

# A Connectivity-Aware Graph Neural Network for Real-Time Drowsiness Classification

Zhuoli Zhuang<sup>1</sup>, Yu-Kai Wang<sup>2</sup>, *Member, IEEE*, Yu-Cheng Chang<sup>3</sup>,  
Jia Liu, and Chin-Teng Lin<sup>4</sup>, *Fellow, IEEE*

**Abstract**—Drowsy driving is one of the primary causes of driving fatalities. Electroencephalography (EEG), a method for detecting drowsiness directly from brain activity, has been widely used for detecting driver drowsiness in real-time. Recent studies have revealed the great potential of using brain connectivity graphs constructed based on EEG data for drowsy state predictions. However, traditional brain connectivity networks are irrelevant to the downstream prediction tasks. This article proposes a connectivity-aware graph neural network (CAGNN) using a self-attention mechanism that can generate task-relevant connectivity networks via end-to-end training. Our method achieved an accuracy of 72.6% and outperformed other convolutional neural networks (CNNs) and graph generation methods based on a drowsy driving dataset. In addition, we introduced a squeeze-and-excitation (SE) block to capture important features and demonstrated that the SE attention score can reveal the most important feature band. We compared our generated connectivity graphs in the drowsy and alert states and found drowsiness connectivity patterns, including significantly reduced occipital connectivity and interregional connectivity. Additionally, we performed a post hoc interpretability analysis and found that our method could identify drowsiness features such as alpha spindles. Our code is available online at <https://github.com/ALEX95GOGO/CAGNN>.

**Index Terms**—Graph neural network, drowsiness detection, functional connectivity.

## I. INTRODUCTION

**D**ROWSY driving is one of the leading causes of traffic accidents [1]. As drowsiness is the first sign of falling asleep, it is believed that early drowsiness detection could be useful for preventing related traffic accidents. Camera-based methods are widely used to monitor driver drowsiness. However, these methods may lead to privacy concerns and are susceptible to facial occlusion and illumination variations [2]. Electroencephalography (EEG) does not have these issues and can be used to directly measure brain activity. Recently, EEG has been integrated into driver monitoring systems for drowsiness detection. For example, researchers at Mercedes-Benz investigated driver fatigue detection with alpha spindles based on the “Attention assist” system [3]. In drowsiness studies, EEG has been used to indicate the transition between the waking and sleeping states [4]. Compared with other frequency bands, the alpha (8-14 Hz) and theta (4-8 Hz) bands have repeatedly been shown to be more closely associated with drowsiness [5], [6], [7], [8].

For instance, alpha spindles [6] and theta bursts [9], [10] are two prominent features that occur before falling asleep. In addition, increased theta and alpha activity and decreased beta activity have been observed in cases of mental fatigue [7]. In particular, power oscillations in the theta and alpha bands are highly associated with monotonous driving tasks [8].

The drowsy and alert states have been shown to be distinguishable via brain connectivity analysis. Brain functional connectivity is defined as the temporal coincidence in neural activities between segregated regions of the brain [11]. Traditionally, functional connectivity graphs based on EEG were constructed via statistical coupling between two different time series using correlation metrics such as the phase lag index (PLI) [12], amplitude locking value (ALV) [13], and weighted phase lag index (WPLI) [14], and the drowsy and alert states were classified using traditional machine learning methods according to the connectivity information. However, traditional machine learning methods usually require handcrafted hyperparameters. This problem can be mitigated through deep learning methods, which use nonlinear universal approximation for feature extraction purposes.

Convolutional neural networks (CNNs) are popular deep learning models in EEG analysis, including the commonly used EEGNet [15] and ShallowConvNet [16]. Driver drowsi-

Manuscript received 19 February 2023; revised 27 June 2023 and 22 October 2023; accepted 21 November 2023. Date of publication 27 November 2023; date of current version 12 January 2024. This work was supported in part by the Australian Research Council (ARC) under Grant DP210101093 and Grant DP220100803, in part by the University of Technology Sydney (UTS) Human-Centric Artificial Intelligence (AI) Centre funded by GrapheneX (2023–2031), in part by the Australia Defence Innovation Hub under Contract P18-650825, in part by the Australian Cooperative Research Centers Projects (CRC-P) Round under Grant 11 CRCPX1000007, in part by the U.S. Office of Naval Research Global under Cooperative Agreement ONRG—NICOP—N62909-19-1-2058, in part by the Air Force Office of Scientific Research and Defence Science and Technology (AFOSR—DST) Australian Autonomy Initiative under Agreement ID10134, and in part by the New South Wales (NSW) Defence Innovation Network and NSW State Government of Australia under Grant DINPP2019 S1-03/09 and Grant PP21-22.03.02. (Corresponding authors: Zhuoli Zhuang; Chin-Teng Lin.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of the Veterans General Hospital, Taipei, Taiwan, under Approval No. 2012-08-019BCY, and performed in line with the Guide for the Committee of Laboratory Care and Use of the National Chiao Tung University, Taiwan.

The authors are with the CIBCI Laboratory, GrapheneX-UTS HAI Centre, AAIL, School of Computer Science, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: zhuoli.zhuang@student.uts.edu.au; YuKai.Wang@uts.edu.au; Yu-Cheng.Chang@uts.edu.au; jia.liu@uts.edu.au; chin-teng.lin@uts.edu.au).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TNSRE.2023.3336897>, provided by the authors. Digital Object Identifier 10.1109/TNSRE.2023.3336897

ness levels have been successfully detected using CNN in several studies [17], [18], [19]. In some driver drowsiness studies, the reaction time could be forecasted with a CNN [19], [20]. Cui et al. developed a compact single-channel CNN [17] and a multichannel interpretable CNN [21] for drowsiness detection, and these methods demonstrated better performance than EEGNet and ShallowConvNet; however, their method was designed for offline analysis only and real-time implementation using their method decreased the accuracy by approximately 6% [17].

In recent years, researchers have explored graph neural networks (GNNs) for EEG classification [22], [23], [24], [25], [26] to address the limitations of CNNs, since the natural geometries and brain networks in EEG studies are non-Euclidean. The graph in a GNN developed for EEG analysis is usually predefined based on some prior knowledge, such as the distance between electrodes, and the correlation or covariance between EEG channels [22], [24], [25], [26]. However, graphs defined in this manner are not trainable or task-relevant. Moreover, the input feature vectors are fed into the neural network blindly as a black box. In this work, to address these problems, we first applied a self-attention mechanism [27], [28] to generate a task-relevant connectivity graph through end-to-end GNN training. Then, we introduced a squeeze-and-excitation (SE) block [29] to selectively highlight important features, thereby improving the interpretability of our model.

We designed a novel architecture for drowsiness detection as follows. First, we applied self-attention mechanisms to take full advantage of the power of GNN to generate connectivity patterns. Self-attention mechanisms generate importance scores that relate different positions in a single sequence to an encoded sequence through end-to-end training [30]. Due to its ability to capture long-range dependencies, the self-attention mechanism has demonstrated good performance in machine translation [31], [32] and computer vision tasks [33]. Inspired by the attention mechanism [27], [28], we designed a method to learn an adjacency matrix based on a weighted feature vector, which can capture the interdependencies between EEG channels through end-to-end training.

Furthermore, we introduced an SE block to make the network focus on more informative features in end-to-end training. Originally, the SE block was designed for channel-wise attention for image classification tasks, and it has been shown to improve EEG classification accuracy in some studies [34], [35], [36]. Instead of exploring the interdependencies between EEG channels, we applied the SE block to extract the interdependencies between EEG features, which has been overlooked in many previous studies. The results show that the SE block not only improved the classification accuracy but also enhanced the interpretability of our model by highlighting the most important frequency features.

The main contributions of our work can be summarized as follows:

- 1) We proposed a novel method to generate connectivity patterns in GNNs that outperforms other state-of-the-art

CNNs and GNNs in drowsiness classification in both balanced and imbalanced datasets.

- 2) We showed that the SE attention score of our model indicated the most important frequency features; moreover, when our model was retrained using the frequency band with the highest attention score, the accuracy improved.
- 3) We demonstrated that our learned connectivity patterns could provide some insights into neuroscience and that our generated connectivity patterns showed weaker parietal and occipital connectivity and decreased intraregional connectivity, indicating local segregation effects in drowsy states.

