESULTS OF DIFFERENT KERNEL SIZES

Subjects	Ksize 1*1	Ksize 3*3	Ksize 5*5	Ksize 7*7	Ksize 9*9
Sub1	0.979	0.979	0.979	0.979	0.979
Sub2	0.979	0.979	0.979	0.979	0.914
Sub3	0.969	0.970	0.971	0.971	0.969
Sub4	0.978	0.974	0.968	0.976	0.927
Sub5	0.979	0.979	0.979	0.979	0.979
Sub6	0.975	0979	0.979	0.979	0.967
Sub7	0.753	0.767	0.735	0.711	0.737
Sub8	0.979	0.979	0.979	0.979	0.976
Sub9	0.979	0.979	0.975	0.978	0.977
Sub10	0.979	0.979	0.979	0.979	0.979
Sub11	0.977	0.975	0.979	0.968	0.963
Sub12	0.978	0.979	0.977	0.979	0.976
Sub13	0.979	0.979	0.979	0.979	0.979
Average	0.960	0.961	0.958	0.957	0.948

Ksize means kernel size

TABLE II
RESULTS OF DIFFERENT FREQUENCY BANDS

Subjects	beta	alpha	theta
Sub1	0.979	0.979	0.979
Sub2	0.976	0.960	0.968
Sub3	0.979	0.973	0.978
Sub4	0.979	0.979	0.978
Sub5	0.979	0.979	0.979
Sub6	0.979	0.979	0.979
Sub7	0.684	0.780	0.769
Sub8	0.966	0.979	0.979
Sub9	0.921	0.975	0.976
Sub10	0.979	0.979	0.979
Sub11	0.970	0.949	0.958
Sub12	0.963	0.966	0.973
Sub13	0.979	0.979	0.979
Average	0.949	0.958	0.960

2) EFFICIENCY METRICS

Efficiency is another fundamental attribute of the network, including global efficiency and local efficiency. As shown in (5), Global efficiency (Eg) is a concept that corresponds to characteristic path length and is numerically approximated by its reciprocal. It is proposed because when there is no connection between two nodes, L is infinity. Both metrics measure the global transmission capability of the network. The shorter the path length, the higher the global efficiency is, and the faster the rate of information transfer in the network. Local efficiency (Eloc) is a concept corresponding to clustering coefficients. The clustering coefficient considers only the direct connection of adjacent nodes, and the local efficiency considers the neighbouring nodes of a node as subgraphs. As shown in (6), when a_{ij} is one means there is an edge between two nodes and zero means no connection. $k_i(k_i-1)$ indicates the maximum number of connected edges that may exist in node i. And $d_{jh}(N_i)$ is the shortest path length from node j to node h, if only node i is included. Local efficiency of the whole network is the mean value of all nodes. Both the clustering coefficient and local efficiency measure the local information transmission capability of the network.

Global efficiency reflects the transport efficiency in a network working in parallel, and local efficiency describes the subgraph composed of the neighbour nodes of a node, which reflects the fault tolerance of the network, i.e., the information transmission rate of the nodes adjacent to the node when it is removed.

$$E_g = \frac{1}{n}\sum_{i\in N} E_{g,i} = \frac{1}{n(n-1)}\sum_{i,j\in N, i\neq j} d_{i,j}^{-1} \tag{5}$$

$$E_{loc} = \frac{1}{n}\sum_{i\in N} E_{loc,i} = \frac{1}{n}\frac{\sum_{j,h\in N, j\neq i} a_{ij}a_{ih}[d_{jh}(N_i)]^{-1}}{k_i(k_i-1)} \tag{6}$$

III. RESULT

A. CLASSIFICATION RESULT

We used different sizes of convolutional kernels for classification. Considering later comparison, we set the size of the convolutional kernels in each layer to be the same, but the number of filters per layer was different, as shown in Fig.6. The accuracy is shown in Table 1. In our experiments, the classification of 3*3 convolutions worked best, and as the size increased, the classification accuracy decreased. The impact of the size of the convolutional kernels on the convolutional process is not yet conclusive. In this article, we used 3*3 convolutional kernels for analysis.

We also studied the effect of different frequency bands on the classification results. We used single-channel data as input and found that the accuracy in the theta band was the highest, followed by the alpha band, and the beta band was relatively low. The results are shown in Table 2. These results are consistent with the findings in previous studies [44, 46, 50, 51] and indicate that theta band plays a vital role in MMN research. For further exploration, we divided the whole brain into six regions of interest (ROIs). We rearranged the electrodes according to the division of ROIs and averaged the electrodes in each region. The location of each ROI and the result after division is shown in Fig. 7. We find that the theta and alpha bands have a more pronounced connection in the bilateral temporal lobe and the central region, and the theta band is stronger than the alpha band. The beta band does not show this, so we assume that the exchange of information between the temporal lobes and the central area had a significant effect in the MMN experiment. In addition, we further studied this through convolutional output.

B. CONVOLUTIONAL LAYERS OUTPUT

To better represent the weight distribution, we de-averaged the output matrix of each convolutional layer. The results are shown in Fig. 8.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

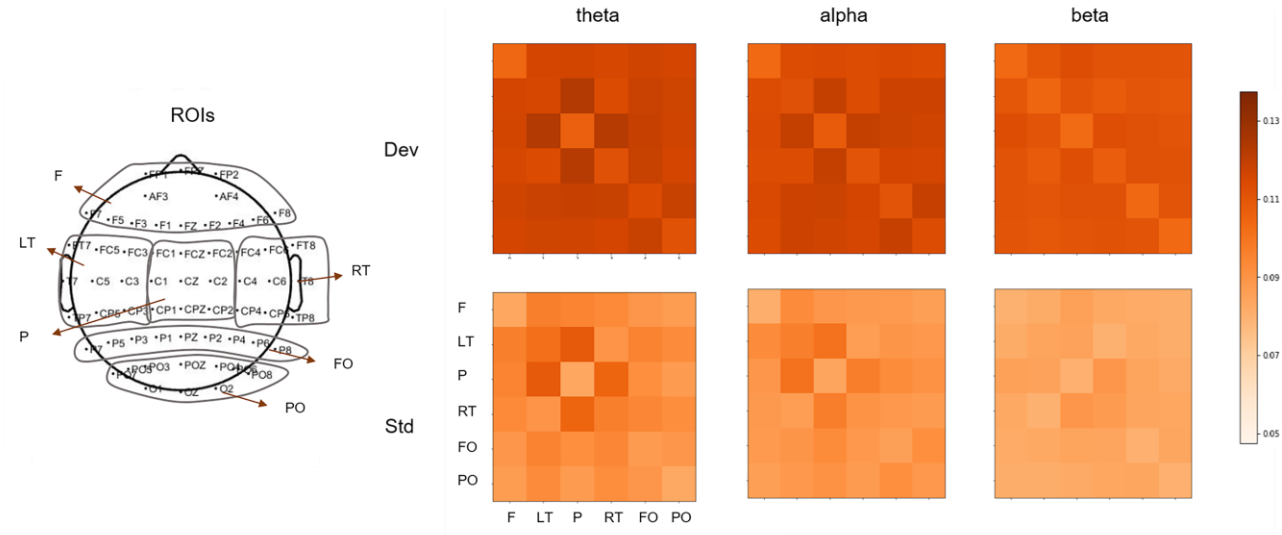


FIGURE 7. **ROI division and functional connection matrices after rearrangement. F represents the frontal lobe, LT and RT represent the left temporal lobe and the right temporal lobe, respectively, P represents the parietal lobe, FO and PO represent the front occipital lobe and the post occipital lobe, respectively. Redder colour represents stronger connection.**

From the output of the first layer, we find that the connections inside the left and right temporal lobes gradually strengthened. Additionally, the connection between the temporal lobe and the central region was preserved. For deviation stimuli, the connections inside the front occipital lobe and the post occipital lobe strengthened, and the connection between them was retained. In addition, the connection weight of the deviation stimuli was more intense than the standard. The output of the second layer shows that for the deviation stimuli, the weight of the internal connection of the temporal and occipital lobes was further enhanced. Two distinct areas with greater weight formed. For standard stimuli, the weight of the temporal lobes and the occipital lobes were obvious in the beginning, but in the training process, the model attempts to weight the connection between the occipital region and the frontal lobe. This phenomenon is worth studying and is related to the analysis of the second convolutional layer output matrices. The third layer shows that the weight of the deviation stimuli and standard stimuli in the temporal and occipital lobes were enhanced more obviously. For deviation stimuli, the connection inside the temporal and occipital lobes was stronger, and for standard stimuli, it was stronger between the temporal and parietal lobes. However, in general, the regions were obvious.

