DSMN-ESS: Dual-Stream Multitask Network for Epilepsy Syndrome Classification and Seizure Detection

Jiuwen Cao, *Senior Member, IEEE*, Yaohui Chen, Runze Zheng, Xiaonan Cui, Tiejia Jiang, and Feng Gao

***Abstract*— Simultaneous childhood epilepsy syndrome classification (ESC) and seizure detection (SD) are both significant in epilepsy analysis. Current research mainly focuses on a single task, mostly on SD. In this article, a novel dual-stream multitask network (DSMN) exploiting multichannel scalp electroencephalograms (EEGs) is developed to simultaneously perform ESC-Task and SD-Task, in short as DSMN-ESS. The close correlation between ESC-Task and SD-Task is explored to achieve better performance. To improve the performance, an information-sharing gate module is designed in DSMN to enable both tasks to fully obtain useful information. Meanwhile, a channel weight update module is developed to well-extract the internal spatial relationship between multichannel EEGs. Furthermore, an area-under-the-curve (AUC)-based loss is proposed to address the data imbalance issue in epilepsy analysis. Studies on EEG data recorded 49 patients from the Children's Hospital, Zhejiang University School of Medicine (CHZU), are carried out to show the effectiveness of DSMN-ESS. The results show that DSMN-ESS can achieve the highest AUC, 99.95% and 99.78% in ESC-Task and SD-Task, respectively, which are superior to several state-of-the-art (SOTA) methods.**

***Index Terms*— Data imbalance, epilepsy syndrome classification (ESC), multitask, seizure detection (SD).**

I. Introduction

EPILEPSY is characterized by repeated sudden seizures. Symptoms may include a brief pause in consciousness, violent convulsions, and sometimes even unconsciousness [1]. Epilepsy warning systems enable to take emergency preventive strategies to reduce the potential harm. Childhood epileptic syndromes, which may include multiple types of seizures, are defined by a group of epileptic characteristics occurring together. Around 0.5%–1% children suffered from epileptic syndromes. Known epileptic syndromes include early infantile epileptic encephalopathy (EIEE) [2], infantile spasms (called WEST syndrome) [3], [4], childhood absence epilepsy (CAE) [5], febrile seizures plus (FS+) [6], benign childhood epilepsy with centrotemporal spikes (BECT) [7], and epileptic encephalopathy with continuous spike-and-wave in sleep (CSWS) [8]. Accurate epilepsy syndrome classification (ESC) and seizure detection (SD) are both crucial [9].

Scalp electroencephalogram (EEG) is the most convenient and effective way in brain function characterization and brain disease analysis [10], [11], [12], [13], [14], [15]. In recent years, machine learning and deep learning technology have been widely applied in SD and ESC using scalp EEGs. For SD, for instance, support vector machine (SVM) and naive Bayesian classifiers with hand-crafted EEG features in [16], SVM with EEG feature optimization in [17], ResNet deep learning model with statistical analysis based EEG features in [18], 3-D convolutional networks in [19], [20], and [21], and transfer learning in [22] and [23] have been presented. In addition, the graph network is used to explore the spatio-temporal dynamics of connections between brain regions [24], [25], capsule network captures spatial information of EEG and allows information transfer between capsules through a dynamic routing mechanism for superior performance [26]. The ability of the transformer to calculate attention between input signal channels in SD has been presented in [27]. For childhood ESC, recent studies on scalp regional EEG pattern analysis [28], deep feature fusion [29], 3-D residual-attention-deep-network [30], and two-stream 3-D deep network [31] have been conducted.

A strong correlation is well known to exist between epilepsy syndromes and seizures. But existing studies generally focused on a single task, either on epilepsy seizure analysis or ESC. Few studies have been presented to jointly explore both two tasks. In fact, different epilepsy syndromes may have different symptoms, including seizure duration, seizure onset frequency, EEG waveform characteristics, and difference between ictal and interictal EEGs. Moreover, for a patient suffering from certain epilepsy syndrome, there may be different seizure types

Manuscript received 19 March 2023; revised 30 July 2023; accepted 8 August 2023. Date of publication 23 August 2023; date of current version 4 September 2023. This work was supported in part by the National Natural Science Foundation of China under Grant U1909209, in part by the National Key Research and Development of China under Grant 2021YFE0100100, in part by the National Key Research and Development Program of China under Grant 2021YFE0205400, in part by the Natural Science Key Foundation of Zhejiang Province under Grant LZ22F030002, and in part by the Research Funding of Education of Zhejiang Province under Grant GK228810299201. The Associate Editor coordinating the review process was Dr. Xun Chen. *(Corresponding author: Jiuwen Cao.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Children's Hospital, School of Medicine, Zhejiang University, and Registered in Chinese Clinical Trail Registry under Application No. ChiCTR1900028804.