## II. RELATED WORK

### A. Popular Architectures for EEG Classification

Convolutional neural networks (CNNs) are the most popular architectures for EEG classification thanks to their strong feature extraction abilities. The ShallowConvNet [16] architecture consists of two convolutional layers (a (1, 13) temporal filter and a ( $C$ , 1) spatial filter, with  $C$  representing the number of channels) and has shown good performance based on motor imaginary datasets. EEGNet [15] is an improved version of ShallowConvNet that introduces a depthwise separable convolutional layer, which reduces the number of parameters and prevents overfitting, and decouples the relationship within and across each feature map. Cui et al. [17], [21] developed InterpretableCNN and further improved EEGNet by introducing a global average pooling (GAP) layer to replace the widely used fully connected layer, which reduced the number of parameters and prevented overfitting. They demonstrated that InterpretableCNN outperforms EEGNet and ShallowConvNet based on a drowsy driving dataset [37]. One of the limitations of CNNs is that they are designed for regular and Euclidean structured data [38], [39], and this is not sufficiently precise since the natural geometry and connectivity of brain networks in EEG data are non-Euclidean [22].

### B. Graph Neural Networks for EEG Classification

In graph neural networks (GNNs), the convolutional layers in CNNs are replaced with graph convolutional layers. GNNs for EEG classification model the brain as a graph, with the electrodes represented by nodes and the connectivity between them represented by edges. One key challenge of GNNs is selecting appropriate graph representations for EEG signals. This is usually performed by exploring some prior knowledge in the EEG data. For instance, EEG-GCNN [24] defined the graph based on the electrode distance and spectral coherence between different EEG channels. Similarly, Tang et al. [22] constructed a graph based on the electrode distance and correlations between spectral features. In contrast, Jingcong et al. showed that a learnable graph generation method called the self-organized graph neural network (SOGNN) [23] could outperform the traditional correlation and covariance graph representation. However, the graph learned by the SOGNN was highly asymmetric, which might not be the case for healthy individuals [40]. We note two key insights based on these architectures. First, we aim to generate the GNN

graph representation through end-to-end training to learn a task-relevant graph for downstream tasks. Second, we aim to generate a biologically meaningful graph that can offer valuable insights into neuroscience.

### III. MATERIALS AND METHODS

#### A. Dataset Description

The dataset in this article was collected at the Brain Research Center (BRC) at National Chiao Tung University (NCTU), Hsinchu, Taiwan. This study was performed in strict accordance with the recommendations in the Guide for the Committee of Laboratory Care and Use of the National Chiao Tung University, Taiwan, with ethics approval (2012-08-019BCY) from the Institutional Review Board of the Veterans General Hospital, Taipei, Taiwan. This dataset contains 32-channel EEG recordings (including one ground electrode and one reference electrode), collected from 27 healthy subjects (age range: 22 to 28 years) in a driving simulator over 62 sessions, with each recording occurring over 90 minutes. The impedances of the EEG electrodes were maintained at less than 5 k $\Omega$ , and the sampling rate was 500 Hz. The experiment was conducted in a 360-degree virtual reality (VR) lab [41]. The scenario was a monotonous VR highway scene aiming to induce drowsiness [42]. During the task, lane departure events were induced randomly to make the car drift towards the left or right sides. The subjects were instructed to respond as quickly as possible to move the car back to the centre of the lane. Between two consecutive trials, there was one resting period that lasted for a random amount of time ranging from 5 – 10s to prevent participants from anticipating a potential deviation. The reaction time (RT) was measured to evaluate the driver’s performance. The RT was defined as the latency between the onset of the deviation and the onset of the response. A high RT indicates that the participant is relatively drowsy, while a short RT indicates that the participant is relatively alert. A detailed description of this dataset can be found in [37]. The dataset is also publicly available from figshare [43].

#### B. Data Processing

The dataset was preprocessed with the following steps. First, the raw data were filtered with a finite impulse response (FIR) filter with frequency cut-offs of 1 Hz and 50 Hz. Second, artefacts were automatically removed via the Automatic Artefact Removal (AAR) plug-in for EEGLAB. The AAR method is suitable for real-time applications [44], [45]. Third, all trials with reaction times of less than 100 ms were removed. Because human reaction times are unlikely to be less than 100 ms [46], in these cases, the subjects might have turned the steering wheel by chance instead of actually responding to the stimuli. We then downsampled the EEG data to 128 Hz and extracted the 5 seconds of samples prior to the stimuli presentation. We included only sessions in which the motion function of the simulator was disabled to minimize motion artefacts. The drowsy samples and alert samples were labelled according to the method proposed in [47]:

TABLE I

SUMMARY OF THE SELECTED NUMBER OF SAMPLES FOR EACH SUBJECT IN THE BALANCED AND UNBALANCED DATASET

ID	Drowsy		Alert	
	Balanced	Unbalanced	Balanced	Unbalanced
1	91	96	91	91
2	65	65	65	353
3	62	179	62	62
4	74	74	74	112
5	112	112	112	141
6	41	41	41	157
7	44	101	44	44
8	132	132	132	222
9	154	154	154	238
10	53	53	53	181
11	112	128	112	112
<b>Total</b>	<b>940</b>	<b>1135</b>	<b>940</b>	<b>1713</b>

- The global reaction time (RT) was calculated by averaging all the local RTs within a 90-second window before the onset of deviation [48]. The alert RTs were calculated as the 5<sub>th</sub> quantile of all the local RTs.
- The 5-second samples were labelled “Alert” if both the local and global RTs were shorter than 1.5 times the alert RT and were labelled “Drowsy” if both the local and global RTs were greater than 2.5 times the alert RT.

The eligible trials were selected by the following procedures:

- 1) Sessions with fewer than 40 samples of either class were discarded.
- 2) The samples from each session were balanced by choosing the most representative samples with the shortest (for the alert class) or longest (for the drowsy class) local RT [17].
- 3) For multiple sessions with the same subject, we selected the session with the most samples.

The first step was used to exclude sessions with too few samples in one or both classes for neural network training. The second step ensured a balanced dataset for both classes. For the imbalanced dataset, this step was not used. The third step was used to balance the number of subjects and ensure that the sessions with the most samples were retained. As summarized in Table I, we obtained 1880 samples from 11 subjects for the balanced dataset and 2848 samples for the naturally imbalanced dataset. We obtained a balanced dataset with a size of [1880 samples  $\times$  30 channels  $\times$  640 timepoints], and an imbalanced dataset with a size of [2848 samples  $\times$  30 channels  $\times$  640 timepoints].

#### C. Network Architecture

The architecture of the proposed method is shown in Figure 1. The graph in the GNN can be formulated as a set of nodes, edges and weights of edges  $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \mathcal{A})$ . The node  $\mathcal{N}$  corresponds to the EEG data of each channel, and the weights of the edges denoted by the adjacency matrix  $\mathcal{A}$  in the GNN represent the connectivity between the EEG channels.