C. GRAPH THEORETICAL ANALYSIS RESULT

In our paper, we focus on the small-world and efficiency of connection matrices, analysing the network from global and nodal aspects. Fig. 9 and Fig. 10 show the changes in the output matrices of each layer during the training process (more details are included in the supplementary materials).

1) SMALL-WORLD METRICS

The clustering coefficient and characteristic path length are two factors that restrict each other in small-world attributes. Fig. 9 shows the changes in small-world metrics during the training process. For deviation stimuli, as the convolutional layers deepen and the number of iterations increase, the clustering coefficient increases gradually. The deviation stimuli are significant between layers and the beginning and ending iterations. This represents that the concentration of the network is strengthened and the capability of information exchange is increasing. However, from the characteristic path length, we also find that there is an increase in the characteristic path length, which may be caused by the deepening convolutional layers that increase the weight distribution within the regions. This represents the overall information processing capability of the network is reduced. For standard stimuli, the deepening of the convolutional layer increases the clustering coefficient significantly, but at the second and third layers, there is a decrease with the increase in the number of iterations. In addition, this trend is significant at the second convolutional layer. Therefore, we speculate that this reduction is related to the weight distribution strategy change at the second convolutional layer, which weakens the weight within regions. At the same time, the path length decreases significantly in the second and third layers. We therefore, further infer that this strategy change strengthens longer connections.

Small-world networks synthetically measure these two attributes, and sigma is a commonly used standard for small-world networks. For the deviation stimuli, the trend is basically consistent with the clustering coefficient. For the standard stimuli, in the second layer, as the number of iterations increases, there is a decrease in sigma, but in the third layer, this situation no longer exists. Small-world

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

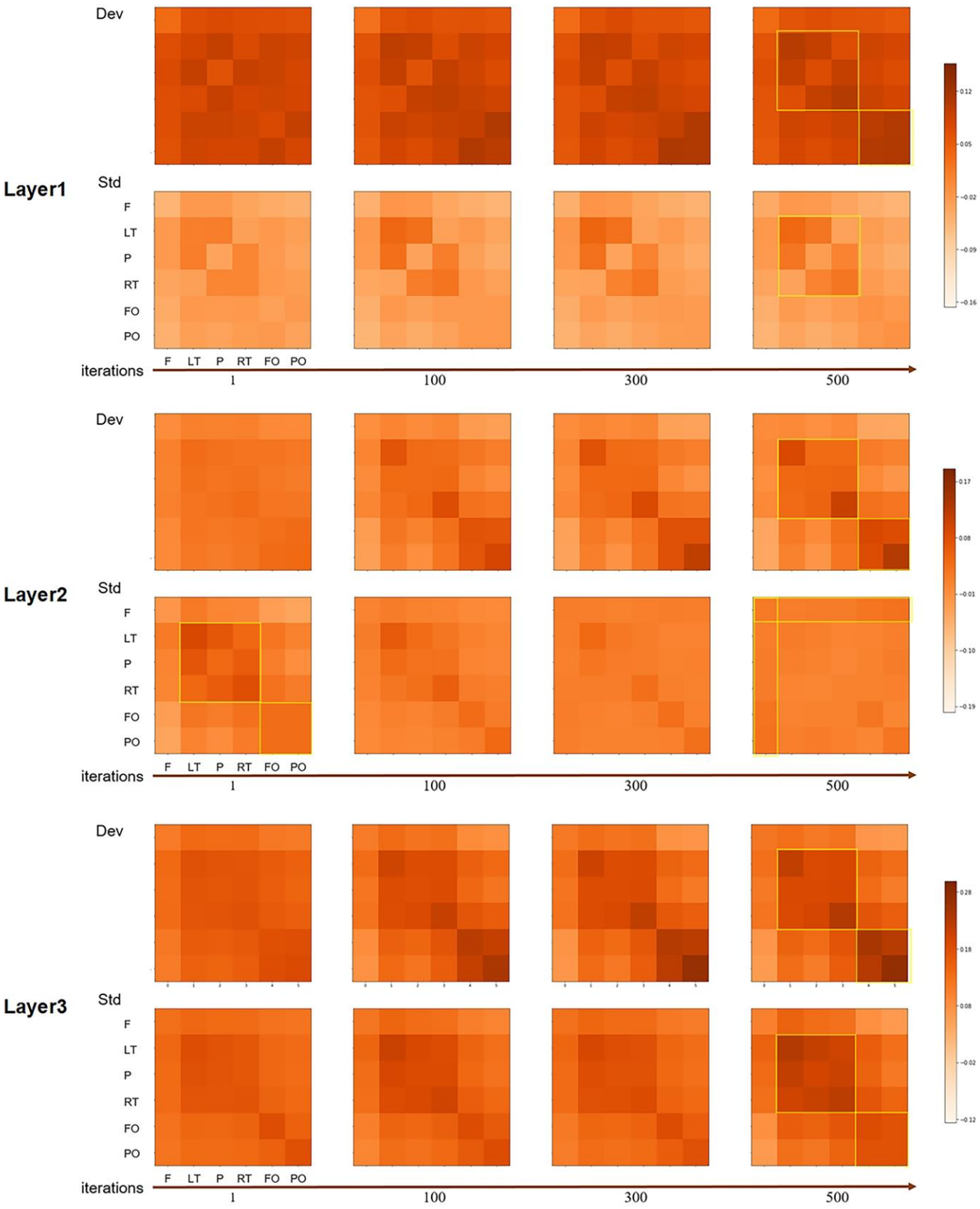


FIGURE 8. **The output matrices of three convolutional layers. The horizontal and vertical coordinates mark the six ROIs. The upper is deviation stimuli, and the lower is standard stimuli. Matrices from left to right are the beginning to the end of iterations, and each iteration represents a training batch of 50 samples.**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

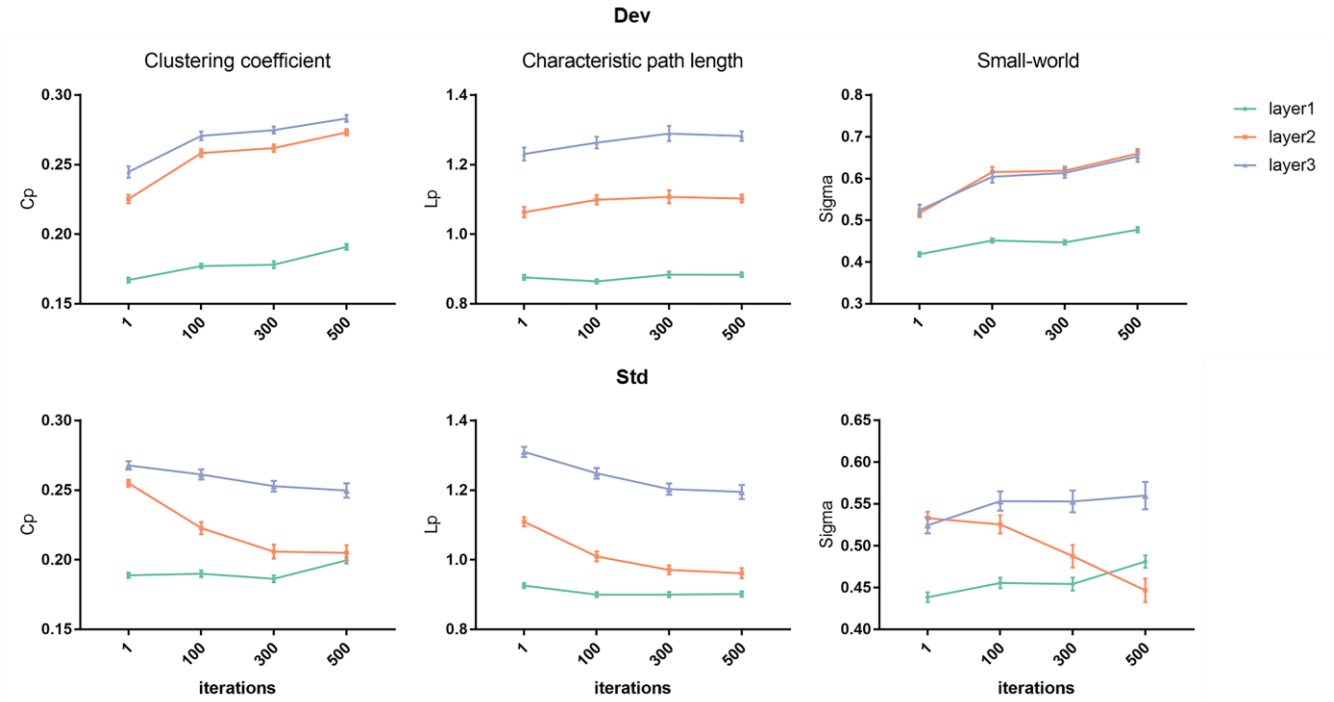


FIGURE 9. Small-world analysis. Different colours represent different layers. Samples are obtained from 65 trainings (5 rounds, 13 folds each round). The vertical axis represents different metrics, and the horizontal axis represents the number of iterations, and 50 samples per iteration.