Jiuwen Cao, Yaohui Chen, Runze Zheng, and Xiaonan Cui are with the Machine Learning and I-Health International Cooperation Base of Zhejiang Province and the Artificial Intelligence Institute, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China (e-mail: jwcao@hdu.edu.cn; 1944168019@qq.com; runzewuyu@hdu.edu.cn; xncui@hdu.edu.cn).

Tiejia Jiang and Feng Gao are with the Department of Neurology, Children's Hospital, School of Medicine, National Clinical Research Center for Child Health, Zhejiang University, Hangzhou 310003, China (e-mail: jiangyouze@zju.edu.cn; epilepsy@zju.edu.cn).

Digital Object Identifier 10.1109/TIM.2023.3307724

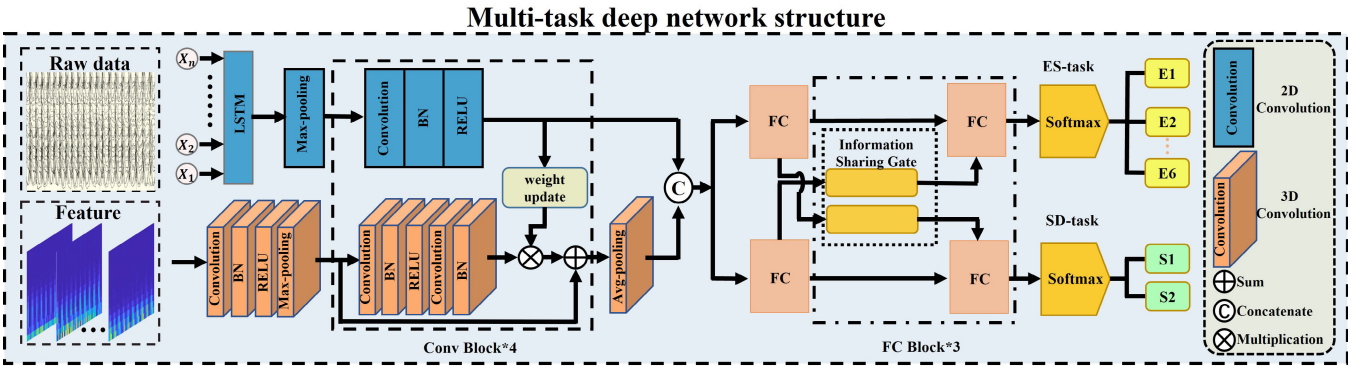


Fig. 1. Network structure of the proposed DSMN-ESS.

and seizure onset properties. Thus, multitask learning on joint ESC and seizure onset detection is crucial, leading to effective epilepsy diagnosis and precise treatment.

Over the past years, multitask learning has been widely used in many fields, and effectiveness is also demonstrated in biomedical engineering applications [32], [33]. For instance, Nie et al. [34] developed a multitask learning method to effectively utilize gender detection to improve microexpressions recognition accuracy. Miranda-Correa and Patras [35] proposed a multitask cascaded deep neural network-based people's emotional level prediction and personal factors analysis framework. Ma et al. [36] proposed the architecture of long-short-term-memory network-based (LSTM-based) multitask learning for seizure prediction and latency regression, where an average prediction accuracy of 89.36% can be achieved. Inspired by the multitask deep networks, we construct a dual-stream multitask network (DSMN) for joint childhood ESC and SD (DSMN-ESS) in this article. The detailed network architecture is shown in Fig. 1. DSMN-ESS applies 3-D convolutions to perform temporal, frequency, and spatial feature analysis on multichannel EEGs for ESC and SD. Particularly, as in Fig. 1, the time-domain EEG features are extracted through an LSTM, which will then be applied to EEG channel weight updating and deep feature fusion. The channel weight update module is developed to study the mapping relationship between EEG epilepsy features and brain regions. Moreover, an information-sharing gate is designed in the model to link the ESC and SD tasks, which effectively shares the features of the two independent tasks to make complementary learning. To address the natural imbalance issue that existed in epilepsy interictal and ictal periods, as well as in different epilepsy syndromes, a novel loss function that can approximately maximize the area under the ROC curve (AUC) is developed to minimize the adverse effects of an imbalanced dataset. The main contributions of the article can be summarized as follows.