Each node in the GNN was constructed by the following steps. First, the 5-second EEG sample was divided into ten 0.5-second EEG segments. Then, to model the EEG spectral features, we used a fast Fourier transform (FFT) to transform

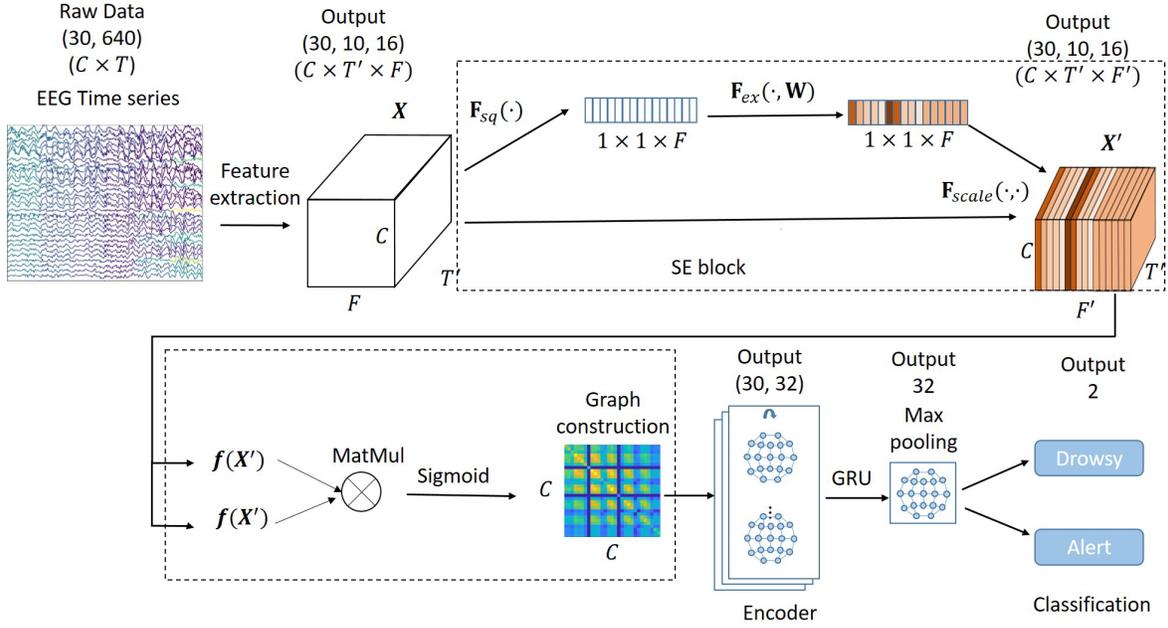


Fig. 1. Architecture of the proposed connectivity-aware graph neural network. First, the frequency features were extracted from the EEG time series through FFTs. Then an SE block was used to highlight important frequency features, and a self-attention mechanism was used to construct the graph from the scaled features. Then, the scaled features and graph were fed into a GNN encoder and decoded via a GRU. Finally, a max-pooling layer and a fully connected layer were applied for classification.

the time series segments into frequency spectra and retained only the positive frequencies in the frequency spectrum to remove redundancy. We included frequency bins only in delta, theta, alpha, and beta bands since these bands are more closely related to drowsiness features. With this approach, we obtained ten frequency spectra with 16 frequency bins:  $[X_1, X_2, \dots, X_{16}]$ . Next, we applied an SE block to obtain a set of modulation weights for the input feature vectors. In the SE block, first, a squeeze operation was applied to aggregate the feature vectors [29]:

$$z_f = \mathbf{F}_{sq}(\mathbf{X}_f) = \frac{1}{C \times T'} \sum_{i=1}^C \sum_{j=1}^{T'} \mathbf{X}_{i,j,f} \quad (1)$$

where  $C$  and  $T'$  are the first two dimensions of the feature vector  $X \in \mathbb{R}^{C \times T' \times F}$ ,  $C$  is the channel dimension,  $T'$  is the time dimension, and  $F$  is the frequency dimension. The squeeze operation is a global pooling of the feature vector in the channel and time dimensions. Then, an excitation block was applied to learn the sample-specific activation for each feature [29]:

$$s = \mathbf{F}_{ex}(z, \mathbf{W}) = (\sigma(\mathbf{W}^2 \delta(\mathbf{W}^1 z))) \quad (2)$$

where  $\delta(\cdot)$  denotes the ReLU [49] function,  $\sigma(\cdot)$  denotes the sigmoid function, and  $\mathbf{W}^1$  and  $\mathbf{W}^2$  represent two fully connected layers, where  $\mathbf{W}^1 \in \mathbb{R}^{\frac{F}{r} \times F}$  and  $\mathbf{W}^2 \in \mathbb{R}^{F \times \frac{F}{r}}$ . In this article, we set the squeeze ratio  $r$  to 8. Next, we designed a method to learn the adjacency matrix based on the weighted feature with a self-attention mechanism. The construction of the edges can be expressed as:

$$\begin{aligned} A &= \sigma(f(s \odot \mathbf{X}) f(s \odot \mathbf{X})^T) \\ &= \sigma(\tanh(\mathbf{W}^3 (s \odot \mathbf{X})) (\tanh(\mathbf{W}^3 (s \odot \mathbf{X})))^T) \end{aligned} \quad (3)$$

where  $\odot$  denotes elementwise multiplication,  $\mathbf{X}$  is the normalized input feature vector of the EEG channels, a  $\tanh$  function was applied to map the input to  $(0, 1)$  and  $\mathbf{W}^3$  denotes a fully connected layer. The key and query are obtained from the same feature vector, and  $\sigma$  denotes the sigmoid function, which is applied to map the attention weight to  $(0, 1)$  while ensuring a symmetric self-attention score. The self-attention score was used as the adjacency matrix of the GNN.

The key and query are the same for the following reasons. We generate the graph using the self-attention mechanism to mimic the functional connectivity. Functional connectivity represents the temporal coincidence of two spatial separate neurophysiological events, which are nondirectional and symmetric [50]. To generate  $A = K Q^T$  as a symmetric functional connectivity matrix, the key  $K$  and the query  $Q$  needed to be the same.

Then, the normalized Laplacian of the GNN was computed by:

$$\mathbf{L} = \mathbf{I} - \hat{\mathbf{D}}^{-1/2} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-1/2} \quad (4)$$

where  $\mathbf{D}$  is a diagonal degree matrix of the normalized adjacency matrix with  $\mathbf{D}_{ii} = \sum_j \mathcal{A}_{ij}$  and  $\mathbf{I}$  is the identity matrix. To capture the dynamic characteristics of functional networks, in this work, we use a variant of the diffusion convolutional recurrent neural network (DCRNN) [51] for brain dynamics modelling, which can model temporal and spectral dynamics at the same time.

The DCRNN has demonstrated impressive performance in tasks such as traffic forecasting, with the model learning complex spatiotemporal patterns and obtaining accurate predictions [51]. The DCRNN has also shown promise in brain dynamics modelling, enabling researchers to understand the complex interactions and dynamics of neural processes in the

brain [22]. In the DCRNN, the encoder encodes the spatiotemporal dependencies of the EEG data as a feature representation using diffusion convolutional gated recurrent units (DCGRUs). Specifically, the encoder consists of two recurrent layers with 32 recurrent units in each layer, and the information is encoded into a set of 30-channel node features ( $\mathbb{R}^{30 \times 32}$ ). The DCGRU was used instead of matrix multiplication operations in the gated recurrent unit (GRU) [52], with the diffusion convolution defined as follows:

$$\mathbf{r}^{(t)} = \sigma(\Theta_{r \star \mathcal{G}}[\mathbf{X}(t), \mathbf{H}^{(t-1)}] + \mathbf{b}_r), \quad (5)$$

$$\mathbf{u}^{(t)} = \sigma(\Theta_{u \star \mathcal{G}}[\mathbf{X}(t), \mathbf{H}^{(t-1)}] + \mathbf{b}_u), \quad (6)$$

$$\mathbf{C}^{(t)} = \tanh(\Theta_{C \star \mathcal{G}}[X^{(t)}, (r(t) \odot \mathbf{H}^{(t-1)})] + \mathbf{b}_c), \quad (7)$$

$$\mathbf{H}^{(t)} = \mathbf{u}^{(t)} \odot \mathbf{H}^{(t-1)} + (1 - \mathbf{u}^{(t)}) \odot \mathbf{C}^{(t)} \quad (8)$$

where  $\mathbf{X}^{(t)}$ ,  $\mathbf{H}^{(t)}$  are the input and output at time  $t$  of the recurrent layer and  $\mathbf{r}^{(t)}$  and  $\mathbf{u}^{(t)}$  are the reset gate and update gate at time  $t$ ,  $\odot$  represents the Hadamard product, and  $\star \mathcal{G}$  denotes the ChebNet spectral convolution.