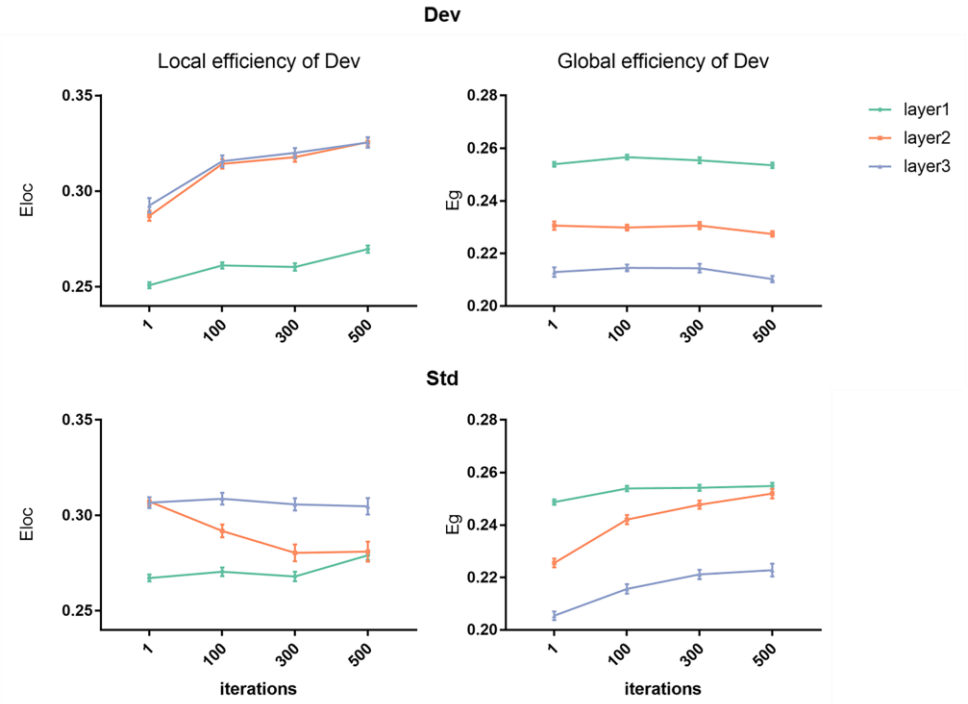
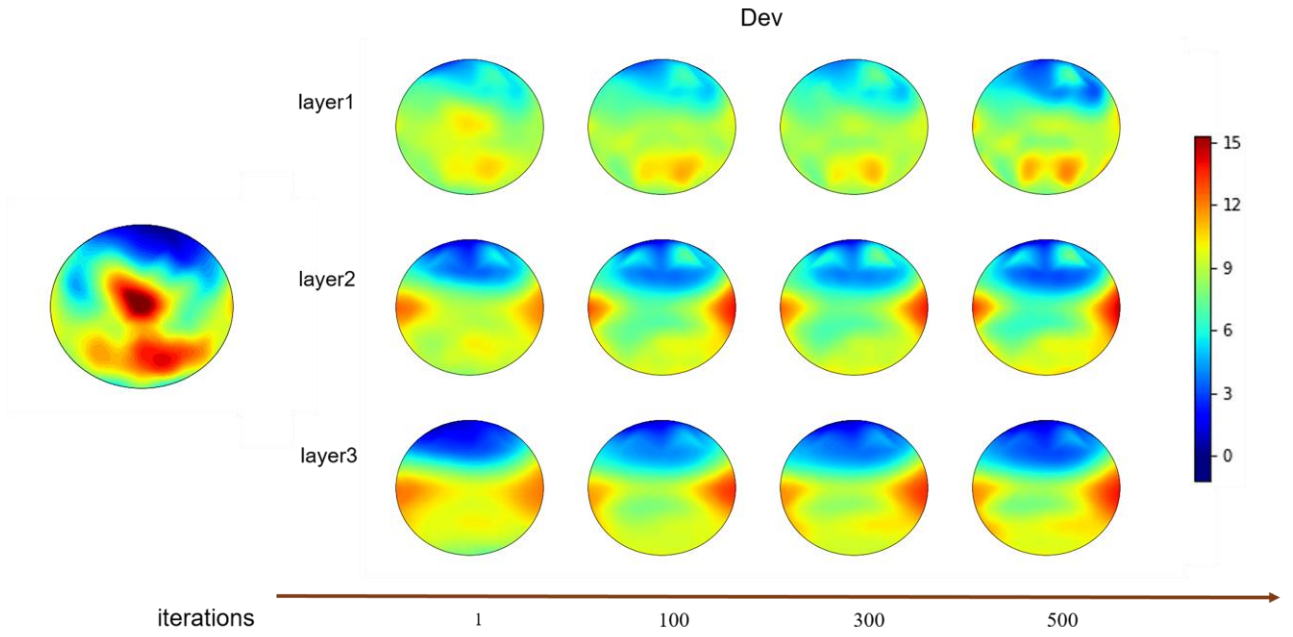
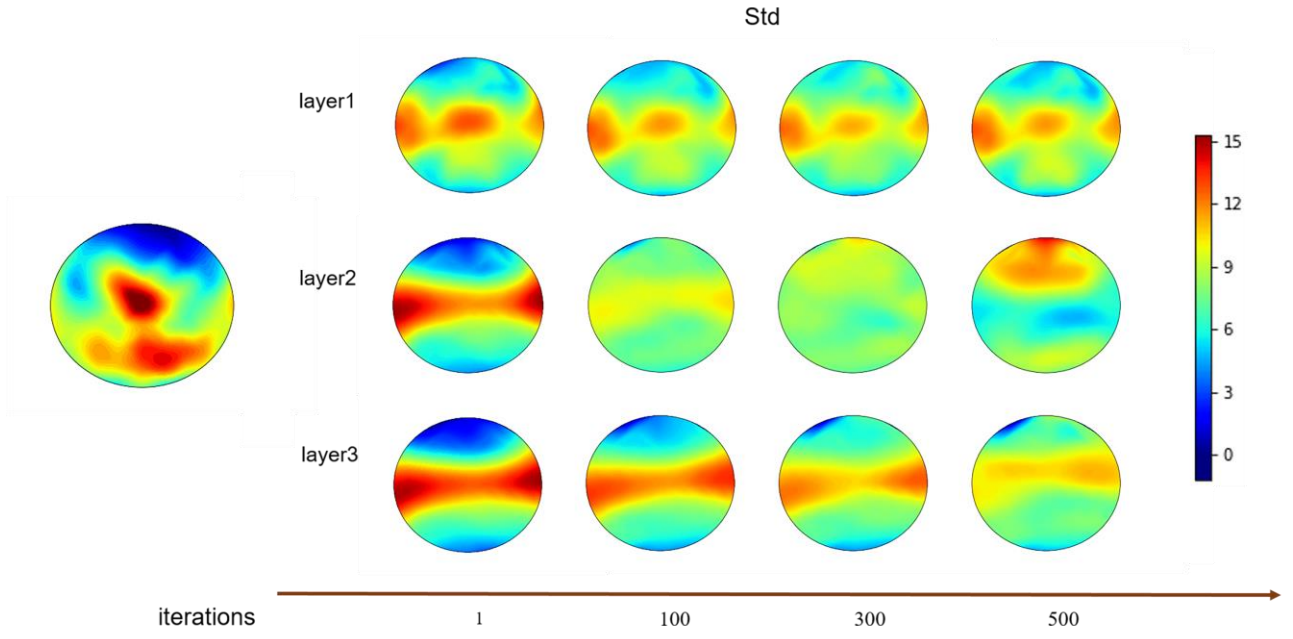


FIGURE 10. Efficiency analysis. Different colours represent different layers. Samples are obtained from 65 trainings (5 rounds, 13 folds each round). The vertical axis represents different metrics, and the horizontal axis represents the number of iterations, and 50 samples per iteration.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access



(a)



(b)

FIGURE 11. Maps of degree centrality. (a) was obtained from classification of deviation stimuli, and (b) was obtained from classification of standard stimuli. Maps from top to bottom are different convolutional layers and from left to right are the beginning to the end of iterations, and each iteration contains 50 samples.

attribute enhancements indicate an increase in the efficiency of information exchange, so we speculate that when classifying standard stimuli, the second convolutional layer attempts to change the strategy, but the benefit is no better than the previous strategy; i.e., the reduction in path length does not compensate for the reduction in clustering coefficients, and therefore, in the third convolutional layer, the network modified this strategy.