1) A novel multitask learning model named DSMN-ESS is developed for joint childhood ESC and SD, where one stream of DSMN-ESS adopts LSTM and 2-D convolutional neural network (CNN) to explore the time and spatial characteristics of multichannel EEGs, and another stream employs 3-D CNN to learn the temporal-spectral-spatial information on EEG spectrum.
2) A weight update module is designed to study the highly correlated EEGs to different epilepsy syndromes, which can help to understand the internal connection between EEG channels and epilepsy from the perspective of brain regions, and further to remove redundant EEGs. An information-sharing gate is developed to share useful information between ESC and SD to benefit both tasks.
3) To address the natural imbalance issue between interictal and ictal in SD, an AUC loss function is developed to minimize the adverse effects by maximizing the AUC score. Particularly, a novel objective function that is differentiable, approximates to the Wilcoxon-Mann–Whitney statistic, and equals AUC, is adopted in DSMN-ESS.

II. Methodology

The proposed DSMN-ESS contains two stream pipelines learning on the filtered multichannel EEGs and spectrogram, respectively. Overall: 1) the LSTM is applied on the filtered EEGs of each channel to extract time characteristics, and the spatial information of 21 channel EEGs is extracted by a 2-D CNN; 2) a 3-D CNN is constructed to extract deep features from the spectrogram of 21 channel EEGs; 3) the extracted time-/space-domain EEG features by LSTM and CNN are adopted to update the weights of EEG channels; and 4) the fused features are finally input into the full connection module with information sharing mechanism to achieve ESC and SD. The proposed DSMN-ESS will be introduced in detail in the following.

A. Childhood Epilepsy Syndrome

There are around 20 known childhood epilepsy syndromes, where different epilepsy syndromes have different characteristics. EIEE mainly manifests as tonic-spasmodic seizures in early infancy with severe psychomotor disturbances, and the patient has a structural brain injury in some cases. The EEGs of EIEE are generally characterized by a suppression-burst (S-B) pattern [37]. The interictal EEG of WEST can be highly irregular, and the ictal EEG shows high-amplitude slow waves followed by 1 to several seconds of voltage decay or low-amplitude fast-wave activity [38]. Disturbed consciousness, loss of consciousness, and cessation of behavior are fundamental characteristics of CAE patients. The ictal EEG pattern of CAE is characterized by high-amplitude, bisynchronous, and symmetrical rhythmic 3 Hz sharp-slow wave

compound discharges [5]. FS+ is a familial hereditary epilepsy syndrome with onset mainly in childhood and adolescence. The EEG varies greatly with different types of seizures [6]. BECT is an idiopathic partial epilepsy with high genetic susceptibility and is most common in childhood. The interictal EEG pattern is high-voltage focal or multifocal sharp waves predominantly in the centrotemporal region, often followed by spike slow complex waves during sleep [39]. The most prominent features of CSWS are electrical persistence during sleep and impairment of higher cortical functions. The peak age of onset is 5–7 years old, more common in boys [8]. We conduct the study on long-term recorded EEGs of 49 childhood patients suffering from the above six typical epilepsy syndromes, respectively. The overall EEG duration used for analysis is more than 288 h length.

B. EEG Preprocessing and Representation

For 21 channel EEGs, the regular preprocessing is performed, there include a notch filter to eliminate 50 Hz power frequency interference, and a bandpass filter from 1 to 70 Hz to select EEGs containing most information. Besides learning on original EEGs, the spectrogram containing the instantaneous frequency information [40] is studied for representation. Short-time Fourier transform (STFT) is a common method to convert time-domain signal to time-frequency domain, which is used to obtain EEG spectrogram as follows:

$$X(t, f) = \int_{-\infty}^{\infty} x(\tau) h(t-\tau) e^{-j2\pi f \tau} d\tau \tag{1}$$

where $X(t, f)$ is the spectrum of the signal at time t and $h(t-\tau)$ is the Hamming window function.