The DCRNN propagates messages between nodes through diffusion convolution. In the case of an undirected graph, diffusion convolutions were performed using ChebNet spectral graph convolution [53]:

$$\mathbf{X}_{:,m \star \mathcal{G}} f_{\theta} = \sum_{k=0}^{k-1} \theta_k \mathbf{L}^k \mathbf{X}_{:,m} = \sum_{k=0}^{k-1} \hat{\theta}_k T_k(\hat{\mathbf{L}}) \mathbf{X}_{:,m} \quad (9)$$

where  $\theta_k$  represents the learnable weights for the  $k_{th}$  order Chebyshev polynomial approximation, and  $L$  is the renormalized Laplacian matrix of the graph,  $T_0(x) = 1$ ,  $T_1(x) = x$ ,  $T_k(x) = xT_{k-1}(x) - T_{k-2}(x)$  are the bases of the Chebyshev polynomial. The diffusion convolutional layer in the encoder works by propagating node features across the graph structure. The diffusion operation captures the local and global dependencies of the nodes in the graph, allowing the network to learn representations that capture the complex interactions and correlations between the different EEG nodes.

Finally, the node features are aggregated through max-pooling along the node dimension, followed by two fully connected layers ( $32 \times 16$ ,  $16 \times 2$ ) that output the probabilities of the two classes. The binary cross-entropy loss function was used for network training:

$$l_n = -w_n [y_n \cdot \log \sigma(x_n) + (1 - y_n) \cdot \log (1 - \sigma(x_n))] \quad (10)$$

where  $x_n$  denotes the  $n_{th}$  sample of the network output,  $y_n$  denotes the  $n_{th}$  sample of the label and  $\sigma(\cdot)$  is the *sigmoid* activation function. The optimization objective is to minimize the binary cross-entropy loss between the network output and the labels.

#### D. Implementation Details

All the code was implemented with Python 3.8.0. All the models were implemented with PyTorch 1.9.0. The models were trained on a computer with NVIDIA Quadro P5000 graphic cards. The Adam optimizer with an initial learning rate of  $1e^{-3}$  and a cosine annealing weight decay were used to train the proposed CAGNN. All the models were trained for 30 epochs.

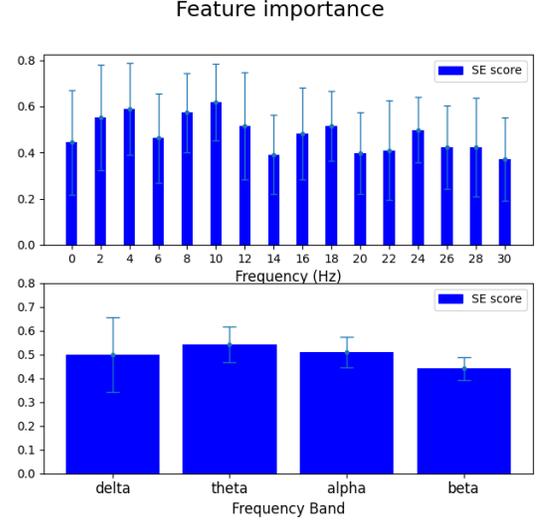


Fig. 2. Attention score of the SE block over different frequency features (upper panel) and frequency bands including delta, theta, alpha, and beta bands (lower panel). (Error bar: standard deviation.)

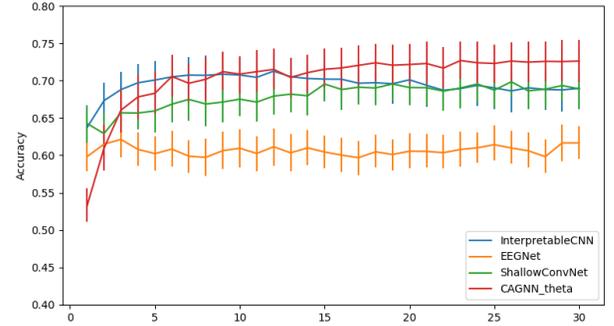


Fig. 3. Training curve of the CAGNN, InterpretableCNN, EEGNet, and ShallowConvNet. (Error bar: standard deviation.)

## IV. RESULTS

### A. Performance Analysis

We used leave-one-out cross-validation since the number of subjects was small. During training, one subject was left out for evaluation, and the data from other subjects were used for training. All experiments were repeated five times with five different random seeds. All the models were evaluated at the 30<sup>th</sup> epoch, and all models converged at approximately 15 epochs, as shown in Figure 3. We compared the performance of our CAGNN model with that of state-of-the-art CNN models, including EEGNet and ShallowConvNet, and a recently developed model called InterpretableCNN [21]. InterpretableCNN was originally developed only for offline analysis, and the sample must be sent to the model as a bundle during evaluation. This is not desirable for real-time applications since new samples are usually obtained individually instead of as a bundle. In this article, all the methods were implemented in such a way that real-time data could be used. Note that InterpretableCNN was reimplemented to handle real-time data by turning off the tracking of the running mean and variance in evaluation mode. The results are summarized in Table II, and the training curve is shown in Figure 3. In addition to accuracy, we used the F1 score,

TABLE II  
PERFORMANCE COMPARISON OF THE CAGNN AND THREE BASELINE METHODS ON THE BALANCED DATASET

ID	EEGNet		ShallowConvNet		InterpretableCNN		CAGNN(w/o SE)		CAGNN	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
1	63.6	46.5	69.2	70.2	77.6	76.1	81.1	80.4	78.5	77.4
2	63.6	46.5	56.8	56.6	48.0	45.8	86.3	86.1	84.9	84.6
3	51.1	6.3	71.8	70.7	71.3	60.1	62.8	60.5	63.7	61.4
4	64.9	54.3	62.4	72.2	66.1	73.7	53.1	49.1	58.3	55.0
5	50.7	4.8	74.3	67.7	66.9	52.6	60.4	53.6	60.7	54.4
6	63.2	46.0	68.3	71.0	77.3	78.0	68.5	68.5	72.7	72.7
7	70.2	62.2	76.4	79.1	62.0	72.1	66.8	62.8	67.3	63.6
8	54.4	25.3	66.8	72.9	72.6	76.7	61.9	57.5	61.4	56.1
9	69.9	56.7	82.1	82.0	87.4	86.3	84.4	84.3	87.1	87.0
10	68.1	53.1	75.1	66.9	71.9	61.4	74.3	72.5	74.9	73.2
11	58.1	39.9	57.9	67.4	59.6	66.2	66.8	66.7	65.9	65.8
All	61.6	40.7	69.2	70.6	69.2	68.1	69.7	67.5	70.5	68.3

TABLE III  
PERFORMANCE OF THE CAGNN IN DIFFERENT FREQUENCY BANDS BASED ON THE BALANCED DATASET

Frequency Band	EEGNet				ShallowConvNet				InterpretableCNN				CAGNN			
	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec
Delta	56.7	47.2	61.7	57.0	64.0	52.7	66.0	64.0	65.4	62.7	70.6	65.4	67.4	65.8	70.0	59.9
Theta	65.4	60.6	70.5	65.4	64.1	61.4	69.1	64.1	69.0	67.2	73.5	69.0	72.6	70.7	77.0	71.2
Alpha	64.9	52.6	69.2	64.9	66.1	64.1	70.0	66.1	65.2	62.8	70.0	65.2	67.1	65.0	75.6	61.3
Beta	58.6	53.8	63.8	58.7	57.1	52.5	62.1	57.1	62.3	58.6	68.8	62.3	65.3	63.2	70.4	58.1

precision, and recall, which are commonly used evaluation metrics for binary classification. The accuracy is defined as the proportion of correct predictions made by the model; the precision measures how many positively labelled instances predicted by the classifier were actually true positives; the recall, also known as the sensitivity, measures how many actual positive instances were correctly identified by the classifier; and the F1 score is the balanced mean of the precision and recall. The accuracy, precision, recall, and F1 score are formulated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (14)$$

$$= \frac{2 \times TP}{2 \times TP + FP + FN}$$

where TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives.

With the balanced dataset, our model (with the SE block) outperformed the other three models in terms of overall accuracy, with a mean accuracy of 70.5. We noticed that without the SE block, the accuracy of our model dropped by approximately 1% to 69.7%. In terms of individual subjects, our model outperformed the other models in Subjects 1, 2 and 11 in accuracy and Subjects 1, 2, and 11 in F1 score. The accuracy of our model was all above 58%, while all the other models had some participants with accuracies below 58%.