2) EFFICIENCY METRICS

Local efficiency is similar to the clustering coefficient and used to measure the local information exchange ability of a network. However, the clustering coefficient considers the direct connection between adjacent nodes, whereas the local efficiency considers the subgraph composed of the adjacent

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

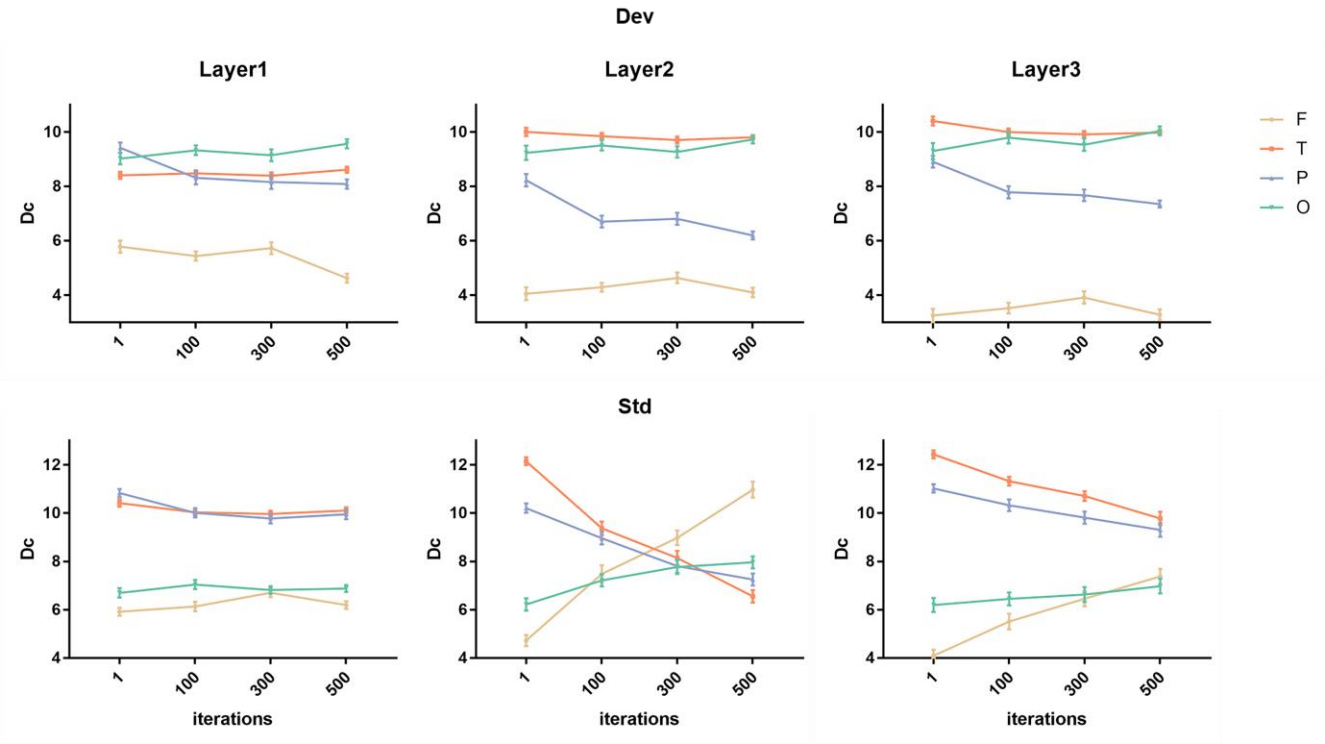


FIGURE 12. Degree centrality analysis of ROI in different layers. Samples are obtained from 65 trainings (5 rounds, 13 folds each round). The vertical axis represents the value of degree centrality, and the horizontal axis represents number of iterations, and 50 samples per iteration.

nodes of the node. Fig. 10 shows the local efficiency and global efficiency change in the connection matrices.

For the deviation stimuli, there is a significant increase between the second convolutional layer and the first convolutional layer. In the second and third convolutional layers, the increase is not as obvious. There is also a significant increase in the starting and the ending of the iterations. For the standard stimuli, there is a significant increase as the convolutional layers deepen, but in the second layer, as the number of iterations increases, there is a significant decrease in local efficiency, and in the third layer, there is no such situation, which is consistent with our previous inferences.

Global efficiency is numerically close to the reciprocal of the path length. Both are widely used to measure the efficiency of the whole network. Networks with high global efficiency are highly integrated, and the path between nodes is always short. Fig. 9 shows the global efficiency of the connection matrices. For the deviation stimuli, the reduction trend can be seen at various thresholds, and this reduction is significant between convolutional layers, except at the first layer, there is an increase at the beginning of the iteration. For the standard stimuli, although there is still a significant reduction between convolutional layers, in the second and third layers, as the number of iterations increases, there is an increase, which is different from the deviation stimuli. This is consistent with the conclusion of the characteristic path length.

3) DEGREE FOR EACH NODE

According to the results obtained by the global metrics of the network, we calculate the nodal metric degree centrality (Dc) for each node and obtain the topology map through the degree value together with the electrode positions. Fig. 11 shows the degree change with the deepening of the convolutional layers and the increase in iterations on each electrode. Degree centrality is the number of edges directly connected to the node. The greater the degree, the greater the number of connected edges, and the more important the node is in the network. From Fig. 11(a) we find that for the deviation stimuli, after the weight distribution in the first convolutional layer, nodes with a large degree value are mainly in the occipital region. After the weight distribution of the second and third convolutional layers, nodes with a large degree value gradually gather in the bilateral temporal lobe. This is consistent with the connection matrices of the convolutional layer output we obtained before. From Fig. 11(b) we find that for the standard stimuli, in the first convolutional layer, the nodes with a larger degree are in the temporal lobe and the central region, and in the second layer, the nodes begin to concentrate on the frontal lobe, and the occipital region also has a slight increase. This is consistent with the connection matrices of the convolutional layer and conforms to our previous inference of the second layer attempting to change the weight distribution strategy. The third layer abandons the strategy in the second layer and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

strengthens the weight between the nodes in the temporal lobe, which results in a degree increase in the temporal lobe.

To verify this change, we performed a statistical analysis of the samples of 65 repeated experiments (13 trainings per round, 5 rounds of training). The results are shown in Fig. 12. We find that for the deviation stimuli, the temporal lobes and occipital regions were significantly higher than the other two regions; for standard stimuli, the degree centrality was concentrated in the temporal and parietal regions. This shows that both the deviation stimuli and the standard stimuli are inextricably linked to the temporal lobe. In the visual stimulation experiment, the visual activation of MMN is reflected as the deviation stimuli have relatively concentrated information processing in the occipital region (more details are included in the supplementary materials).

IV. CONCLUSION AND DISCUSSION

The purpose of this paper is to combine the functional connection matrix with deep learning and to explore the work performed by the convolutional layer in the classification process. Based on the EEG from the visual MMN experiment, this paper constructs functional connection networks through the PLI by regarding the electrodes as nodes and using the matrices as inputs for the convolutional neural network. In addition to the classification result, we also analyse the weight distribution in the outputs of the convolutional layers to explore the rule of feature extraction of convolutional layers. Our classification results are above 95%, which can guarantee the validity of features extracted by the convolutional layers. To further analyse the weight distribution matrices, we use the graph theory method. All our graph analysis results are statistically significant. Finally, we combine the results of graph theory with the position information of the electrodes and present them as topological maps.