We use the Shannon function to measure the spectrum power distribution. The spectrum entropy can reflect the power distribution after EEG conversion, calculated as

$$\begin{aligned} X(f) &= \sum_{i=1}^{t} X(i, f) \\ \hat{X}(f) &= \frac{X(f)}{\sum X(f)} \\ SE &= -\sum \hat{X}(f) \log(\hat{X}(f)) \end{aligned} \tag{2}$$

where $X(f)$ represents the normalized spectrum. The Shannon algorithm is then applied to get the spectrum entropy (SE). Fig. 2 depicts the spectrum entropy distributions of different epilepsy syndromes in interictal and seizure periods, respectively. As observed, SE of different epilepsy syndromes has various distribution, the same observation can be found between interictal and seizure periods.

The spectrogram is further computed using a 4-s EEG segment with 50% overlapping. It is resized to 32×32 to fit our proposed network. Finally, the spectrogram of all 21 channels is combined into the final 3D feature with the size of $21 \times 32 \times 32$ (channel $\times$ time $\times$ frequency).

C. LSTM Network

LSTM is widely used in time series learning and has also been studied in seizure pattern learning [41]. LSTM can effectively extract the EEG features over time. The unit of LSTM usually includes the input gate, output gate, and forget gate. The unit is similar to a memory machine that constantly remembers useful information, and forgets useless information. In LSTM unit, x_{tra}^{t} is the input of the current layer, h_{tra}^{t-1} is the output of the hidden node of the previous layer. As shown in (3), the function of the forget gate is to determine the information retention degree of the cell state C^t of the previous layer. The input gate is responsible for updating the information selectively, $\tilde{C}^t$ is the memory cell candidate value, as shown in (4) and (5). C^t is a cell state of long-term memory to carry memory information, as shown in (6). The output information is controlled by the output gate, as shown in (7)

$$f^t = \sigma\left(W_f \cdot \left[h_{\text{tra}}^{t-1}, x_{\text{tra}}^{t}\right] + b_f\right) \tag{3}$$
$$i^t = \sigma\left(W_i \cdot \left[h_{\text{tra}}^{t-1}, x_{\text{tra}}^{t}\right] + b_i\right) \tag{4}$$
$$\tilde{C}^t = \tanh\left(W_C \cdot \left[h_{\text{tra}}^{t-1}, x_{\text{tra}}^{t}\right] + b_C\right) \tag{5}$$
$$C^t = f^t * C^{t-1} + i^t * \tilde{C}^t \tag{6}$$
$$o^t = \sigma\left(W_o \cdot \left[h_{\text{tra}}^{t-1}, x_{\text{tra}}^{t}\right] + b_o\right) \tag{7}$$
$$h_{\text{tra}}^{t} = o^t * \tanh\left(C^t\right) \tag{8}$$

where W and b correspond to weight parameters and offset, respectively, σ is the Sigmoid function.

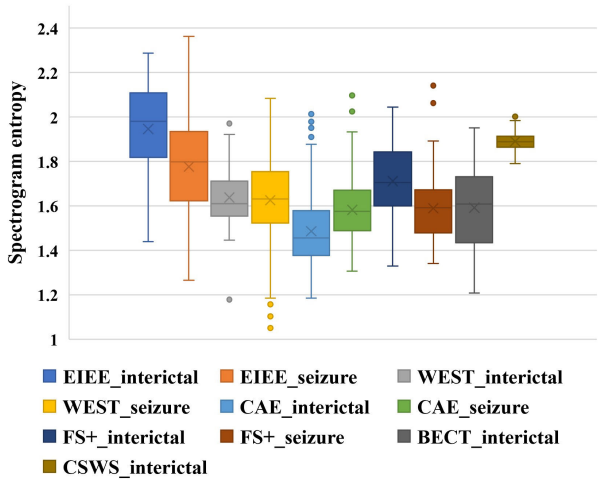


Fig. 2. Distribution of the mean spectrum entropy of 21 channels for different periods of different epilepsy syndromes.