We show the attention scores learned by the SE block in Figure 3. Next, we calculated the average attention scores for four feature bands: delta, theta, alpha, and beta. As shown in Figure 3, the theta band had the highest attention score, followed by the alpha, delta, and beta bands. We then retrained our CAGNN model with different frequency features. Feature clipping with the theta band yielded the highest accuracy, 72.6%, and F1 score, 70.7%, which correspond to the frequency band with the highest SE attention score. Feature clipping with the delta band and the alpha band resulted in an accuracy of approximately 67%. Feature clipping in the beta band resulted in the lowest accuracy, and the beta features also had the lowest SE attention score. This might be because the theta band features are unique to drowsiness, and the alpha band features coexist in drowsy and alert samples, as illustrated in Figure 6. We also evaluated the performance of EEGNet, ShallowConvNet, and InterpretableCNN with a 5<sup>th</sup> order Butterworth filter applied in the delta, theta, alpha, and beta bands. However, the performance of these comparison methods did not improve after applying the filtering process. This may have been due to their original design, which incorporated a temporal convolutional layer with a size of (1, T) meant to handle a wide range of EEG frequency bands. Specifically, ShallowConvNet was inspired by the filter bank common spatial patterns (FBCSP) pipeline [54], which involves separating an EEG signal into different frequency bands for feature extraction. Similarly, in ShallowConvNet [16], the first layer performed a temporal convolution operation that mimicked the bandpass CSP filter in FBCSP. This design has been adopted by many CNN architectures for EEG classification, including EEGNet and InterpretableCNN. In this case, bandpass filtering might have disrupted their ability to capture spatial patterns across the

TABLE IV  
PERFORMANCE COMPARISON OF THE CAGNN AND THREE BASELINE METHODS BASED ON THE UNBALANCED DATASET

ID	EEGNet				ShallowConvNet				InterpretableCNN				CAGNN			
	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec
1	58.0	36.0	24.2	78.5	75.5	72.1	62.1	86.4	73.3	70.1	72.1	72.1	78.7	81.7	70.4	97.4
2	79.4	18.0	17.0	27.8	34.8	21.0	51.4	13.4	55.4	17.6	21.0	21.0	61.9	70.7	98.7	55.6
3	25.9	0.7	0.3	30.0	50.7	50.6	37.0	92.1	37.8	31.0	50.6	50.6	39.9	46.9	36.6	91.0
4	69.2	45.4	38.6	67.2	63.1	65.3	87.0	52.4	61.6	65.1	65.3	65.3	60.7	56.3	84.6	42.7
5	56.0	3.1	1.6	50.0	70.7	55.5	44.6	83.3	65.1	38.8	55.5	55.5	59.0	72.4	57.9	96.5
6	83.1	35.2	24.4	85.3	64.9	44.0	67.3	32.8	78.0	56.5	44.0	44.0	66.4	75.8	87.6	67.9
7	53.5	50.0	34.1	97.8	78.5	82.5	73.9	94.1	80.1	85.5	82.5	82.5	77.5	65.7	62.3	70.0
8	65.6	26.3	18.2	59.3	66.8	61.5	72.9	55.1	71.2	60.4	61.5	61.5	60.3	60.9	79.7	49.3
9	71.3	34.4	22.7	79.7	84.3	66.1	60.3	74.0	86.0	64.9	66.1	66.1	84.9	86.9	91.1	83.3
10	82.9	40.1	27.2	92.9	86.2	62.9	53.6	79.9	84.5	53.1	62.9	62.9	79.3	87.3	83.2	91.9
11	48.1	7.1	3.8	87.1	69.8	72.1	73.1	72.7	63.0	58.1	72.1	72.1	67.8	73.1	60.0	93.6
All	63.0	26.9	19.3	68.7	67.8	59.4	62.1	66.9	<b>68.7</b>	54.6	59.4	59.4	66.9	<b>70.7</b>	<b>73.8</b>	<b>76.3</b>

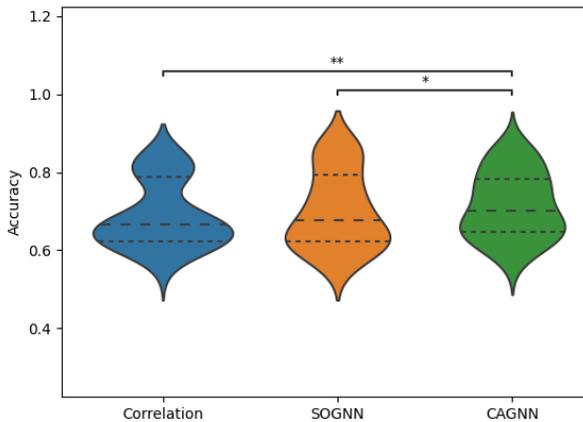


Fig. 4. Comparison of the accuracies of the CAGNN model and other graph generation methods. (\*,  $p < 0.05$ . \*\*,  $p < 0.01$ . Middle dashed line in the violin plot: median. Lower dash line: the 1<sub>st</sub> quantile. Upper dashed line: 3<sub>rd</sub> quantile).

broader frequency spectra, as these models were designed to work effectively without prior bandpass filtering.

We also compared the performance of the CAGNN model with other models based on the naturally unbalanced dataset, as shown in Table IV. Our method with feature clipping in the theta band exhibited a slightly lower accuracy than InterpretableCNN and ShallowConvNet. Nevertheless, our method outperformed the other comparison methods in terms of various metrics for unbalanced datasets, such as the F1 score, precision, and recall, outperforming the other models by more than 5%. These results demonstrate the superior ability of our method in identifying dangerous positive drowsy samples compared to other approaches.

We then compared the accuracy of the CAGNN with the accuracies of other graph generation methods, as shown in Figure 4. We selected a static graph generation method using the correlations between EEG segments proposed by [22] and a learnable graph generation method (SOGNN) [23]. The CAGNN had the highest average accuracy among the three methods. Paired T-tests with Bonferroni correction were used for statistical analysis. The accuracy of the CAGNN model was significantly higher than those of the SOGNN ( $p = 0.029$ ) and the correlation graph generation method ( $p = 0.004$ ), while there was no significant difference between the accuracies of the SOGNN and the correlation graph methods.

The SOGNN used a different weight matrix for the key and query, which was the same as that of the transformer, to construct the attention matrix. This alternation led to a suboptimal result and can be attributed to the fact that the transformer architecture was originally designed for large-scale datasets such as those in language and image processing domains. In such contexts, employing distinct key and query matrices can introduce additional parameters and facilitate better fitting. However, when working with smaller datasets such as EEG data, this approach may lead to overfitting.

### B. Interpretability Analysis

We next investigated the learned connectivity graphs generated by the CAGNN with features in the theta band in the drowsy and alert conditions as shown Figure 5.

To compare the difference between drowsy and alert connectivity, we performed a paired t-test with Bonferroni correction and subtracted the alert connectivity from drowsy connectivity. All the non-significant connectivity ( $p \geq 0.05$ ) was set to 0. We found that there were some differences between these two states in terms of connectivity. First, there was lower connectivity in the occipital region in the drowsy state. This finding was consistent with the findings in [55], [56], [57], and [58], which might be an indication of fading of consciousness [59]. Second, there was higher connectivity in the parietal region in the alert state. This was consistent with the findings in [13] and [60], where significantly weaker connectivity was found within the parietal and occipital regions in drowsy states. In addition, there was lower interregional (e.g., frontal-occipital) connectivity in the drowsy state, which might be due to the local segregation effect and is consistent with the findings in [12], this might indicate that brain connectivity was transforming to a more efficient architecture to maintain brain function in the drowsy states.