From our classification results, we can easily find that the theta band has a greater contribution to the classification of MMN than the alpha and beta bands, which means that the theta band is more valuable in MMN. This result is also consistent with the analysis of spectral power in our previous study [46]. From the input connection matrices (see Fig. 10), we find that there is stronger connection in the theta and alpha bands in the temporal region both for the deviation stimuli and the standard stimuli. We assume the classification accuracy of the beta band being lower than the others may be associated with no significant connection in the temporal lobe. The weight distribution matrices of each layer in the iterative process shows that the weight of the temporal region gradually strengthens with the deepening of convolutional layers and iterations, accompanied by the reinforcement of weight in the occipital region, which eventually forms regions with significantly greater weight in the temporal and occipital regions, except for a special case when classifying standard stimuli, in which the second layer changes the weighting strategy, and the connection weight of the frontal and occipital regions are strengthened. The connection between the occipital region and the frontal lobe is longer than the interior of the temporal lobe and the interior of the occipital region.

We conducted further research through graph theory to study the effect of greater weight given to the connections in the regions and greater weight given to the longer connections, which shows that for the deviation stimuli, the matrices obtained after the weight distribution are greatly enhanced in local information communication, which is manifested by increases in the clustering coefficient and local efficiency, and necessarily leads to the decline in the global information exchange capability by an increase in path length and a decrease in the global efficiency. However, in general, the connection matrices after weight distribution have a stronger information exchange capability and show an increase in small-world attributes. For standard stimuli, in the first convolutional layer, the output matrix is the same as the deviation stimuli; but the second layer tends to change this strategy of strengthening local edges and gives weight to long connections. The global information exchange is enhanced, which is manifested by a decrease in path length, and an increase in the global efficiency, but the local information exchange is also reduced, represented by the reduction in clustering coefficients and local efficiency. The small-world network of the connection matrices is reduced. Because of this, in the third layer, the network no longer strengthens the long connection. Therefore, we believe that in a three-layer convolutional network, the first convolutional layer will find and initially strengthen the connection of important areas, and the second layer may continue the strategy of the first layer or find a new weight-allocation strategy. The third layer network will make the final decision based on the first two layers. However, in general, the convolutional network is still inclined to strengthen the weight of connections in local areas [52-54].

In this paper, from the final output of the connection matrix by convolutional layers, we can easily find that the temporal lobe and the occipital region have a large weight, and when we focus on the nodal degree, the results show that the important nodes are mainly concentrated in the temporal lobe and the occipital area. The temporal lobe region is related to the origin of the MMN mechanism [55, 56], so auditory MMN is the easiest to induce, and it is also the most classic MMN experiment. From this paper, we infer that even visual MMN experiments require the temporal lobe region. Additionally, the difference between the deviation stimuli and the standard stimuli are mainly reflected in the visual area in which the deviation stimuli has more strengthened weight and important nodes in the visual area.

In summary, in this article, we innovatively combined the functional connection matrices with deep learning, explored the convolutional layers in weight distribution, and determined that the occipital region and temporal lobe are pivotal in the MMN experiment.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

However, the current research also has some limitations. Due to the lack of relevant research in the combination of functional connectivity and deep learning, there are some places we deal with relatively simply. For example, to maintain the size of the matrices of the convolutional layer, we simply set the step size to 1, and to simplify the analysis, the obtained multiple convolutional kernels were averaged. However, we believe that applying the deep learning algorithm to the functional connection matrices and analysing the weight distribution of the convolutional layer by means of graph theory may be a new method for exploring EEG and deep learning.

APPENDIX

Appendix A and B are statistical analyses of results in Fig.9 to Fig.10, respectively. The results in tables are p-value between two layers from different iterations. Appendix C is statistical analysis of results of Fig.11. The result in table is p-value in comparison between two regions in different iterations..

ACKNOWLEDGMENT

We gratefully acknowledge all the participants in this experiment for their cooperation. We thank Nan Mu and other researchers for their contribution of data collection and preprocessing. We thank all the members of our lab for their suggestions and support to this project.

REFERENCES

[1] Krizhevsky, A., I. Sutskever, and G.E. Hinton, "*ImageNet Classification with Deep Convolutional Neural Networks.*" Communications of the Acm, vol. 60, no. 6, pp. 84-90, 2017.
[2] Shin, H.C., et al., "*Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning.*" Ieee Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285-1298, 2016.
[3] Kamnitsas, K., et al., "*Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation.*" Medical Image Analysis, vol. 36, no. pp. 61-78, 2017.
[4] Kshirsagar, G.B. and N.D. Londhe, "*Improving Performance of Devanagari Script Input-Based P300 Speller Using Deep Learning.*" IEEE transactions on bio-medical engineering, vol. 66, no. 11, pp. 2992-3005, 2019.
[5] Lawhern, V.J., et al., "*EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces.*" Journal of Neural Engineering, vol. 15, no. 5, pp. 17, 2018.
[6] Zhang, X. and D.R. Wu, "*On the Vulnerability of CNN Classifiers in EEG-Based BCIs.*" Ieee Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 5, pp. 814-825, 2019.
[7] Li, X.W., et al., "*EEG-based mild depression recognition using convolutional neural network.*" Medical & Biological Engineering & Computing, vol. 57, no. 6, pp. 1341-1352, 2019.
[8] Ullah, I., et al., "*An automated system for epilepsy detection using EEG brain signals based on deep learning approach.*" Expert Systems with Applications, vol. 107, no. pp. 61-71, 2018.
[9] Truong, N.D., et al., "*Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram.*" Neural Networks, vol. 105, no. pp. 104-111, 2018.
[10] Khan, H., et al., "*Focal Onset Seizure Prediction Using Convolutional Networks.*" Ieee Transactions on Biomedical Engineering, vol. 65, no. 9, pp. 2109-2118, 2018.
[11] Ansari, A.H., et al., "*Neonatal Seizure Detection Using Deep Convolutional Neural Networks.*" International Journal of Neural Systems, vol. 29, no. 4, pp. 20, 2019.
[12] Kwon, Y.H., S.B. Shin, and S.D. Kim, "*Electroencephalography Based Fusion Two-Dimensional (2D)-Convolution Neural Networks (CNN) Model for Emotion Recognition System.*" Sensors, vol. 18, no. 5, pp. 13, 2018.
[13] Li, J.P., Z.X. Zhang, and H.G. He, "*Hierarchical Convolutional Neural Networks for EEG-Based Emotion Recognition.*" Cognitive Computation, vol. 10, no. 2, pp. 368-380, 2018.
[14] Wei, L.J., et al., *Time-Frequency Convolutional Neural Network for Automatic Sleep Stage Classification Based on Single-Channel EEG*, in *2017 Ieee 29th International Conference on Tools with Artificial Intelligence*. 2017, Ieee: New York. p. 88-95.
[15] Zhang, J.M. and Y. Wu, "*Automatic sleep stage classification of single-channel EEG by using complex-valued convolutional neural network.*" Biomedical Engineering-Biomedizinische Technik, vol. 63, no. 2, pp. 177-190, 2018.
[16] Smit, D.J.A., et al., "*Heritability of "Small-World" Networks in the Brain: A Graph Theoretical Analysis of Resting-State EEG Functional Connectivity.*" Human Brain Mapping, vol. 29, no. 12, pp. 1368-1378, 2008.
[17] Babiloni, F., et al., "*Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function.*" Neuroimage, vol. 24, no. 1, pp. 118-131, 2005.
[18] Stam, C.J., G. Nolte, and A. Daffertshofer, "*Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources.*" Human Brain Mapping, vol. 28, no. 11, pp. 1178-1193, 2007.
[19] Liu, F.Y., G.S. Lin, and C.H. Shen, "*CRF learning with CNN features for image segmentation.*" Pattern Recognition, vol. 48, no. 10, pp. 2983-2992, 2015.
[20] Li, G.B. and Y.Z. Yu, "*Visual Saliency Detection Based on Multiscale Deep CNN Features.*" Ieee Transactions on Image Processing, vol. 25, no. 11, pp. 5012-5024, 2016.
[21] Zhang, R.K., et al., "*Automatic Detection and Classification of Colorectal Polyps by Transferring Low-Level CNN Features From Nonmedical Domain.*" Ieee Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 41-47, 2017.
[22] Li, X.Q., et al., "*Tooth-Marked Tongue Recognition Using Multiple Instance Learning and CNN Features.*" Ieee Transactions on Cybernetics, vol. 49, no. 2, pp. 380-387, 2019.
[23] Schirrmeister, R.T., et al., "*Deep Learning With Convolutional Neural Networks for EEG Decoding and Visualization.*" Human Brain Mapping, vol. 38, no. 11, pp. 5391-5420, 2017.
[24] Gunes, S., et al., "*Sleep spindles recognition system based on time and frequency domain features.*" Expert Systems with Applications, vol. 38, no. 3, pp. 2455-2461, 2011.
[25] Mohammadi, S.M., et al., "*Improving time-frequency domain sleep EEG classification via singular spectrum analysis.*" Journal of Neuroscience Methods, vol. 273, no. pp. 96-106, 2016.
[26] Srinivasan, V., C. Eswaran, and N. Sriraam, "*Artificial neural network based epileptic detection using time-domain and frequency-domain features.*" Journal of medical systems, vol. 29, no. 6, pp. 647-60, 2005.
[27] Vidaurre, C., et al., "*Time Domain Parameters as a feature for EEG-based Brain-Computer Interfaces.*" Neural Networks, vol. 22, no. 9, pp. 1313-1319, 2009.
[28] Galantucci, S., et al., "*Structural Brain Connectome and Cognitive Impairment in Parkinson Disease.*" Radiology, vol. 283, no. 2, pp. 515-525, 2017.

