TIE-EEGNet: Temporal Information Enhanced EEGNet for Seizure Subtype Classification

Ruimin Peng, Changming Zhao, Jun Jiang, Guangtao Kuang, Yuqi Cui, Yifan Xu, Hao Du, Jianbo Shao, and Dongrui Wu, *Senior Member, IEEE*

***Abstract*— Electroencephalogram (EEG) based seizure subtype classification is very important in clinical diagnostics. However, manual seizure subtype classification is expensive and time-consuming, whereas automatic classification usually needs a large number of labeled samples for model training. This paper proposes an EEGNet-based slim deep neural network, which relieves the labeled data requirement in EEG-based seizure subtype classification. A temporal information enhancement module with sinusoidal encoding is used to augment the first convolution layer of EEGNet. A training strategy for automatic hyper-parameter selection is also proposed. Experiments on the public TUSZ dataset and our own CHSZ dataset with infants and children demonstrated that our proposed TIE-EEGNet outperformed several traditional and deep learning models in cross-subject seizure subtype classification. Additionally, it also achieved the best performance in a challenging transfer learning scenario. Both our code and the CHSZ dataset are publicized.**

***Index Terms*— EEG, seizure subtype classification, EEGNet, deep learning, positional encoding.**

I. Introduction

EPILEPSY, a chronic non-communicable disease caused by the abnormal supersynchronous electrical activity of brain neurons, is one of the most common neurological diseases affecting around 70 million people of all ages around the world [1], [2]. Repeated epileptic seizures have persistent negative effects on the mental and cognitive functions of the patients, life-threatening in severe situations. Moreover, the etiological and clinical symptoms of epilepsy could be diverse and complicated due to the initial lesion and propagation mode differences among individuals.

Manuscript received 17 May 2022; revised 4 August 2022 and 12 August 2022; accepted 14 August 2022. Date of publication 5 September 2022; date of current version 15 September 2022. This work was supported by the Wuhan Science and Technology Bureau under Grant 2020020601012240. *(Corresponding authors: Jianbo Shao; Dongrui Wu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the Wuhan Children's Hospital under Application No. 2022R034-E01.

Ruimin Peng, Changming Zhao, Yuqi Cui, Yifan Xu, and Dongrui Wu are with the Key Laboratory of the Ministry of Education for Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430000, China (e-mail: drwu@hust.edu.cn).

Jun Jiang, Guangtao Kuang, Hao Du, and Jianbo Shao are with Children's Hospital, Wuhan 430016, China (e-mail: shaojb2002@sina.com).

Digital Object Identifier 10.1109/TNSRE.2022.3204540

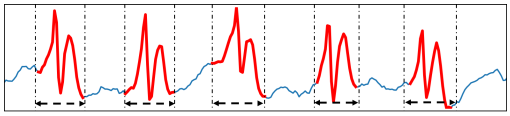

Fig. 1. An example of epileptic EEG signal with spike-and-wave discharges (in red).

Electroencephalogram (EEG), the electrophysiological signal generated by synchronizing neuronal activities in the brain, is the most popular and successful non-invasive signal for epilepsy-related disease diagnosis. Epileptic EEG signals are characterized by spike-and-wave discharges (SWDs) [3], as shown in Fig. 1. Reading EEG signals for epilepsy diagnosis requires experienced clinicians. Such labor-intensive work not only demands a significant time investment but also relies heavily on the experience and subjective judgments of the labelers. To relieve the burden on clinicians and improve the treatment efficiency, automated epilepsy diagnosis algorithms are desired.

More specifically, for clinical epilepsy diagnosis, seizure detection is the first step. The next is to determine the subtype, which is important for identifying epilepsy syndromes, targeted therapies, and eligibility for epilepsy surgery [4]. According to the 2017 International League Against Epilepsy guideline [5], [6], subtypes of epilepsy include:

1) Generalized seizures, e.g., absence seizures and tonic-clonic seizures, which affect both sides of the brain.
2) Focal seizures, e.g., simple and complex focal seizures, which are distinguished by the intensity of patients' consciousness located in only one area of the brain.
3) Generalized and focal seizures, which begin at one part of the brain, then spread to both sides.

Machine learning algorithms are promising for automatic seizure subtype classification. In a typical scenario with traditional machine learning approaches, EEG artifacts, e.g., eye/muscle movements and electrical noise, are first removed by band-pass filtering and detrending [7], [8]. Then, time domain features [9], frequency domain features [10], temporal-spatial features [11], [12], or nonlinear features [13], can be extracted, selected [14], and finally sent to a classifier, e.g., support vector machine (SVM) [10], [15], logistic regression (LR) [12], or multilayer perceptron (MLP) [16], for classification [17].

To alleviate the difficulty of manual feature extraction, deep learning, particularly Convolutional Neural Networks (CNNs), have become popular [18], [19], [20]. However, there are two challenges in CNN-based approaches for seizure subtype classification:

1) Insufficient labeled data. Epilepsy datasets are usually constructed from a small number of patients with limited number of seizures. A large neural network may easily overfit them [21]. Therefore, slim networks are preferred. There are generally two strategies to obtain slim neural networks: i) Design new network architectures, especially convolution structures, which can achieve satisfactory classification performance with fewer parameters [22], [23]; and, ii) compress a large neural network by knowledge distilling [24], [25], network pruning [26], [27], or low-rank decomposition [28].
2) Difficulty to use the raw EEG signals directly. Different from images or audio, EEG signals are two dimensional (2D) time series with spatial correlations among channels. Many CNN-based approaches for EEG classification first transform the 1D signal of each channel into 2D by wavelet transform or spectrogram, which is then input to a conventional CNN [29]. However, the design of transformation relies on expert knowledge, which may not always be available or optimal. Thus, deep learning models that can take the raw EEG signals as input are more desirable. Unfortunately, it is very challenging to design network architectures to effectively extract both spatial and temporal information [30].

This paper proposes a novel temporal information enhancement module, which can be added to EEGNet [31] for cross-patient seizure subtype classification. Our main contributions are:

1) We propose a novel model named Temporal Information Enhanced EEGNet (TIE-EEGNet), which augments the standard EEGNet with a novel TIE module. With specially designed auxiliary sinusoidal encoding, the TIE module augments temporal patterns to the original convolution layer by adding time positional embedding. In addition, a training strategy is proposed for automatic hyper-parameter selection.
2) We further consider a transfer learning scenario, and use fine-tuning to boost the pre-trained model's performance.
3) We collected an epilepsy dataset of infant and children patients with subtype annotations, which was used in our experiments, and is also available to the research community by contacting the authors.

Our Python code is publicized at https://github.com/rmpeng/TIE-EEGNet, so that others can easily duplicate our results and further improve them.

The rest of this paper is organized as follows. Section II briefly reviews related works. Section III introduces the details of our proposed TIE-EEGNet. Section IV compares the performance of TIE-EEGNet with those of existing approaches on two datasets. Section V draws conclusions and outlines some future research directions.

II. Related Work

This section briefly reviews existing approaches for seizure subtype classification.