D. Convolution Network Architecture

CNN achieved great success in epileptic EEG analysis and SD in the past [22], [30], [42], [43], [44], [45], [46], [47]. Since EEG contains rich temporal, spectral, and spatial features, we design a 3-D CNN to fully learn the temporal–spectral–spatial information on the spectrum of multichannel EEGs. At the same time, a 2-D CNN is applied to extract the temporal-spatial information of EEG features taken by LSTM.

For the 2-D CNN, there contains convolution layer, pooling layer, and activation layer. The feature map in the convolution layer is obtained by convolution among multiple adjacent location points in the previous layer. Particularly, the

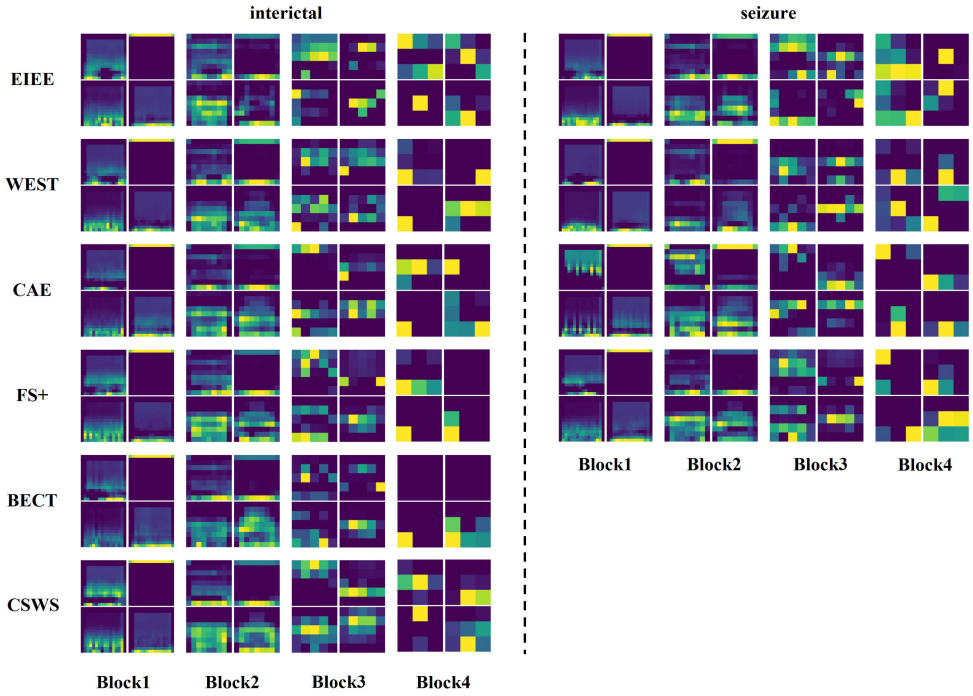


Fig. 3. Visualization of the feature map of the four 3-D convolution block learned from the spectrogram.

jth feature map of the ith layer at position (x, y), denoted as f_{ij}^{xy}, is obtained by

$$f_{ij}^{xy} = \tanh\left(b_{ij} + \sum_{c}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1} w_{ijc}^{pq} f_{(i-1)c}^{(x+p)(y+q)}\right) \tag{9}$$

where b_{ij} is the bias, w_{ijc}^{pq} is the weight of convolution kernel at the (p, q) position connected to the cth feature map in the previous layer, P_i and Q_i are the length and width of the convolution kernel, respectively.

For the 3-D CNN, the convolution operation is similar to 2-D CNN, expressed as

$$f_{ij}^{xyz} = \tanh\left(b_{ij} + \sum_{c}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1} w_{ijc}^{pqr} f_{(i-1)c}^{(x+p)(y+q)(z+r)}\right) \tag{10}$$

where f_{ij}^{xyz} represents the jth feature map of the ith later at position (x, y, z). Compared with the 2-D convolution kernel, the 3-D convolution kernel adds the height dimension R_i. Similar to conventional deep networks, the max pooling operation is further adapted to filter out redundant information as well as to reduce network size. Therefore, the network learning can be accelerated and the overfitting issue can be effectively avoided.