[29] Uhlhaas, P.J., et al., "*Neural synchrony and the development of cortical networks.*" Trends in Cognitive Sciences, vol. 14, no. 2, pp. 72-80, 2010.
[30] Wang, L.Y., et al., "*Beta- Band Functional Connectivity Influences Audiovisual Integration in Older Age: An EEG Study.*" Frontiers in Aging Neuroscience, vol. 9, no. pp. 11, 2017.
[31] Toth, B., et al., "*Large-scale network organization of EEG functional connectivity in newborn infants.*" Human Brain Mapping, vol. 38, no. 8, pp. 4019-4033, 2017.
[32] Yang, Y., Y.H. Qiu, and A.C. Schouten, "*Dynamic Functional Brain Connectivity for Face Perception.*" Frontiers in Human Neuroscience, vol. 9, no. pp. 12, 2015.
[33] Mumtaz, W., et al., "*A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD).*" Medical & Biological Engineering & Computing, vol. 56, no. 2, pp. 233-246, 2018.
[34] Mumtaz, W., et al., "*An EEG-based functional connectivity measure for automatic detection of alcohol use disorder.*" Artificial Intelligence in Medicine, vol. 84, no. pp. 79-89, 2018.
[35] Bassett, D.S. and E.T. Bullmore, "*Small-world brain networks.*" Neuroscientist, vol. 12, no. 6, pp. 512-523, 2006.
[36] Hoekzema, E., et al., "*An Independent Components and Functional Connectivity Analysis of Resting State FMRI Data Points to Neural Network Dysregulation in Adult ADHD.*" Human Brain Mapping, vol. 35, no. 4, pp. 1261-1272, 2014.
[37] Dunne, J.A., R.J. Williams, and N.D. Martinez, "*Network structure and robustness of marine food webs.*" Marine Ecology Progress Series, vol. 273, no. pp. 291-302, 2004.
[38] Iturria-Medina, Y., et al., "*Characterizing brain anatomical connections using diffusion weighted MRI and graph theory.*" Neuroimage, vol. 36, no. 3, pp. 645-660, 2007.
[39] Iturria-Medina, Y., et al., "*Studying the human brain anatomical network via diffusion-weighted MRI and Graph Theory.*" Neuroimage, vol. 40, no. 3, pp. 1064-1076, 2008.
[40] Polania, R., et al., "*Introducing graph theory to track for neuroplastic alterations in the resting human brain: A transcranial direct current stimulation study.*" Neuroimage, vol. 54, no. 3, pp. 2287-2296, 2011.
[41] Vecchio, F., et al., "*Cortical Connectivity Modulation During Sleep Onset: A Study via Graph Theory on EEG Data.*" Human Brain Mapping, vol. 38, no. 11, pp. 5456-5464, 2017.
[42] Hesse, P.N., et al., "*Preattentive Processing of Numerical Visual Information.*" Frontiers in Human Neuroscience, vol. 11, no. pp. 14, 2017.
[43] Javitt, D.C., et al., "*Mismatch negativity as a biomarker of theta band oscillatory dysfunction in schizophrenia.*" Schizophrenia Research, vol. 191, no. pp. 51-60, 2018.
[44] Stothart, G. and N. Kazanina, "*Oscillatory characteristics of the visual mismatch negativity: what evoked potentials aren't telling us.*" Frontiers in Human Neuroscience, vol. 7, no. pp. 9, 2013.
[45] Langrova, J., et al., "*Gender Impact on Electrophysiological Activity of the Brain.*" Physiological Research, vol. 61, no. pp. S119-S127, 2012.
[46] Yan, T.Y., et al., "*Theta Oscillations Related to Orientation Recognition in Unattended Condition: A vMMN Study.*" Frontiers in Behavioral Neuroscience, vol. 11, no. pp. 8, 2017.
[47] Lee, H., G.J.F. Brekelmans, and G. Roks, "*The EEG as a diagnostic tool in distinguishing between dementia with Lewy bodies and Alzheimer's disease.*" Clinical Neurophysiology, vol. 126, no. 9, pp. 1735-1739, 2015.
[48] van der Maaten, L. and G. Hinton, "*Visualizing Data using t-SNE.*" Journal of Machine Learning Research, vol. 9, no. pp. 2579-2605, 2008.
[49] Watts, D.J. and S.H. Strogatz, "*Collective dynamics of 'small-world' networks.*" Nature, vol. 393, no. 6684, pp. 440-442, 1998.
[50] Choi, J.W., et al., "*Fronto-temporal interactions in the theta-band during auditory deviant processing.*" Neuroscience Letters, vol. 548, no. pp. 120-125, 2013.
[51] Hsiao, F.J., et al., "*Theta oscillation during auditory change detection: An MEG study.*" Biological Psychology, vol. 81, no. 1, pp. 58-66, 2009.
[52] Chen, Y.S., et al., "*Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks.*" Ieee Transactions on Geoscience and Remote Sensing, vol. 54, no. 10, pp. 6232-6251, 2016.
[53] Keren, G., B. Schuller, and Ieee, *Convolutional RNN: an Enhanced Model for Extracting Features from Sequential Data*, in *2016 International Joint Conference on Neural Networks*. 2016, Ieee: New York. p. 3412-3419.
[54] Lu, X.J., et al., "*Feature Extraction and Fusion Using Deep Convolutional Neural Networks for Face Detection.*" Mathematical Problems in Engineering, vol. no. pp. 9, 2017.
[55] Garrido, M.I., et al., "*The mismatch negativity: A review of underlying mechanisms.*" Clinical Neurophysiology, vol. 120, no. 3, pp. 453-463, 2009.
[56] Gu, F., et al., "*Lateral Inhibition is a Neural Mechanism Underlying Mismatch Negativity.*" Neuroscience, vol. 385, no. pp. 38-46, 2018.

JINGLONG WU received the B.S. degree in electrical engineering from Jilin Vocational Teachers College, China, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from Kyoto University, Japan, in 1991 and 1994.

From 1994 to 2008, he was a professor at the School of Intelligent Mechanical Systems Engineering at National Xiangshan University, Japan. From 2008 to 2014, he was a professor at the National Academy of Natural Sciences, Okayama University, Japan. Since 2014, he has been a professor at College of Mechatronic Engineering, Beijing Institute of Technology. He hosted the research on bionic tactile sensation of humanoid robot based on interactive tactile brain representation mechanism, supported by National Natural Science Foundation, in 2015 to 2018. He hosted the investigation on the international frontier field of early diagnosis of cognitive disorders, supported by Japan Society for the Promotion of Science (JSPS) Scientific Research Fund Project B, in 2009 to 2012. He is also the author of 16 SCI papers.

Prof. Wu organized the IEEE CME 2009, CME 2010, CME 2011 International Conference on Medical Engineering as the President of the Conference and founded the International Journal of Neuroscience and Biomedical Engineering and is the editor-in-chief of the magazine.

PEIWEN HUANG received the B.S. degree in mechatronic engineering from Beijing Institute of Technology, Beijing, China, in 2017. She is currently pursuing the M.S. degree in mechanical engineering in Beijing Institute of Technology.

She joined the Brain and Neurotechnology Laboratory in Beijing Advanced Innovation Center for Intelligent Robots and Systems in 2017 and has studied in machine learning and EEG signal processing for three years. Now she is working on the functional connection of EEG and the deep learning algorithms.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

Tiantian Liu was born in Songyuan, Jilin, China in 1995. She received the B.S. degree in biomedical engineering from Beijing Institute of Technology, Beijing, China, in 2017. She is currently pursuing the Ph.D. degree in biomedical engineering at Beijing Institute of Technology, Beijing, China. Her research interest includes multimodal data processing for neurodegenerative diseases.