A. EEG-Based Seizure Subtype Classification

Though seizure subtype classification is crucial for precise medical treatment, few studies have attempted to explore it, due to the complexity of multi-class classification and the lack of labeled clinical datasets.

For traditional machine learning, feature extraction and classification are performed separately. Roy *et al.* [32] adopted a KNN model to classify the frequency features generated by Fast Fourier Transform, achieving an F_1 score of 0.907. Saputro *et al.* [33] used Mel Frequency Cepstral Coefficients features and Hjorth Descriptor in an SVM classifier, achieving an accuracy of 0.914.

Deep learning eliminates the burden of manual feature extraction, and hence becomes popular in seizure classification. Asif *et al.* [34] proposed SeizureNet, a CNN-based neural network, which obtained 0.620 weighted F_1 score for within-patient seizure type classification. Li *et al.* [35] proposed CE-stSENet, a multi-scale model embedded with group convolution, to process the temporal and spatial data in parallel, achieving an average F_1 score of 0.937 and an accuracy of 0.920.

Most existing approaches for automatic seizure classification are either patient-specific (build a separate machine learning model for each patient, where the training and test data come from the same patient) or patient-independent (all patients' data are mixed together, and then partitioned randomly into training and test sets, without distinguishing among individuals or considering the temporal causality). These scenarios are under the assumption that distributions of the training and test data are consistent, which unfortunately does not always hold in reality due to individual differences and non-stationarity of EEG signals. As a result, these models have limited generalization and robustness. It is desirable to learn a model from already collected data and generalize well to new patients.

B. CNN-Based Slim Neural Networks

In data scarcity applications, e.g., seizure classification and brain-computer interfaces (BCIs) [36], slim neural networks are preferred to reduce overfitting.

Multiple slim CNN architectures have been proposed. For example, Ding *et al.* [37] proposed ACNet, which introduces an asymmetric convolution block to decompose one traditional square-kernel convolutional layer into three parallel branches of convolutional layers with extra horizontal and vertical kernels to emphasize the skeleton position. SqueezeNet [21] embeds a Fire module into CNN to reduce the number of parameters. The Fire module consists of a squeeze layer with kernel size 1×1 to compress the feature maps, and an expand layer with kernel size 1×1 and 3×3 to recover the feature maps' resolution. Hu *et al.* [38] developed Squeeze-and-Excitation Network, which embeds a parallel

squeeze-and-excitation block into the original network. The squeeze-and-excitation block obtains a global distribution of channel-wise feature representation, improving the feature maps' quality by incorporating spatial attention. MobileNet [22] substitutes the traditional convolutional layer with depth-wise separable convolutional layers to reduce the number of parameters and to increase the efficiency. First, depth-wise convolution learns features from each channel. The following point-wise convolution helps each feature map to obtain global feature information. Similarly, ShuffleNet [23] and Xception [39] adopt depth-wise convolution to slim CNN-based neural networks.

EEGs are also frequently used in BCIs, and hence different deep learning classifiers for EEG-based BCIs have been proposed. EEGNet [31] is one of the most popular among them. It adopts depth-wise and separable convolutional layers to learn the temporal patterns of each channel and correlations among different channels separately. EEGNet has achieved promising performances in multiple BCI paradigms, with only about 1000 parameters.

C. Learning Temporal Information

Due to the fact that square convolution filters have small receptive fields, traditional CNNs may not be good at processing time-series data.

To expand the receptive fields, Bai *et al.* [40] proposed Temporal Convolutional Network, which combines causal convolutions and dilated convolutions hierarchically. Therefore, neurons of deep layers receive messages from a long sequence of neurons in shallow layers, enabling them to learn long-term temporal information. CE-stSENet [35] and multi-scale CNN [41] reconstruct the raw time series data into multiple scales and obtain different fine-grained temporal information, expanding the receptive fields. Inspired by recurrent neural networks, Liang and Hu [42] proposed recurrent CNN to learn EEG feature representations. After transforming the EEG series into topographs, it employs the CNN-based ConvNet to extract local features, and a long short-term memory layer to learn the temporal information. Similarly, Zhang *et al.* [43] proposed a convolution recurrent attention model for BCI tasks. Singh and Lin [44] proposed time encoding kernel (EnK) to embed time information into the data during horizontal shifting of a kernel in convolution operations.

III. Method

This section proposes TIE-EEGNet, which enhances EEGNet on learning temporal information with a TIE module.

A. EEGNet

EEGNet is a CNN-based compact network widely used in EEG-based BCIs. Inspired by Xception [39], EEGNet introduces depth-wise convolution and separable convolution that encapsulate the temporal and spatial filers to substitute the traditional square convolution, reducing the number of parameters. Its effectiveness in multiple BCI paradigms has been demonstrated in [31]. This paper adopts EEGNet as the basic model to improve upon.

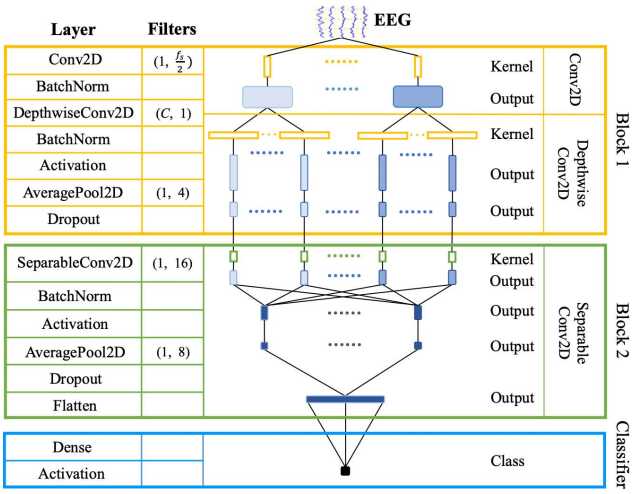

Fig. 2. Architecture of the standard EEGNet.

The architecture of the basic EEGNet is shown in Fig. 2. In the first block, two convolutional layers are arranged sequentially. The first layer is dedicated to capturing the input EEG signal's feature maps containing different band-pass frequencies. Its kernel size is $(1, f_s/2)$, where f_s is the sampling frequency. The next layer executes depth-wise convolution with kernel size $(C, 1)$, where C is the number of channels. Without full connection to previous feature maps, this layer not only reduces the number of trainable parameters, but also explicitly learns spatial filters for each temporal filter. The outputs of Block 1 are frequency-specific spatial feature maps.

Block 2 employs separable convolution, with a depth-wise convolutional layer followed by a point-wise convolutional layer. In addition to reducing the number of parameters, relationships within and across feature maps are obtained in a decoupled way. Finally, the classifier is a fully-connected layer with softmax activation.

B. Encoding Kernel (EnK)

Without increasing the number of parameters, EnK [44] embeds linear increasing time encoding into translation-invariant kernels of several CNN-based models, including EEGNet.