We conducted post hoc interpretability analysis via an occlusion-based approach [61]. During each iteration, we set one segment of EEG data (half a second per channel) to zero and calculate the relative change in the model output. This resulted in an occlusion map  $O \in \mathbb{R}^{T \times C}$ , with  $T = 10$  corresponding to the 10 segments of EEG data and  $C = 30$  corresponding to the number of channels. The occlusion map was normalized to the range 0 to 1. The strength of the

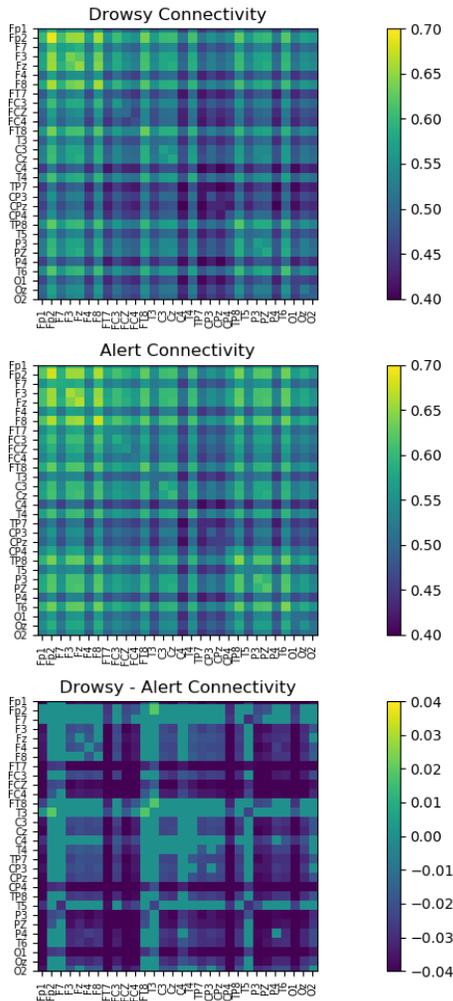


Fig. 5. Generated CAGNN connectivity in the drowsy state (upper panel), alert state (middle panel), and the difference between the drowsy and alert state. All the non-significant differences ( $p \geq 0.05$ ) were set to 0.

occlusion map indicates the saliency of a specific EEG clip. Four examples of selected occlusion maps are shown in Figure 6. As shown in figure 6a, the CAGNN model could recognize the alpha spindles and give a correct prediction of a drowsy sample with a high probability of 0.92. In Figure 6b, the CAGNN could capture the beta oscillation and predict an alert sample with a probability of 0.7. In figure 6c, CAGNN gave an incorrect prediction of a drowsy sample with a low probability of 0.51 since there were also alert features, which were the beta oscillations in this sample. In figure 6d, the CAGNN gave an incorrect prediction since there were also coexisting alpha spindles and beta oscillations in this example. For the other subjects, the CAGNN successfully captured alpha spindles in 10 of the 11 subjects with drowsy samples, and identified beta oscillations in all 11 subjects with alert samples, as demonstrated in the Supplementary Materials.

In addition, we analyzed the band power of the highlighted segments (occlusion map score  $> 0.5$ ) in Supplementary Figure 3. We found that when our model predicted drowsiness, it focused on the segments with higher alpha power, with an

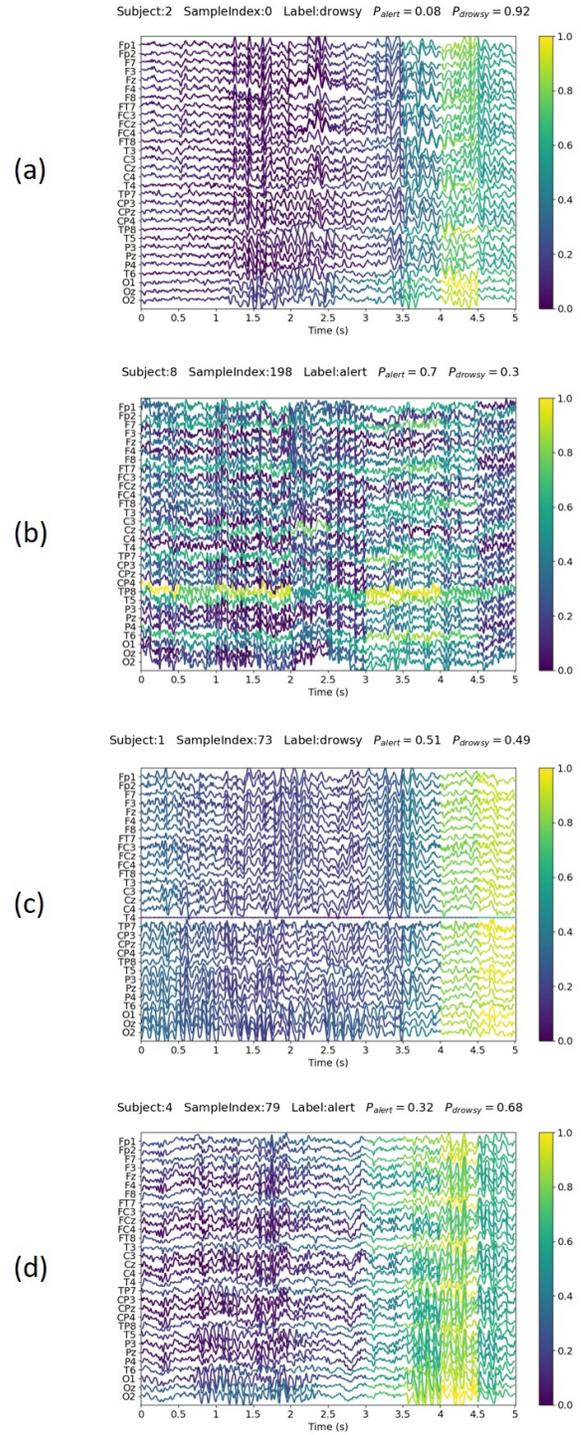


Fig. 6. Occlusion map analysis of the CAGNN for selected samples: (a) correctly classified drowsy sample, (b) correctly classified alert sample, (c) incorrectly classified drowsy sample, and (d) incorrectly classified alert sample. The channels are ordered in the same way as in Figure 5.

average alpha power of 2.43 dB when the model predicted drowsiness and an average value of 2.05 dB when the model predicted alertness; in contrast, when our model predicted alertness, it highlighted segments with higher beta power, with an average beta power of -0.60 dB when the model predicted drowsiness and a value of -0.29 dB when the model predicted alertness. A statistical analysis was performed via an

TABLE V  
PERFORMANCE COMPARISON AMONG DIFFERENT VARIANTS OF THE CAGNN

Encoder	Decoder	Accuracy	F1	Precision	Recall
CNN	GRU	63.0%	64.3%	68.8%	63.3%
Correlation Graph	GRU	69.0%	64.9%	75.4%	66.7%
Self-attention Graph	FFN	66.0%	63.4%	71.1%	60.5%
(Ours) Self-attention Graph	GRU	72.6%	70.7%	77.0%	71.2%

independent t-test, and both the alpha and beta power levels were significantly different ( $p < 0.01$ ).

## V. DISCUSSION

In this article, we proposed a task-relevant graph-generation method that can outperform other state-of-the-art CNN baselines in a driver drowsiness dataset. First, we applied a self-attention mechanism to generate connectivity in the GNN and showed that it can outperform other graph generation methods. Second, we utilized an SE attention block to improve the overall performance by 1% and showed that the SE attention could indicate the informative features correctly since when we retrained the model using the band with the highest SE attention score the accuracy improved by 2%. In this work, we first tested our model in the balanced dataset setting and showed that our method can surpass other methods by 3% in terms of accuracy, which helps researchers to establish the baseline understanding of this technology. After that, we tested on the imbalanced dataset and showed that our model could surpass other methods in terms of unbalanced dataset metrics including F1 score, precision, and recall, which means our method identifies the dangerous positive drowsy class well. Remarkably, the high precision and recall of our method revealed that our method can avoid false positives and false negatives effectively. By combining both balanced and imbalanced data testing, we can gain a more comprehensive understanding of the technology’s strengths, weaknesses, and potential for real-world application. Finally, we showed that our generated connectivity had lower occipital and parietal connectivity and lower interregional connectivity and was consistent with some neuroscience findings; it could provide some biological insights into EEG research.