GO RITSU received the B.S. degree in electrical engineering from Jilin Vocational Teachers College, China, in 1984, and the M.S. degree in electronic information from Kyoto Craft Fiber University, Japan, in 1993.

She has worked as senior experimenter at the College of Mechatronic Engineering, Beijing Institute of Technology. She assisted in the creation of EEG Labs, Cognitive Psychology Labs, and EEG, Brain Image Data Analysis Lab, and engaged in cognitive neuroscience, brain-computer interface, brain function imaging.

Duanduan Chen was born in Beijing, China, in 1982. She received the DPhil degree in Biomedical Engineering in 2009, at the University of Oxford, UK, and the B.Sc. degree in Mechanical Engineering in 2005, at Fudan University, China.

She is now a professor at the School of Life Science in Beijing Institute of Technology, China. Her research focuses on biomechanics and computer simulation. The developed techniques could be applied in surgery planning of various vascular and brain diseases.

Tianyi Yan was born in Jilin, Jilin, China in 1981. He received the B.S. degree in electronic information engineering from Changchun University of Science and Technology, Changchun, China, in 2005, the M.S. degree in biomedical engineering from Kagawa University, Kagawa, Japan, in 2008, and the Ph.D. degree in biomedical engineering from Okayama University, Okayama, Japan, in 2011.

From 2011 to 2012, he was a Part-time Researcher in Okayama University. Since 2016, he has been a Professor with School of Life Science, Beijing Institute of Technology. He is the author of more than 35 articles. His research interests include cognitive neuroscience, hardware and software development, machine learning, and neuromodulation. He is an Associate Editor of the journal Frontiers in Neuroinformatics, Journal of Alzheimer's Disease, and holds more than 10 patents.

Professor Yan is the Executive Director of Beijing Neurodegenerative Disorders Society and the Standing Committee Member of Chinese Society of Microcirculation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

APPENDIX

A. STATISTICAL ANALYSIS OF SMALL-WORLD

TABLE I
CLUSTERING COEFFICIENT IN DIFFERENT LAYERS

	I-Layer	J-Layer	Dev			Std		
			MD (I-J)	MSE	P value	MD (I-J)	MSE	P value
Step1	layer1	layer2	-.05809*	.00349	**.000**	-.06612*	.00311	**.000**
		layer3	-.07760*	.00447	**.000**	-.07883*	.00354	**.000**
	layer2	layer1	.05809*	.00349	**.000**	.06612*	.00311	**.000**
		layer3	-.01951*	.00507	**.001**	-.01271*	.00387	**.004**
	layer3	layer1	.07760*	.00447	**.000**	.07883*	.00354	**.000**
		layer2	.01951*	.00507	**.001**	.01271*	.00387	**.004**
Step2	layer1	layer2	-.08106*	.00318	**.000**	-.03275*	.00498	**.000**
		layer3	-.09348*	.00353	**.000**	-.07123*	.00428	**.000**
	layer2	layer1	.08106*	.00318	**.000**	.03275*	.00498	**.000**
		layer3	-.01242*	.00412	**.009**	-.03849*	.00571	**.000**
	layer3	layer1	.09348*	.00353	**.000**	.07123*	.00428	**.000**
		layer2	.01242*	.00412	**.009**	.03849*	.00571	**.000**
Step3	layer1	layer2	-.08365*	.00364	**.000**	-.01957*	.00554	**.002**
		layer3	-.09658*	.00364	**.000**	-.06653*	.00451	**.000**
	layer2	layer1	.08365*	.00364	**.000**	.01957*	.00554	**.002**
		layer3	-.01293*	.00364	**.001**	-.04696*	.00627	**.000**
	layer3	layer1	.09658*	.00364	**.000**	.06653*	.00451	**.000**
		layer2	.01293*	.00364	**.001**	.04696*	.00627	**.000**
Step4	layer1	layer2	-.08222*	.00319	**.000**	-.00534	.00597	**.754**
		layer3	-.09225*	.00319	**.000**	-.05009*	.00559	**.000**
	layer2	layer1	.08222*	.00319	**.000**	.00534	.00597	.754
		layer3	-.01003*	.00319	**.006**	-.04475*	.00756	**.000**
	layer3	layer1	.09225*	.00319	**.000**	.05009*	.00559	**.000**
		layer2	.01003*	.00319	**.006**	.04475*	.00756	**.000**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

TABLE II
CHARACTERISTIC PATH LENGTH IN DIFFERENT LAYERS

	I-Layer	J-Layer	Dev			Std		
			MD (I-J)	MSE	P value	MD (I-J)	MSE	P value
Step1	layer1	layer2	-.18708*	.01693	**.000**	-.18346*	.01506	**.000**
		layer3	-.35453*	.01984	**.000**	-.38387*	.01630	**.000**
	layer2	layer1	.18708*	.01693	**.000**	.18346*	.01506	**.000**
		layer3	-.16745*	.02408	**.000**	-.20041*	.01991	**.000**
	layer3	layer1	.35453*	.01984	**.000**	.38387*	.01630	**.000**
		layer2	.16745*	.02408	**.000**	.20041*	.01991	**.000**
Step2	layer1	layer2	-.23459*	.01475	**.000**	-.10968*	.01601	**.000**
		layer3	-.39879*	.01745	**.000**	-.34843*	.01677	**.000**
	layer2	layer1	.23459*	.01475	**.000**	.10968*	.01601	**.000**
		layer3	-.16420*	.02125	**.000**	-.23876*	.02100	**.000**
	layer3	layer1	.39879*	.01745	**.000**	.34843*	.01677	**.000**
		layer2	.16420*	.02125	**.000**	.23876*	.02100	**.000**
Step3	layer1	layer2	-.22330*	.02048	**.000**	-.07070*	.01490	**.000**
		layer3	-.40524*	.02346	**.000**	-.30269*	.01751	**.000**
	layer2	layer1	.22330*	.02048	**.000**	.07070*	.01490	**.000**
		layer3	-.18194*	.02854	**.000**	-.23199*	.02065	**.000**
	layer3	layer1	.40524*	.02346	**.000**	.30269*	.01751	**.000**
		layer2	.18194*	.02854	**.000**	.23199*	.02065	**.000**
Step4	layer1	layer2	-.21871*	.01319	**.000**	-.05993*	.01657	**.001**
		layer3	-.39811*	.01534	**.000**	-.29307*	.02191	**.000**
	layer2	layer1	.21871*	.01319	**.000**	.05993*	.01657	**.001**
		layer3	-.17941*	.01786	**.000**	-.23314*	.02507	**.000**
	layer3	layer1	.39811*	.01534	**.000**	.29307*	.02191	**.000**
		layer2	.17941*	.01786	**.000**	.23314*	.02507	**.000**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

TABLE III
SIGMA IN DIFFERENT LAYERS

	I-Layer	J-Layer	Dev			Std		
			MD (I-J)	MSE	P value	MD (I-J)	MSE	P value
Step1	layer1	layer2	-.09805*	.01135	**.000**	-.09455*	.00973	**.000**
		layer3	-.10542*	.01429	**.000**	-.08605*	.01125	**.000**
	layer2	layer1	.09805*	.01135	**.000**	.09455*	.00973	**.000**
		layer3	-.00737	.01656	.960	.00850	.01235	.869
	layer3	layer1	.10542*	.01429	**.000**	.08605*	.01125	**.000**
		layer2	.00737	.01656	.960	-.00850	.01235	.869
Step2	layer1	layer2	-.16403*	.01359	**.000**	-.06999*	.01256	**.000**
		layer3	-.15272*	.01525	**.000**	-.09791*	.01317	**.000**
	layer2	layer1	.16403*	.01359	**.000**	.06999*	.01256	**.000**
		layer3	.01131	.01895	.910	-.02792	.01591	.226
	layer3	layer1	.15272*	.01525	**.000**	.09791*	.01317	**.000**
		layer2	-.01131	.01895	.910	.02792	.01591	.226
Step3	layer1	layer2	-.17142*	.01151	**.000**	-.03339	.01551	.098
		layer3	-.16614*	.01278	**.000**	-.09893*	.01523	**.000**
	layer2	layer1	.17142*	.01151	**.000**	.03339	.01551	.098
		layer3	.00527	.01533	.981	-.06554*	.01875	**.002**
	layer3	layer1	.16614*	.01278	**.000**	.09893*	.01523	**.000**
		layer2	-.00527	.01533	.981	.06554*	.01875	**.002**
Step4	layer1	layer2	-.18200*	.01292	**.000**	.03443	.01589	.095
		layer3	-.17536*	.01459	**.000**	-.07898*	.01785	**.000**
	layer2	layer1	.18200*	.01292	**.000**	-.03443	.01589	.095
		layer3	.00664	.01734	.974	-.11341*	.02155	**.000**
	layer3	layer1	.17536*	.01459	**.000**	.07898*	.01785	**.000**
		layer2	-.00664	.01734	.974	.11341*	.02155	**.000**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