For an input $\mathrm{X} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}}$, let

$$\mathrm{P}^{u,v} = \begin{bmatrix} \mathrm{X}_{u-m,v-n} & \cdots & \mathrm{X}_{u-m,v+n} \\ \vdots & \ddots & \vdots \\ \mathrm{X}_{u+m,v-n} & \cdots & \mathrm{X}_{u+m,v+n} \end{bmatrix} \in \mathbb{R}^{(2m+1)\times(2n+1)}, \quad (1)$$

and $\mathrm{K} \in \mathbb{R}^{(2\mathrm{m}+1)\times(2\mathrm{n}+1)}$ be the convolutional kernel. Then, the (u, v)th element of the output feature map $\mathrm{Z} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}}$, $\mathrm{Z}_{u,v}$, obtained from a typical dot-product convolution, is:

$$\mathrm{Z}_{u,v} = \mathrm{P}^{u,v} \cdot \mathrm{K} = \sum_{i=-m}^{m} \sum_{j=-n}^{n} \mathrm{X}_{u+i,v+j} * \mathrm{K}_{i,j}. \quad (2)$$

In EnK, the convolutional kernel uses linear time encoding to embed temporal information into the convolution operation:

$$\mathrm{Z}_{u,v} = \mathrm{P}^{u,v} \cdot (\mathrm{K} + b * t_v), \tag{3}$$

where $t_v = v$ denotes the time position of $\mathrm{P}^{u,v}$. The trainable scaling factor b controls the intensity of the time information. By updating the kernel with intuitive linear chronological order during convolution dynamically, EnK enhances the time-sequence-related representation of the feature maps.

(3) shows that a later time point, which has a larger t_v, contributes more to $\mathrm{Z}_{u,v}$. However, a more uniform contribution to the time positional embedding is desired. Additionally, the magnitude of the time positional embedding should be similar to that of the elements in $\mathrm{P}^{u,v}$.

C. Temporal Information Enhancement (TIE)

To further enhance EEGNet's ability to learn temporal information for seizure subtype classification, we propose a novel TIE module.

(3) in EnK is equivalent to (4) below, where Z is decomposed into a convolution with kernel K and a time positional embedding generated by linear encoding,

$$\begin{aligned}\mathrm{Z}_{u,v} &= \mathrm{P}^{u,v} \cdot \mathrm{K} + \mathrm{P}^{u,v} \cdot (b * t_v) \\ &= \mathrm{P}^{u,v} \cdot \mathrm{K} + (b * t_v) * \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \mathrm{X}_{u+i,\,v+j}\end{aligned} \tag{4}$$

According to (4), $\mathrm{Z}_{u,v}$ is gradually dominated by t_v when t_v increases, so EnK's ability to detect SWDs decreases gradually. To resolve this problem, TIE designs time positional encoding to satisfy the following three requirements:

1) It should be bounded.
2) It should be periodic, to simulate the repeated SWDs.
3) It should preserve the time sequence information that reflects the temporal features such as wave shape in a period.

Sinusoidal functions are used in TIE, as in Transformer [45]. More specifically, the proposed sinusoidal encoding (SE) is:

$$SE(t_v) = \begin{cases} \sin\left(\dfrac{t_v}{\omega}\right), & t_v \text{ is even} \\ \cos\left(\dfrac{t_v}{\omega}\right), & t_v \text{ is odd,} \end{cases} \tag{5}$$

where ω is the SE period regulator.

Additionally, TIE uses the average of all elements in $\mathrm{P}^{u,v}$, instead of their sum as in EnK, in temporal information encoding. As a result, the encoding has the same magnitude as the elements in $\mathrm{P}^{u,v}$.

In summary, the proposed TIE module is:

$$\mathrm{Z}_{u,v} = \mathrm{P}^{u,v} \cdot \mathrm{K} + b * SE(t_v) * \mathrm{R}_{u,v}, \tag{6}$$

where the element of representation matrix R is calculated by:

$$\mathrm{R}_{u,v} = \frac{1}{(2m+1) * (2n+1)} * \sum_{i=-m}^{m} \sum_{j=-n}^{n} \mathrm{X}_{u+i,v+j}. \tag{7}$$

$\mathrm{R}_{u,v}$, implemented by average pooling, represents the background information of $\mathrm{P}^{u,v}$.

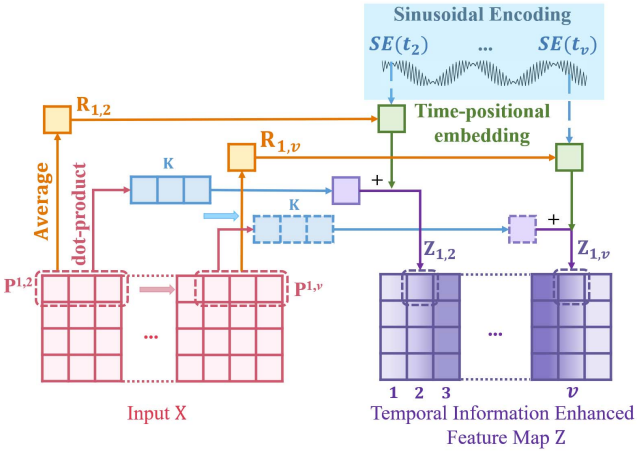

Fig. 3. Workflow of the proposed TIE module with sinusoidal encoding for convolution. For simplicity, $m = 0$ and $n = 1$ are used. The kernel K in blue is sliding during convolution, and its generated original feature representation pixels are marked in light purple. The yellow pixels $\mathrm{R}_{u,v}$ represent pixels of the representation matrix, which are the average of $\mathrm{P}^{u,v}$. Then, sinusoidal encoding generates time positional embedding from $\mathrm{R}_{u,v}$, shown as the green blocks. Next, the time positional embedding is added to the original feature representation corresponding to the location. The resulting TIE feature map Z is marked in purple.

TABLE I
EEG FREQUENCY BANDS

Frequency Band (Hz)	δ	θ	α	β	γ	μ
Lower bound	0	4	8	12∼16*	32	8
Upper bound	4	8	12∼16*	32	64	12

* The literature has not reached a consensus on the upper bound of the α wave and the lower bound of the β wave[1].

Fig. 3 illustrates the workflow of TIE with sinusoidal encoding.

D. TIE-EEGNet

According to previous studies [46], EEG is typically described in terms of rhythmic activities and transients. The rhythmic activities are divided into bands by frequencies, as shown in Table I. Note that infants and children's brains are in development, and hence the rhythmic activities of their EEG signals are significantly different from adults [47]. Also, EEG activities vary across individuals. Thus, it is of practical significance to establish a reliable and general model that is robust to the variations among patients.

A fixed ω in (5) may not give the optimal performance. So, we try to select the optimal ω from a candidate list $\boldsymbol{\omega}$.