### A. Effects of Each Building Block of the CAGNN

Our method leverages several deep learning architectures, including self-attention, a graph neural network, and a gated recurrent unit (GRU). To investigate the benefit of each individual component of our network, we analyzed the performance of several variants of our proposed method as shown in Table V. First, we investigated the effect of self-attention by replacing it with the correlation between channels, as proposed in [22]. This substitution led to an observed decrease in accuracy of approximately 3%, which might have been due to the self-attention mechanism that could explore the rich context between EEG channels to adaptively learn the brain network according to the downstream task. Second, we investigate the effect of the graph neural network by replacing the encoder with a CNN. The implementation followed the CNN-LSTM proposed in [62], with a replacement of LSTM to a GRU. The

CNN encoder proposed in this work consists of two convolutional layers (32 kernels of size  $3 \times 3$ ) to extract features from EEG time series. This decreased the accuracy significantly. This outcome implies that the graph neural network performs better when encoding the brain network because it has the ability to effectively capture and encode the intricate relationships between EEG channels. Third, we examine the effects of GRU by removing GRU and directly perform classification after the graph convolution via a feedforward network which is two fully connected layers ( $32 \times 16, 16 \times 2$ ). This also decreased the accuracy significantly, which illustrated the GRU was beneficial for EEG classification as it can model the dynamic temporal information.

### B. The Information Learned by the SE Block

In conventional graph neural networks [22], the input feature vectors are fed into the graph convolutional layer directly. In our approach, we introduced the SE block before the convolutional layer, and thus the interdependencies between different frequency features can be captured by the SE block, and the sensitivity of the model to more informative features increases. To exploit the dependencies between the features, the squeeze part of the SE block has a global average pooling layer to learn statistics between features, as depicted in equation (1). The excitation part uses a bottleneck gating mechanism, as depicted in equation (2) to generate the nonlinear interactions between features. We visualized the attention weight learned by the SE block and showed that it could be useful for feature selection. The theta band feature had the highest SE attention score, and when we retrained the model with only the theta band feature, the accuracy improved by approximately 2%. Therefore, the attention score of the SE block can reveal informative features.

### C. Biological Insights of the Generated Connectivity and Our Model

Our model could potentially reveal more connectivity information than traditional CNNs through end-to-end training. In contrast to static connectivity methods, our learnable connectivity patterns are relevant to downstream tasks and thus could increase classification accuracy. In addition, we found some drowsy connectivity features were identifiable from our generated connectivity. For instance, we found lower connectivity in the occipital region in the drowsy state. This fulfills our hypothesis since our task was visual sustained attention, and the occipital lobe is strongly correlated with visual information processing [63]. The lower connectivity in the occipital region might suggest the cortical gate mechanism that prevents the brain from the external sensory output [64].

Furthermore, our model revealed lower interregional connectivity in drowsiness, which might indicate that the state of the brain network changed to allow more efficient local information processing in drowsy states. This might indicate a local segregation effect, as depicted in [65]. We also explained why the model sometimes made a wrong prediction. As drowsy features are very complex and the coexistence of drowsy and alert features (e.g., alpha spindles and beta oscillations in Figure 6c) is very common, sometimes the model becomes confused and makes an incorrect decision. This may indicate that the participant was struggling against drowsiness and tried to perform the task well.

#### D. Challenges and Further Research Directions

EEG data is intrinsically more difficult to explain than image or natural language data. Therefore, in the future, we will explore some explainable model methods for graphs, such as graph attention networks [66], to improve explainability. Moreover, although our model had a smaller intersubject variance than other CNN-based models, the subject variance was still high. The leave-one-out accuracy in the balanced dataset ranged from 58.3 to 87.1, which means there was still a great difference in data distributions across different subjects. Meta-learning may be a way to address this problem in the future [67]. Furthermore, we only performed experiments on one dataset. Thus, we will test whether our method can be transferred to other neuroscience datasets as well. Another limitation of our dataset is that only sessions without motion have sufficient data for analysis. In the future, we will include more sessions with motion to evaluate our model in a more realistic setting. Additionally, in this work, to capture the entire brain network, a high-density 32-channel EEG system was utilized. After this study, we may limit EEG measurements to only the most critical regions. This strategy has the potential to significantly reduce the amount of setup time in future studies.

## VI. CONCLUSION

In this work, we proposed a novel connectivity-aware graph neural network (CAGNN) that is both connectivity-aware and task-aware. Our method is a new state-of-the-art approach for real-time drowsiness classification in both balanced and imbalanced datasets. We introduced the SE block to highlight the most important features, and we generated the graphs with our GNN model through self-attention mechanisms via end-to-end training. Furthermore, we showed that this approach not only improved the performance of our model but also provided meaningful neurophysiological explanations in terms of brain connectivity (e.g., local segregation) in drowsy states. We also demonstrated that our model could identify drowsy features such as alpha spindles. Our method could be applied in real-time driver drowsiness monitoring systems in the future.

## REFERENCES

- [1] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3017–3032, Dec. 2015.
- [2] A. Guettas, S. Ayad, and O. Kazar, "Driver state monitoring system: A review," in *Proc. 4th Int. Conf. Big Data Internet Things*, 2019, pp. 1–7.
- [3] M. Simon et al., "EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions," *Clin. Neurophysiol.*, vol. 122, no. 6, pp. 1168–1178, Jun. 2011.
- [4] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019.
- [5] M. S. Clayton, N. Yeung, and R. Cohen Kadosh, "The roles of cortical oscillations in sustained attention," *Trends Cognit. Sci.*, vol. 19, no. 4, pp. 188–195, Apr. 2015.
- [6] L. De Gennaro and M. Ferrara, "Sleep spindles: An overview," *Sleep Med. Rev.*, vol. 7, no. 5, pp. 423–440, Oct. 2003.
- [7] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci. Biobehavioral Rev.*, vol. 44, pp. 58–75, Jul. 2014.
- [8] T. Åkerstedt, G. Kecklund, and A. Knutsson, "Manifest sleepiness and the spectral content of the EEG during shift work," *Sleep*, vol. 14, no. 3, pp. 221–225, May 1991.
- [9] R. Hebert and D. Lehmann, "Theta bursts: An EEG pattern in normal subjects practising the transcendental meditation technique," *Electroencephalogr. Clin. Neurophysiol.*, vol. 42, no. 3, pp. 397–405, Mar. 1977.
- [10] J. W. Britton et al., "Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants," Amer. Epilepsy Soc., Chicago, IL, USA, 2016.
- [11] V. Sakkalis, "Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG," *Comput. Biol. Med.*, vol. 41, no. 12, pp. 1110–1117, Dec. 2011.
- [12] H. Wang et al., "Driving fatigue recognition with functional connectivity based on phase synchronization," *IEEE Trans. Cognit. Develop. Syst.*, vol. 13, no. 3, pp. 668–678, Sep. 2021.
- [13] R. Zheng, Z. Wang, Y. He, and J. Zhang, "EEG-based brain functional connectivity representation using amplitude locking value for fatigue-driving recognition," *Cognit. Neurodynamics*, vol. 16, no. 2, pp. 325–336, Apr. 2022.
- [14] I. M. Comsa, T. A. Bekinschtein, and S. Chennu, "Transient topographical dynamics of the electroencephalogram predict brain connectivity and behavioural responsiveness during drowsiness," *Brain Topography*, vol. 32, no. 2, pp. 315–331, Mar. 2019.
- [15] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.
- [16] R. T. Schirmer et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.
- [17] J. Cui et al., "A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel EEG," *Methods*, vol. 202, pp. 173–184, Jun. 2022.
- [18] J. Cui et al., "Subject-independent drowsiness recognition from single-channel EEG with an interpretable CNN-LSTM model," in *Proc. Int. Conf. Cyberworlds (CW)*, Sep. 2021, pp. 201–208.
- [19] C.-T. Lin, J. Liu, C.-N. Fang, S.-Y. Hsiao, Y.-C. Chang, and Y.-K. Wang, "Multistream 3-D convolution neural network with parameter sharing for human state estimation," *IEEE Trans. Cognit. Develop. Syst.*, vol. 15, no. 1, pp. 261–271, Mar. 2023.
- [20] C.-T. Lin, C.-H. Chuang, Y.-C. Hung, C.-N. Fang, D. Wu, and Y.-K. Wang, "A driving performance forecasting system based on brain dynamic state analysis using 4-D convolutional neural networks," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 4959–4967, Oct. 2021.
- [21] J. Cui, Z. Lan, O. Sourina, and W. Müller-Wittig, "EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, Oct. 2022.
- [22] S. Tang et al., "Self-supervised graph neural networks for improved electroencephalographic seizure analysis," in *Proc. Int. Conf. Learn. Represent.*, 2022. [Online]. Available: <https://openreview.net/forum?id=k9bx1EfHI->
- [23] J. Li, S. Li, J. Pan, and F. Wang, "Cross-subject EEG emotion recognition with self-organized graph neural network," *Frontiers Neurosci.*, vol. 15, Jun. 2021, Art. no. 611653. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/34177441>
- [24] N. Wagh and Y. Varatharajah, "EEG-GCNN: Augmenting electroencephalogram-based neurological disease diagnosis using a domain-guided graph convolutional neural network," in *Proc. Mach. Learn. Health*, 2020, pp. 367–378.