B. STATISTICAL ANALYSIS OF EFFICIENCY

TABLE I
LOCAL EFFICIENCY IN DIFFERENT LAYERS

	I-Layer	J-Layer	Dev			Std		
			MD (I-J)	MSE	P value	MD (I-J)	MSE	P value
Step1	layer1	layer2	-.03624*	.00305	**.000**	-.04007*	.00285	**.000**
		layer3	-.04171*	.00412	**.000**	-.03949*	.00339	**.000**
	layer2	layer1	.03624*	.00305	**.000**	.04007*	.00285	**.000**
		layer3	-.00547	.00462	.559	.00058	.00362	.998
	layer3	layer1	.04171*	.00412	**.000**	.03949*	.00339	**.000**
		layer2	.00547	.00462	.559	-.00058	.00362	.998
Step2	layer1	layer2	-.05309*	.00295	**.000**	-.02143*	.00401	**.000**
		layer3	-.05450*	.00342	**.000**	-.03828*	.00381	**.000**
	layer2	layer1	.05309*	.00295	**.000**	.02143*	.00401	**.000**
		layer3	-.00141	.00393	.978	-.01685*	.00455	**.001**
	layer3	layer1	.05450*	.00342	**.000**	.03828*	.00381	**.000**
		layer2	.00141	.00393	.978	.01685*	.00455	**.001**
Step3	layer1	layer2	-.05737*	.00304	**.000**	-.01241*	.00506	**.047**
		layer3	-.05967*	.00320	**.000**	-.03780*	.00406	**.000**
	layer2	layer1	.05737*	.00304	**.000**	.01241*	.00506	**.047**
		layer3	-.00230	.00346	.881	-.02539*	.00544	**.000**
	layer3	layer1	.05967*	.00320	**.000**	.03780*	.00406	**.000**
		layer2	.00230	.00346	.881	.02539*	.00544	**.000**
Step4	layer1	layer2	-.05599*	.00290	**.000**	-.00197	.00574	.981
		layer3	-.05575*	.00334	**.000**	-.02566*	.00491	**.000**
	layer2	layer1	.05599*	.00290	**.000**	.00197	.00574	.981
		layer3	.00025	.00353	1.000	-.02369*	.00676	**.002**
	layer3	layer1	.05575*	.00334	**.000**	.02566*	.00491	**.000**
		layer2	-.00025	.00353	1.000	.02369*	.00676	**.002**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

TABLE II
GLOBAL EFFICIENCY IN DIFFERENT LAYERS

	I-Layer	J-Layer	Dev			Std		
			MD (I-J)	MSE	P value	MD (I-J)	MSE	P value
Step1	layer1	layer2	.02334*	.00182	**.000**	.02318*	.00193	**.000**
		layer3	.04097*	.00200	**.000**	.04326*	.00190	**.000**
	layer2	layer1	-.02334*	.00182	**.000**	-.02318*	.00193	**.000**
		layer3	.01763*	.00238	**.000**	.02008*	.00232	**.000**
	layer3	layer1	-.04097*	.00200	**.000**	-.04326*	.00190	**.000**
		layer2	-.01763*	.00238	**.000**	-.02008*	.00232	**.000**
Step2	layer1	layer2	.02679*	.00157	**.000**	.01186*	.00201	**.000**
		layer3	.04210*	.00157	**.000**	.03826*	.00209	**.000**
	layer2	layer1	-.02679*	.00157	**.000**	-.01186*	.00201	**.000**
		layer3	.01531*	.00157	**.000**	.02640*	.00250	**.000**
	layer3	layer1	-.04210*	.00157	**.000**	-.03826*	.00209	**.000**
		layer2	-.01531*	.00157	**.000**	-.02640*	.00250	**.000**
Step3	layer1	layer2	.02481*	.00173	**.000**	.00641*	.00196	**.004**
		layer3	.04100*	.00199	**.000**	.03299*	.00211	**.000**
	layer2	layer1	-.02481*	.00173	**.000**	-.00641*	.00196	**.004**
		layer3	.01619*	.00210	**.000**	.02659*	.00237	**.000**
	layer3	layer1	-.04100*	.00199	**.000**	-.03299*	.00211	**.000**
		layer2	-.01619*	.00210	**.000**	-.02659*	.00237	**.000**
Step4	layer1	layer2	.02609*	.00154	**.000**	.00289	.00220	.471
		layer3	.04324*	.00154	**.000**	.03204*	.00272	**.000**
	layer2	layer1	-.02609*	.00154	**.000**	-.00289	.00220	.471
		layer3	.01714*	.00154	**.000**	.02915*	.00306	**.000**
	layer3	layer1	-.04324*	.00154	**.000**	-.03204*	.00272	**.000**
		layer2	-.01714*	.00154	**.000**	-.02915*	.00306	**.000**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

C. STATISTICAL ANALYSIS OF DEGREE CENTRALITY

DEGREE CENTRALITY IN DIFFERENT AREAS

			Dev			Std		
	I-Layer	J-Layer	MD (I-J)	MSE	P value	MD (I-J)	MSE	P value
Layer1	F	T	-3.99053*	.19993	**.000**	-3.91495*	.20787	**.000**
		P	-3.46153*	.23976	**.000**	-3.75886*	.25612	**.000**
		O	-4.94409*	.23663	**.000**	-.68473*	.21022	**.009**
	T	F	3.99053*	.19993	**.000**	3.91495*	.20787	**.000**
		P	.52900	.20464	.065	.15609	.24587	.989
		O	-.95356*	.20097	**.000**	3.23021*	.19760	**.000**
	P	F	3.46153*	.23976	**.000**	3.75886*	.25612	**.000**
		T	-.52900	.20464	.065	-.15609	.24587	.989
		O	-1.48256*	.24062	**.000**	3.07412*	.24785	**.000**
	O	F	4.94409*	.23663	**.000**	.68473*	.21022	**.009**
		T	.95356*	.20097	**.000**	-3.23021*	.19760	**.000**
		P	1.48256*	.24062	**.000**	-3.07412*	.24785	**.000**
Layer2	F	T	-5.70987*	.19009	**.000**	4.41909*	.41924	**.000**
		P	-2.09895*	.22637	**.000**	3.72368*	.41311	**.000**
		O	-5.63443*	.22407	**.000**	3.01396*	.41420	**.000**
	T	F	5.70987*	.19009	**.000**	-4.41909*	.41924	**.000**
		P	3.61092*	.17076	**.000**	-.69540	.35522	.276
		O	.07544	.16770	.998	-1.40513*	.35649	**.001**
	P	F	2.09895*	.22637	**.000**	-3.72368*	.41311	**.000**
		T	-3.61092*	.17076	**.000**	.69540	.35522	.276
		O	-3.53548*	.20792	**.000**	-.70972	.34926	.238
	O	F	5.63443*	.22407	**.000**	-3.01396*	.41420	**.000**
		T	-.07544	.16770	.998	1.40513*	.35649	**.001**
		P	3.53548*	.20792	**.000**	.70972	.34926	.238
Layer3	F	T	-6.70912*	.21752	**.000**	-2.39921*	.41817	**.000**
		P	-4.07508*	.23651	**.000**	-1.92510*	.41817	**.000**
		O	-6.76925*	.25743	**.000**	.40453	.41817	1.000
	T	F	6.70912*	.21752	**.000**	2.39921*	.41817	**.000**
		P	2.63404*	.15872	**.000**	.47411	.41817	1.000
		O	-.06013	.18848	1.000	2.80374*	.41817	**.000**
	P	F	4.07508*	.23651	**.000**	1.92510*	.41817	**.000**
		T	-2.63404*	.15872	**.000**	-.47411	.41817	1.000
		O	-2.69417*	.21011	**.000**	2.32963*	.41817	**.000**
	O	F	6.76925*	.25743	**.000**	-.40453	.41817	1.000
		T	.06013	.18848	1.000	-2.80374*	.41817	**.000**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2968717, IEEE Access

	P	2.69417*	.21011	**.000**	-2.32963*	.41817	**.000**