Let f_s be the sampling rate, and $\boldsymbol{f_c} = \{0.5, 4, 8, 12, 16, 32, 64\}$ the crucial frequency list consisting of the lower and upper frequency bounds of all EEG bands in Table I. To correlate with EEG frequency bands, a candidate ω of sinusoidal encoding should satisfy:

$$f_c * \omega = f_s, \quad f_c \in \boldsymbol{f}_c. \tag{8}$$

Thus, $\boldsymbol{\omega}$ is generated by:

$$\boldsymbol{\omega} = \frac{f_s}{\boldsymbol{f}_c}. \tag{9}$$

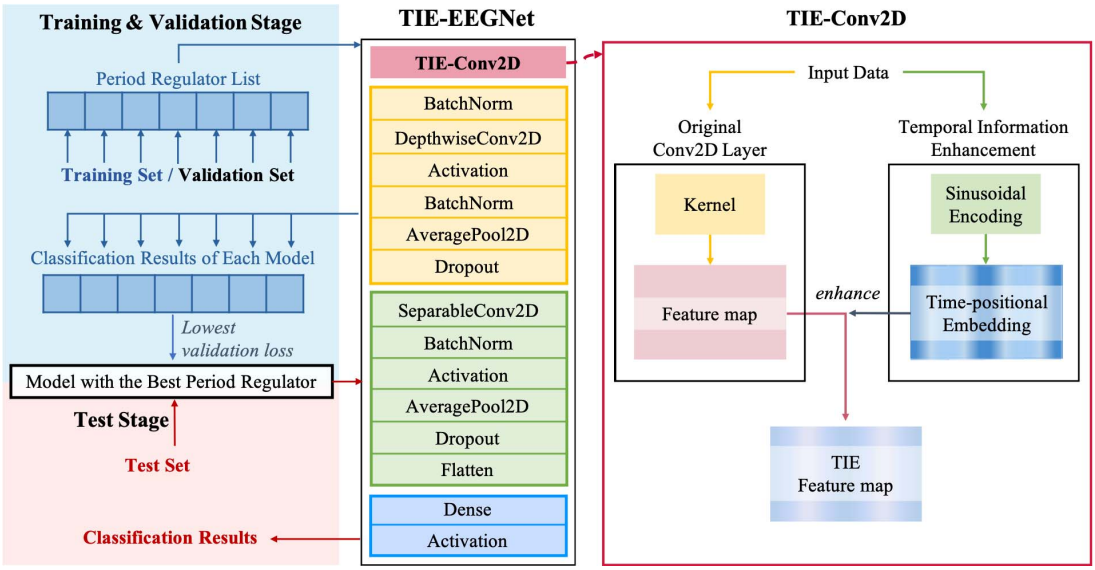

Fig. 4. Framework of our proposed TIE-EEGNet.

We then train the TIE module with different ω in parallel, and select the best model according to the lowest validation loss.

Fig. 4 shows the framework of our proposed TIE-EEGNet. In the standard EEGNet, the first Conv2D layer extracts temporal feature representations of the raw EEG samples with different pass-bands by temporal kernels (filters). To enhance the temporal information, we augment the TIE module to the Conv2D layer to construct a TIE-Conv2D layer, which adds the time positional embedding generated by sinusoidal encoding to the feature maps of the original Conv2D. Hence, the TIE module enhances the (periodical) temporal information of the feature maps. The feature maps from TIE-Conv2D have the same size as those from the original Conv2D, which can readily be used by subsequent layers.

In the training and validation stage, models with different candidate period regulator ω from $\boldsymbol{\omega}$ are trained in parallel. The model with the lowest validation loss is used in test.

IV. Experimental Study

This section evaluates the performance of TIE-EEGNet in two seizure subtype classification tasks. We also tested TIE-EEGNet in a source-free transfer learning scenario, which has not been considered before. Our Python code is available at https://github.com/rmpeng/TIE-EEGNet.

A. Datasets

Experiments were conducted on both the widely used public TUSZ dataset and our own CHSZ dataset.

The TUSZ dataset, which is a portion of the Temple University Hospital EEG Corpus (TUH EEG Corups), is one of the largest publicly available epileptic EEG datasets with seizure subtype annotations [48]. Table II summarizes the distribution of seizure types in it.

To be consistent with our collected dataset and for the application of transfer learning, Generalized Seizure (GNSZ), Focal Non-Specific Seizure (FNSZ) and Myoclonic Seizure (MYSZ) were not considered. Besides, Complex Partial Seizure (CPSZ) and Simple Partial Seizure (SPSZ) were combined into Focal Seizure (FSZ) according to the ILAE guideline [5]. Thus, 4-class [FSZ, Absence Seizure (ABSZ), Tonic Seizure (TNSZ), Tonic-Clonic Seizure (TCSZ)] seizure subtype classification was performed on the TUSZ dataset.

The CHSZ seizure dataset was collected from the Wuhan Children's Hospital affiliated to Tongji Medical College of the Huazhong University of Science and Technology, and

TABLE II
THE TUSZ DATASET

Seizure Type	Number	Description*
Focal Non-specific Seizure (FNSZ)	988	A large category of seizures occurring in a specific focality
Generalized Seizure (GNSZ)	413	A large category of seizures occurring in most if not all of the brain
Simple Partial Seizure (SPSZ)	44	Brief seizures starting in one location of the brain (and may spread). The patient is fully aware and able to interact
Complex Partial Seizure (CPSZ)	342	Partial seizure with impaired awareness
Absence Seizure (ABSZ)	76	Brief, sudden seizure involving lapse in attention. Usually lasts no more than 5 seconds and commonly seen in children
Tonic Seizure (TNSZ)	67	Seizure involving the stiffening of muscles. Usually associated with and annotated as tonic-clonic seizures, but not always (rarely there is no clonic phase)
Tonic-Clonic Seizure (TCSZ)	50	Seizure involving losing consciousness and violent muscle contractions
Myoclonic Seizure (MYSZ)	3	Seizure associated with brief involuntary twitching or myoclonus

* Annotation descriptions are from the documentation on the TUH EEG Corpus website https://isip.piconepress.com/projects/tuh_eeg/html/downloads.shtml#d_anno.

TABLE III
THE CHSZ DATASET

Patient ID	Sampling Rate (Hz)	Seizure Events	Type
000	698	1 1 1	FSZ TNSZ TCSZ
001	500	1 1 1	FSZ TNSZ TCSZ
003	500	5	FSZ
004	500	2	FSZ
005	500	3 3	TNSZ TCSZ
006	500	1	FSZ
007	500	4	FSZ
008	1000	1 1	TNSZ TCSZ
009	500	2 2	TNSZ TCSZ
010	500	2 2	TNSZ TCSZ
011	500	2	ABSZ
012	698	18	ABSZ
013	698	2	ABSZ
014	500	4	FSZ
015	500	30	FSZ
016	500	3	FSZ
017	698	16	FSZ
018	500	24	ABSZ
019	500	1	FSZ
020	698	1	FSZ
021	698	16	ABSZ
022	500	8	FSZ
023	500	21	ABSZ
024	500	1 1 1	FSZ TNSZ TCSZ
026	698	1 1 1	FSZ TNSZ TCSZ

approved by the Ethics Committee of the Wuhan Children's Hospital (2022R034-E01). It includes 22-channel EEG signals from 25 infant and children patients. Different sampling rates, e.g., 500 Hz, 698 Hz and 1000 Hz, were used for different patients. Table III shows the details. Researchers can request this dataset by contacting the authors.

B. Preprocessing

For both datasets, EEG recordings were first downsampled to 128 Hz. 20 common channels [35] listed in Table IV were selected. Then, 50 Hz notch filtering, 0-64 Hz low-pass filtering, and detrending were employed to remove artifacts and noise.