- [25] Y. Gao, X. Fu, T. Ouyang, and Y. Wang, "EEG-GCN: Spatio-temporal and self-adaptive graph convolutional networks for single and multi-view EEG-based emotion recognition," *IEEE Signal Process. Lett.*, vol. 29, pp. 1574–1578, 2022.
- [26] A. Demir, T. Koike-Akino, Y. Wang, and D. Erdoğan, "EEG-GAT: Graph attention networks for classification of electroencephalogram (EEG) signals," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 30–35.
- [27] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.
- [28] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2015, *arXiv:1508.04025*.
- [29] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [30] D. Hu, "An introductory survey on attention mechanisms in NLP problems," in *Proc. SAI Intell. Syst. Conf.* Springer, 2019, pp. 432–448.
- [31] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [32] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
- [33] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [34] Y. Li, Y. Liu, W.-G. Cui, Y.-Z. Guo, H. Huang, and Z.-Y. Hu, "Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 782–794, Apr. 2020.
- [35] B. Sun, X. Zhao, H. Zhang, R. Bai, and T. Li, "EEG motor imagery classification with sparse spectrotemporal decomposition and deep learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 541–551, Apr. 2021.
- [36] W. Tao et al., "EEG-based emotion recognition via channel-wise attention and self attention," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 382–393, Jan. 2023.
- [37] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Sci. Data*, vol. 6, no. 1, pp. 1–8, Apr. 2019.
- [38] J. B. Estrach, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and deep locally connected networks on graphs," in *Proc. 2nd Int. Conf. Learn. Represent. (ICLR)*, 2014.
- [39] Y. Dong, Q. Liu, B. Du, and L. Zhang, "Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 31, pp. 1559–1572, 2022.
- [40] C. Luo et al., "Disrupted functional brain connectivity in partial epilepsy: A resting-state fMRI study," *PLoS ONE*, vol. 7, no. 1, Jan. 2012, Art. no. e28196.
- [41] C.-T. Lin, I.-F. Chung, L.-W. Ko, Y.-C. Chen, S.-F. Liang, and J.-R. Duann, "EEG-based assessment of driver cognitive responses in a dynamic virtual-reality driving environment," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 7, pp. 1349–1352, Jul. 2007.
- [42] K. Huang, C. Chuang, Y. Wang, C. Hsieh, J. King, and C. Lin, "The effects of different fatigue levels on brain-behavior relationships in driving," *Brain Behav.*, vol. 9, no. 12, p. e01379, Dec. 2019.
- [43] Z. Cao, M. Chuang, J. King, and C.-T. Lin. (2019). *Multi-Channel EEG Recordings During a Sustained-Attention Driving Task*. [Online]. Available: [https://figshare.com/articles/Multi-channel\\_EEG\\_recordings\\_during\\_a\\_sustained-attention\\_driving\\_task\\_preprocessed\\_dataset\\_/7666055](https://figshare.com/articles/Multi-channel_EEG_recordings_during_a_sustained-attention_driving_task_preprocessed_dataset_/7666055)
- [44] O. Rosanne, I. Albuquerque, J.-F. Gagnon, S. Tremblay, and T. H. Falk, "Performance comparison of automated EEG enhancement algorithms for mental workload assessment of ambulant users," in *Proc. 9th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Mar. 2019, pp. 61–64.
- [45] Y. Liu et al., "Fusion of spatial, temporal, and spectral EEG signatures improves multilevel cognitive load prediction," *IEEE Trans. Human-Mach. Syst.*, vol. 53, no. 2, pp. 357–366, Apr. 2023.
- [46] R. Luce, "Response time distributions in memory search: A caution," in *Human Memory & Cognitive Capabilities: Mechanisms & Performances: Part A*. Amsterdam, The Netherlands: North-Holland, 1986.
- [47] C.-S. Wei, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "Toward drowsiness detection using non-hair-bearing EEG-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 400–406, Feb. 2018.
- [48] C.-T. Lin, R.-C. Wu, S.-F. Liang, W.-H. Chao, Y.-J. Chen, and T.-P. Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 12, pp. 2726–2738, Dec. 2005.
- [49] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010.
- [50] S. Eickhoff and V. Müller, "Functional connectivity," in *Brain Mapping*, A. W. Toga, Ed. Waltham, MA, USA: Academic, 2015, pp. 187–201. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780123970251002128>
- [51] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*.
- [52] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.
- [53] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016.
- [54] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw., IEEE World Congr. Comput. Intell.*, Jun. 2008, pp. 2390–2397.
- [55] X. Niu et al., "Effects of sleep deprivation on functional connectivity of brain regions after high-intensity exercise in adolescents," *Sustainability*, vol. 14, no. 23, p. 16175, Dec. 2022.
- [56] C.-S. Huang, N. R. Pal, C.-H. Chuang, and C.-T. Lin, "Identifying changes in EEG information transfer during drowsy driving by transfer entropy," *Frontiers Hum. Neurosci.*, vol. 9, p. 570, Oct. 2015.
- [57] J. Chen, H. Wang, Q. Wang, and C. Hua, "Exploring the fatigue affecting electroencephalography based functional brain networks during real driving in young males," *Neuropsychologia*, vol. 129, pp. 200–211, Jun. 2019.
- [58] Y. Qin et al., "Directed brain network analysis for fatigue driving based on EEG source signals," *Entropy*, vol. 24, no. 8, p. 1093, Aug. 2022.
- [59] M. Massimini, F. Ferrarelli, R. Huber, S. K. Esser, H. Singh, and G. Tononi, "Breakdown of cortical effective connectivity during sleep," *Science*, vol. 309, no. 5744, pp. 2228–2232, Sep. 2005.
- [60] S. Zhang, J. Sun, and X. Gao, "The effect of fatigue on brain connectivity networks," *Brain Sci. Adv.*, vol. 6, no. 2, pp. 120–131, Jun. 2020.
- [61] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2014, pp. 818–833.
- [62] D. Ahmedt-Aristizabal, T. Fernando, S. Denman, L. Petersson, M. J. Aburn, and C. Fookes, "Neural memory networks for seizure type classification," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 569–575.
- [63] K. C. Bettencourt and Y. Xu, "Decoding the content of visual short-term memory under distraction in occipital and parietal areas," *Nature Neurosci.*, vol. 19, no. 1, pp. 150–157, Jan. 2016.
- [64] S. K. Esser, S. Hill, and G. Tononi, "Breakdown of effective connectivity during slow wave sleep: Investigating the mechanism underlying a cortical gate using large-scale modeling," *J. Neurophysiol.*, vol. 102, no. 4, pp. 2096–2111, Oct. 2009.
- [65] G. N. Dimitrakopoulos et al., "Functional connectivity analysis of mental fatigue reveals different network topological alterations between driving and vigilance tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 740–749, Apr. 2018.
- [66] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," *arXiv:1710.10903*.
- [67] T. Duan, M. Chauhan, M. Abuzar Shaikh, J. Chu, and S. Srihari, "Ultra efficient transfer learning with meta update for cross subject EEG classification," 2020, *arXiv:2003.06113*.