TABLE IV
THE 20 COMMON CHANNELS USED IN OUR EXPERIMENTS

FP1 - F7	F7 - T3	T3 - T5	T5 - O1	FP2 - F8
F8 - F4	T4 - T6	T6 - O2	T3 - C3	C3 - CZ
CZ - C4	C4 - T4	FP1 - F3	F3 - C3	C3 - P3
P3 - O1	FP2 - F4	F4 - C4	C4 - P4	P4 - O2

We considered a more realistic cross-patient classification scenario. Fig. 5 depicts the experimental setup, including patient subset partition and training/validation/test sample generation.

We used 3-fold cross-validation. Patients were shuffled and partitioned into three subsets, ensuring that the number of events of each seizure subtype in the three subsets are similar. Each of the three subsets took turns to be the test set, and the other two were merged as the training and validation sets. To ensure that the training and validation sets do not overlap, the first half events were used for training, and the last half for validation. Then, the seizure events of all sets were empirically sliced by a 4-second sliding window to generate samples. Two successive sliding windows had 50% overlap to increase the number of training samples, but no overlap at all in generating the validation and test samples.

C. Performance Measures

As shown in Tables II and III, different seizure subtypes had significantly different number of events. Therefore, we used the balanced classification accuracy (BCA) and the weighted F_1 score as our performance measures. Both were already implemented in the scikit-learn[1] package of Python.

To simulate real-world applications, we considered event-level classification rather than sample-level in the final evaluation. More specifically, seizure event recordings were sliced into 4-second fragments as training/validation/test samples, and each fragment was classified with probabilities of different seizure types. Since labels were assigned for each seizure event, fragments from the same event should have the same class label, even though they may be classified as different types by the model. Thus, majority voting was used to aggregate the classifications of the fragments in the same event into a final class. In performance evaluation, both BCA and F_1 score were calculated by comparing the classified and ground-truth labels of each seizure event, rather than each sample (fragment).

D. Experimental Results in Seizure Subtype Classification

TIE-EEGNet was compared with both end-to-end deep learning approaches and traditional machine learning approaches using handcrafted features.

Deep learning baselines are EEGNet, EnK-EEGNet, EEGWaveNet [49], and CE-stSENet. In addition, we improved EnK-EEGNet by replacing its summation representation matrix R with the average representation matrix in (7), named

[1] https://scikit-learn.org/

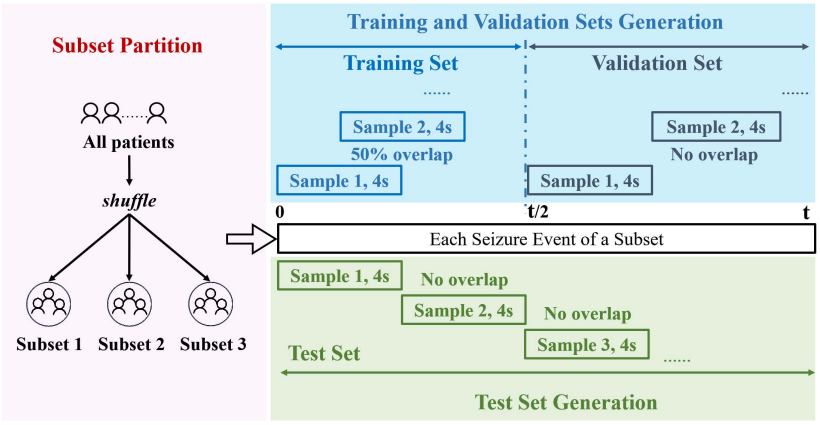

Fig. 5. Patient subset partition and training/validation/test sample generation.

TABLE V
AVERAGE PERFORMANCES ON THE TWO DATASETS FOR 4-CLASS SEIZURE SUBTYPE CLASSIFICATION

	TUSZ		CHSZ	
	BCA	F_1	BCA	F_1
SVM	$0.532_{\pm 0.112}$	$0.641_{\pm 0.068}$	$0.461_{\pm 0.083}$	$0.498_{\pm 0.109}$
RC	$0.512_{\pm 0.028}$	$0.652_{\pm 0.143}$	$0.474_{\pm 0.066}$	$0.475_{\pm 0.183}$
LR	$0.512_{\pm 0.035}$	$0.663_{\pm 0.125}$	$0.449_{\pm 0.084}$	$0.450_{\pm 0.199}$
GBDT	$0.459_{\pm 0.077}$	$0.704_{\pm 0.088}$	$0.562_{\pm 0.025}$	$0.606_{\pm 0.067}$
EEGWaveNet	$0.579_{\pm 0.055}$	$0.694_{\pm 0.014}$	$0.494_{\pm 0.036}$	$0.457_{\pm 0.021}$
CE-stSENet	$0.545_{\pm 0.028}$	$0.701_{\pm 0.050}$	$0.567_{\pm 0.055}$	$0.615_{\pm 0.054}$
EEGNet	$0.557_{\pm 0.035}$	$0.588_{\pm 0.051}$	$0.482_{\pm 0.032}$	$0.290_{\pm 0.049}$
EnK-EEGNet	$0.373_{\pm 0.022}$	$0.151_{\pm 0.054}$	$0.375_{\pm 0.033}$	$0.340_{\pm 0.045}$
Avg-EnK-EEGNet	$0.558_{\pm 0.026}$	$0.574_{\pm 0.053}$	$0.395_{\pm 0.051}$	$0.346_{\pm 0.071}$
TIE-EEGWaveNet	$0.591_{\pm 0.036}$	$0.700_{\pm 0.014}$	$0.509_{\pm 0.013}$	$0.459_{\pm 0.032}$
TIE-CE-stSENet	$\mathbf{0.604}_{\pm 0.004}$	$\mathbf{0.758}_{\pm 0.008}$	$0.574_{\pm 0.043}$	$0.588_{\pm 0.022}$
TIE-EEGNet	$0.561_{\pm 0.027}$	$0.655_{\pm 0.019}$	$\mathbf{0.575}_{\pm 0.036}$	$\mathbf{0.615}_{\pm 0.032}$

as Avg-EnK-EEGNet. We also implemented TIE-CE-stSENet and TIE-EEGWaveNet, which added a TIE module to the convolution layer of the multiScale temporal block in CE-stSENet and the multiscale convolution module in EEGWaveNet, respectively. When training the neural networks, we employed a weighted sampler to ensure each subtype has the same number of samples in each batch. All experiments used Adam optimizer with learning rate 0.001, batch size 64, weight decay 0.0001, cross-entropy loss, and early stopping with patience 30.

For traditional machine learning approaches, we took 3-level discrete wavelet transform using the *db*5 wavelets, computed the average and kurtosis of each component as features, and used them in SVM, LR, gradient boosting decision tree (GBDT), and ridge classifier (RC). 10 repeats of 3-fold cross-validation were performed, and the average results are reported in Table V.

TIE-EEGNet obtained the best BCA and F_1 score on CHSZ. With the help of the TIE module, TIE-CE-stSENet achieved the best BCA and F_1 score on TUSZ. The TIE module improved the performances of EEGNet, EEGWaveNet, and CE-stSENet on both datasets.

TIE-EEGNet consistently outperformed all EEGNet-based deep learning baselines on both datasets. By using the

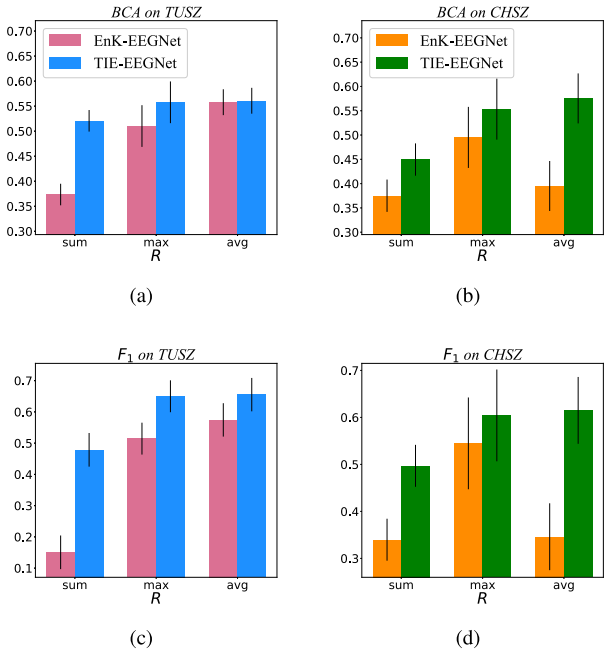

Fig. 6. Performances of TIE-EEGNet and EnK-EEGNet with different representation matrix R. (a) BCA on TUSZ; (b) BCA on CHSZ; (c) F_1 score on TUSZ; (d) F_1 score on CHSZ.

average representation matrix, Avg-EnK-EEGNet achieved large performance improvements over EnK-EEGNet, especially on TUSZ. Comparing with EEGNet, TIE-EEGNet had comparable or much smaller standard deviations. These results indicate that TIE-EEGNet is more accurate and robust than EEGNet and EnK-EEGNet.

E. Effectiveness of the Average Representation Matrix

To evaluate the effectiveness of the average representation matrix R in (7), experiments were conducted to compare the performances of TIE-EEGNet and EnK-EEGNet with different representation matrices: summation (R_{sum}), maximum (R_{max}), and average (R_{avg}). Fig. 6 shows the results.

For EnK-EEGNet, both R_{max} and R_{avg} outperformed R_{sum} on both datasets. Between the former two, R_{avg} outperformed

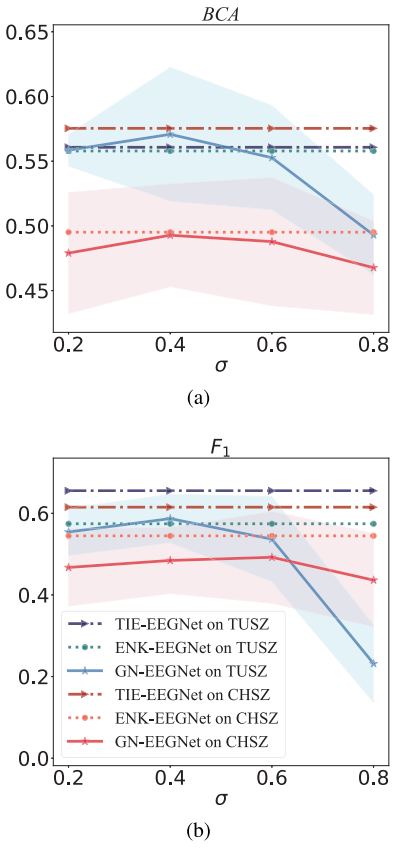

Fig. 7. Performances of TIE-EEGNet, Improved EnK-EEGNet, and GN-EEGNet. (a) BCA; (b) F_1 score.

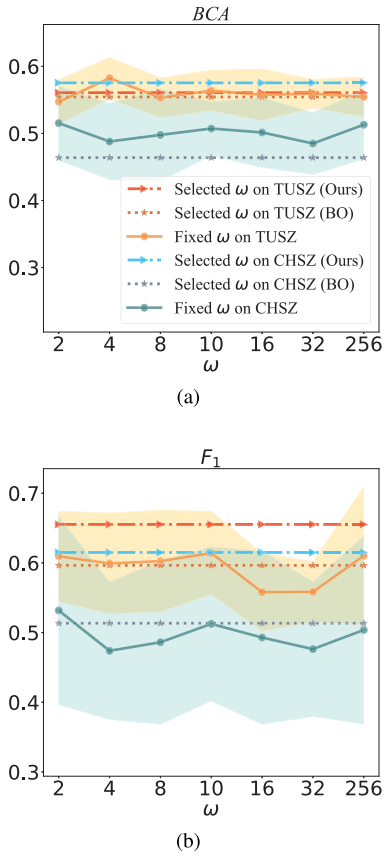

Fig. 8. Performances with selected and fixed ω. (a) BCA; (b) F_1 score.

R_{max} on TUSZ, but the opposite was observed on CHSZ. Anyway, changing R was effective for EnK-EEGNet, though it is difficult to conclude whether R_{max} or R_{avg} was better.

For TIE-EEGNet, again both R_{max} and R_{avg} outperformed R_{sum} on both datasets, and R_{avg} was always slightly better than R_{max}.

In summary, replacing R_{sum} in the original EnK-EEGNet with R_{avg} was effective.

F. Effectiveness of Sinusoidal Encoding

To further demonstrate the validity of sinusoidal encoding of TIE, we carried out an additional experiment, where the sinusoidal encoding was replaced by Gaussian noise. The corresponding model is denoted as GN-EEGNet. More specifically, Gaussian noise was added to the output of the first Conv2D layer of the standard EEGNet with a detached gradient that did not participate in training.

Fig. 7 compares the performances of GN-EEGNet, TIE-EEGNet and Improved EnK-EEGNet (R corresponding to the best result in the previous subsection was used). The horizontal axis represents the standard deviation (σ) of the Gaussian noise. Our proposed TIE-EEGNet always outperformed EnK-EEGNet. Additionally, TIE-EEGNet almost always outperformed GN-EEGNet by a large margin, except BCA on TUSZ when $\sigma = 0.4$. These results demonstrate that the sinusoidal encoding in TIE-EEGNet is indeed better than the linear encoding in EnK-EEGNet, and random Gaussian noise.

G. Effectiveness of Period Regulator Selection

This subsection validates the effectiveness of our strategy for selecting the period regulator ω in sinusoidal encoding of TIE-EEGNet.

TIE-EEGNet was trained with all candidate $\omega \in \{2, 4, 8, 10, 16, 32, 256\}$ in parallel, and the model with the smallest validation loss was selected. We also trained models with a fixed ω in $\{2, 4, 8, 10, 16, 32, 256\}$, and a selected ω by Bayesian Optimization (BO). Fig. 8 shows their performances. The BCA and F_1 score varied with ω, and there was not a single best ω. The ω selection approach in TIE-EEGNet almost always outperformed a fixed ω, or 20 BO iterations for searching $\omega \in [2, 256]$. Furthermore, our proposed ω selection approach only needed to evaluate 7 ω candidates, so the computational cost was low.

H. Transfer Learning

A common assumption in machine learning is that the distributions of the training and test data are identical or similar, which may not hold in practice. For example, the CHSZ dataset was acquired from children and infants, whereas the TUSZ dataset covered patients of all ages. A seizure subtype classification model trained on TUSZ may perform poorly on CHSZ, due to the significant differences between adults' and children's EEG signals. In the ideal case, we should train a separate model for each patient group. However, this

TABLE VI
CLASSIFICATION ACCURACIES AND THE IMPROVEMENTS WHEN TRANSFERRING FROM TUSZ TO CHSZ

	BCA		F_1	
	Result	Imp.*	Result	Imp.*
EEGWaveNet	$0.521_{\pm 0.039}$	0.027	$0.518_{\pm 0.049}$	0.061
CE-stSENet	$0.617_{\pm 0.020}$	0.050	$0.563_{\pm 0.019}$	0.024
EEGNet	$0.542_{\pm 0.070}$	0.060	$0.580_{\pm 0.118}$	0.290
EnK-EEGNet	$0.374_{\pm 0.026}$	-0.001	$0.339_{\pm 0.072}$	-0.001
TIE-EEGWaveNet	$0.536_{\pm 0.050}$	0.027	$0.518_{\pm 0.039}$	0.060
TIE-CE-stSENet	$\mathbf{0.647}_{\pm 0.014}$	0.073	$0.613_{\pm 0.025}$	0.026
TIE-EEGNet (R_{max})	$0.581_{\pm 0.042}$	0.027	$0.641_{\pm 0.070}$	0.037
TIE-EEGNet (R_{avg})	$0.583_{\pm 0.028}$	0.007	$\mathbf{0.672}_{\pm 0.049}$	0.057

* indicates the performance improvements when comparing the fine-tuned and non-fine-tuned models on CHSZ.

requires a large amount of labeled data for each patient group, which is expensive and time-consuming to collect.

A promising solution to the above problem is transfer learning [50], [51], which uses data/knowledge in some auxiliary domains (source domains) to help the learning in a new domain (target domain). There are many different implementations of transfer learning [52]. We employed fine-tuning to transfer the model pre-trained on TUSZ to CHSZ.

Specifically, a seizure subtype classification model was first trained on TUSZ, with data from all subjects combined. Then, the model was fine-tuned on a subset of data from CHSZ for adaptation. In this way, we can utilize information from TUSZ to expedite the training on CHSZ.

Table VI shows the test results on CHSZ. The improvements were computed against the corresponding results in Table V. Clearly, TIE-EEGNet and TIE-CE-stSENet achieved much better performances in transfer learning, especially when compared with EnK-EEGNet. Additionally, using transfer learning resulted in better performance than not using transfer learning. These results again demonstrated the advantages of our proposed TIE module.

V. CONCLUSION

This paper has proposed an EEGNet-based slim deep neural network for EEG-based seizure subtype classification. A temporal information enhancement module with sinusoidal encoding is used to augment the first convolution layer of EEGNet. A training strategy for automatic hyper-parameter (period regulator) selection is also proposed. Experiments on the public TUSZ dataset and our own CHSZ dataset with infants and children demonstrated that our proposed TIE-EEGNet outperformed several traditional and deep learning models in cross-subject seizure subtype classification. Additionally, it also achieved the best performance in a challenging transfer learning scenario, i.e., from TUSZ to CHSZ.

There are two practical considerations in using the TIE module. First, since it uses periodic sinusoidal positional encoding, the TIE module is more suitable for tasks relying on temporal information, e.g., epileptic seizure detection/classification, and sleep stage classification. Second, the TIE module should be applied to the convolution layers that extract temporal (instead of spatial) information from the raw EEG signals.

The following directions will be considered in our future research:

1) TIE focuses on enhancing the learning of temporal information. However, the location of epilepsy, especially the onset region, and the transmission mode, are also important in indicating the epilepsy subtype. Therefore, incorporating the spatial information may further improve the classification performance.
2) Many informative features for seizure classification have been proposed in the literature [17]. Combining these expert-designed features with those automatically extracted in TIE-EEGNet may result in better seizure classification performance.
3) A very simple fine-tuning transfer learning approach was used in this paper. However, there are many other more sophisticated approaches [52], e.g., instance/feature/knowledge transfer, which could be used to further enhance the transfer learning performance. Additionally, it has been shown that data augmentation [53] and data alignment [54], [55] can significantly improve the transfer learning performance in other paradigms of BCIs, e.g., motor imagery. Similar approaches may also be developed for seizure classification.
4) It is interesting to study if TIE-EEGNet can also be extended to other BCI paradigms, e.g., motor imagery [14], tactile P300 event-related potential classification [56], and P300-based BCI speller [57].

REFERENCES

[1] J. R. Stevens, B. L. Lonsbury, and S. L. Goel, "Seizure occurrence and interspike interval: Telemetered electroencephalogram studies," *Arch. Neurol.*, vol. 26, no. 5, pp. 409–419, 1972.
[2] A. J. Casson, D. C. Yates, S. J. M. Smith, J. S. Duncan, and E. Rodriguez-Villegas, "Wearable electroencephalography," *IEEE Eng. Med. Biol. Mag.*, vol. 29, no. 3, pp. 44–56, May/Jun. 2010.
[3] J. E. Motelow and H. Blumenfeld, "Functional neuroimaging of spike-wave seizures," *Methods Mol. Biol.*, vol. 489, pp. 189–209, Mar. 2009.
[4] S. Tang *et al.*, "Self-supervised graph neural networks for improved electroencephalographic seizure analysis," in *Proc. Int. Conf. Learn. Represent.*, 2022, pp. 1–23.
[5] I. E. Scheffer *et al.*, "ILAE classification of the epilepsies: Position paper of the ILAE commission for classification and terminology," *Epilepsia*, vol. 58, no. 4, pp. 512–521, Apr. 2017.
[6] R. S. Fisher *et al.*, "Operational classification of seizure types by the International League Against Epilepsy: Position paper of the ILAE commission for classification and terminology," *Epilepsia*, vol. 58, no. 4, pp. 522–530, 2017.
[7] V. P. Oikonomou, A. T. Tzallas, and D. I. Fotiadis, "A Kalman filter based methodology for EEG spike enhancement," *Comput. Methods Programs Biomed.*, vol. 85, no. 2, pp. 101–108, Feb. 2007.
[8] N. Ille, P. Berg, and M. Scherg, "Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies," *J. Clin. Neurophysiol.*, vol. 19, no. 2, pp. 113–124, 2002.
[9] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, Sep. 1970.
[10] S. Li, W. Zhou, Q. Yuan, S. Geng, and D. Cai, "Feature extraction and recognition of ictal EEG using EMD and SVM," *Comput. Biol. Med.*, vol. 43, no. 7, pp. 807–816, Aug. 2013.
[11] H. R. Mohseni, A. Maghsoudi, and M. B. Shamsollahi, "Seizure detection in EEG signals: A comparison of different approaches," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, New York, NY, USA, Aug./Sep. 2006, pp. 6724–6727.
[12] K. Samiee, P. Kovács, and M. Gabbouj, "Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 541–552, Feb. 2015.

[13] D. Conigliaro, P. Manganotti, and G. Menegaz, "Multiscale sample entropy for time resolved epileptic seizure detection and fingerprinting," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Florence, Italy, May 2014, pp. 3582–3585.

[14] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and Dempster–Shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4814–4825, Nov. 2020.

[15] S. Chandaka, A. Chatterjee, and S. Munshi, "Cross-correlation aided support vector machine classifier for classification of EEG signals," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1329–1336, Mar. 2009.

[16] A. R. Naghsh-Nilchi and M. Aghashahi, "Epilepsy seizure detection using eigen-system spectral estimation and multiple layer perceptron neural network," *Biomed. Signal Process. Control*, vol. 5, no. 2, pp. 147–157, Apr. 2010.

[17] R. Peng, J. Jiang, G. Kuang, H. Du, D. Wu, and J. Shao, "EEG-based automatic epilepsy detection: Review and outlook," *Acta Automatica Sinica*, vol. 48, pp. 335–350, Apr. 2022.

[18] J. Cao, J. Zhu, W. Hu, and A. Kummert, "Epileptic signal classification with deep EEG features by stacked CNNs," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 4, pp. 709–722, Dec. 2020.

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

[20] Z. Wei, J. Zou, J. Zhang, and J. Xu, "Automatic epileptic EEG detection using convolutional neural network with improvements in time-domain," *Biomed. Signal Process. Control*, vol. 53, Aug. 2019, Art. no. 101551.

[21] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*.

[22] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[23] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 6848–6856.

[24] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.

[25] W. Park, D. Kim, Y. Lu, and M. Cho, "Relational knowledge distillation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 3967–3976.

[26] Y. Wang *et al.*, "Pruning from scratch," in *Proc. 34th AAAI Conf. Artif. Intell.*, New York, NY, USA, Feb. 2020, vol. 34, no. 7, pp. 12273–12280.

[27] X. Dong and Y. Yang, "Network pruning via transformable architecture search," in *Proc. 33rd Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2019, pp. 759–770.

[28] S. Lin, R. Ji, C. Chen, D. Tao, and J. Luo, "Holistic CNN compression via low-rank decomposition with knowledge transfer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 2889–2905, Dec. 2019.

[29] A. Emami, N. Kunii, T. Matsuo, T. Shinozaki, K. Kawai, and H. Takahashi, "Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images," *NeuroImage: Clin.*, vol. 22, Jan. 2019, Art. no. 101684.

[30] A. Shoeibi *et al.*, "Epileptic seizures detection using deep learning techniques: A review," *Int. J. Environ. Res. Public Health*, vol. 18, no. 11, p. 5780, May 2021.

[31] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, p. 056013, 2018.

[32] S. Roy, U. Asif, J. Tang, and S. Harrer, "Seizure type classification using EEG signals and machine learning: Setting a benchmark," 2019, *arXiv:1902.01012*.

[33] I. R. D. Saputro, N. D. Maryati, S. R. Solihati, I. Wijayanto, S. Hadiyoso, and R. Patmasari, "Seizure type classification on EEG signal using support vector machine," in *Proc. Int. Conf. Electron. Represent. Algorithm*, Yogyakarta, Indonesia, Jan. 2019, p. 012065.

[34] U. Asif, S. Roy, J. Tang, and S. Harrer, "SeizureNet: Multi-spectral deep feature learning for seizure type classification," 2019, *arXiv:1903.03232*.

[35] 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.

[36] B. J. Lance, S. E. Kerick, A. J. Ries, K. S. Oie, and K. McDowell, "Brain–computer interface technologies in the coming decades," *Proc. IEEE*, vol. 100, no. 3, pp. 1585–1599, May 2012.

[37] X. Ding, Y. Guo, G. Ding, and J. Han, "AcNet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks," in *Proc. IEEE Interface Conf. Comput. Vis. (ICCV)*, Long Beach, CA, USA, Oct. 2019, pp. 1911–1920.

[38] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, May 2018, pp. 7132–7141.

[39] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1251–1258.

[40] S. Bai, J. Zico Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[41] W. Ko, E. Jeon, S. Jeong, and H.-I. Suk, "Multi-scale neural network for EEG representation learning in BCI," *IEEE Comput. Intell. Mag.*, vol. 16, no. 2, pp. 31–45, May 2021.

[42] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 3367–3375.

[43] D. Zhang, L. Yao, K. Chen, and J. Monaghan, "A convolutional recurrent attention model for subject-independent EEG signal analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 715–719, May 2019.

[44] A. Kumar Singh and C.-T. Lin, "EnK: Encoding time-information in convolution," 2020, *arXiv:2006.04198*.

[45] A. Vaswani *et al.*, "Attention is all you need," in *Proc. 31st Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, vol. 30, Dec. 2017, pp. 5998–6008.

[46] W. O. Tatum, "Ellen r. grass lecture: Extraordinary EEG," *Neurodiagnostic J.*, vol. 54, no. 1, pp. 3–21, 2014.

[47] A. L. Anderson and M. E. Thomason, "Functional plasticity before the cradle: A review of neural functional imaging in the human fetus," *Neurosci. Biobehav. Rev.*, vol. 37, no. 9, pp. 2220–2232, Nov. 2013.

[48] V. Shah, "The Temple University hospital seizure detection corpus," *Frontiers Neuroinform.*, vol. 12, p. 83, Nov. 2018.

[49] P. Thuwajit *et al.*, "EEGWaveNet: Multiscale CNN-based spatiotemporal feature extraction for EEG seizure detection," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5547–5557, Aug. 2022.

[50] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain–computer interfaces: A review of progress made since 2016," *IEEE Trans. Cogn. Developmental Syst.*, vol. 14, no. 1, pp. 4–19, Mar. 2022.

[51] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[52] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Jan. 2009.

[53] K. Xia, L. Deng, W. Duch, and D. Wu, "Privacy-preserving domain adaptation for motor imagery-based brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, early access, Apr. 19, 2022, doi: 10.1109/TBME.2022.3168570.

[54] H. He and D. Wu, "Transfer learning for brain–computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, Feb. 2020.

[55] H. He and D. Wu, "Different set domain adaptation for brain-computer interfaces: A label alignment approach," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1091–1108, May 2020.

[56] J. Jin, Z. Chen, R. Xu, Y. Miao, X. Wang, and T.-P. Jung, "Developing a novel tactile P300 brain-computer interface with a cheeks-stim paradigm," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 9, pp. 2585–2593, Sep. 2020.

[57] Y. Yu, Y. Liu, E. Yin, J. Jiang, Z. Zhou, and D. Hu, "An asynchronous hybrid spelling approach based on EEG–EOG signals for Chinese character input," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1292–1302, Jun. 2019.