Fig. 3 visualizes the feature maps extracted by the 3-D convolution blocks on different epilepsy syndromes in interictal and seizure periods. It can be seen that when the layers become deepen, the extracted features become more abstract. It is possible to extract the contour features of the input image from the beginning, and to extract more details later, such as information about a certain brain region or a certain frequency band. So, the deeper the number of layers, the less it represents the visual and intuitive content of the image, and the more information about the category. In initial blocks, the kernels can learn the global feature information, where the extracted feature is similar to the original input. From the extracted feature on different epilepsy period, obvious differences between interictal and seizure EEG frames of CAE can be found, while other syndromes are not obvious. After multilayer convolution, some local information in the spectrogram is enlarged, significant differences between interictal and seizure EEG frames can be found in the features taken by most blocks. For six different epilepsy syndromes, obvious feature differences in different layers can be observed.

E. Weight Update Module and Information Sharing Gate

Early epileptic encephalopathies are related to the dynamic characteristics of brain network [48]. Different scalp regions, such as temporal and occipital regions, occupy different weights in ESC as well as SD. Therefore, the effect of EEG epilepsy recognition will be affected by the enhancement or inhibition of a certain scalp region. The distribution of importance between channels can reveal brain network patterns in different time ranges, which can provide important information for understanding the neurobiological process behind EEG phenotype and the whole brain dysfunction of epilepsy syndrome. The weight map of different EEG channels can be learned through the weight update module, as shown in Fig. 4. Assume the size of 2-D input matrix $X = [x_1, x_2, \ldots, x_h]$, $x_h \in \mathbb{R}^{1\times w}$ is the time feature of the hth EEG channel. The size of 3-D input matrix Y is $C \times D \times H \times W$ (RGB channels, EEG channels, time, and frequency), the weight update module aims to squeeze the time dimension information to obtain the

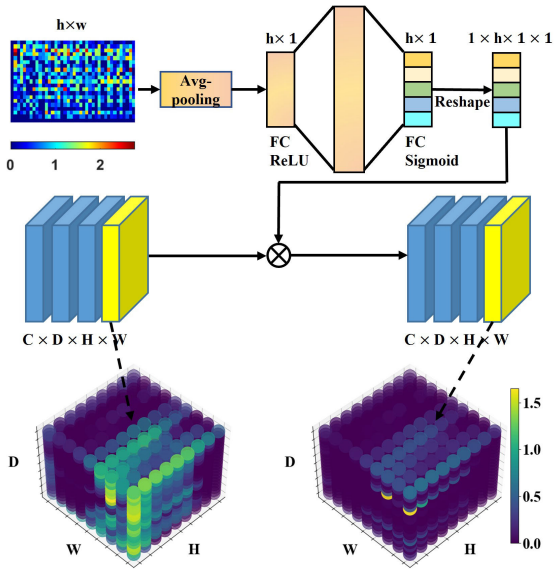


Fig. 4. Weight update module: The EEG channel (D) dimension of 3-D feature is updated, where the module aims to select EEG channels to suppress useless information.

initial weight vector $z \in \mathbb{R}^{1\times h}$, the hth EEG channel initial weight z_h is calculated by

$$z_h = \text{Avgpool}(x_h) = \frac{1}{1 \times w} \sum_{i=1}^{w} x_h(i). \tag{11}$$

Then, the channel dimension is amplified by s times for channel information learning, where s is the expansion rate. However, along with an increment of s, the computing cost will increase. To balance the performance and the network complexity, we set s to be 16. Finally, the value range of the channel weight vector is mapped to [0, 1] through a Sigmoid and the 3-D feature is updated to $\hat{Y}$ through multiplication by

$$\begin{aligned} r &= \sigma(W_2\delta(W_1 z)) \\ \hat{Y} &= Y \cdot r \end{aligned} \tag{12}$$

where δ and σ are the ReLU and Sigmoid functions, respectively, W_1 is the weight parameter of the previous fully connected layer.

To better understand the feature selection mechanism, Fig. 5 shows the weight W_2 of the second full connection layer, which can be used to verify the weight update module in the first convolution block. Each small grid in Fig. 5 represents a weight from the hidden layer to h channels. The redder the grid, the higher the weight. It can be seen that some weights are quite different and some are slightly different. It verifies that the layer can provide a very flexible function combination for each channel, which can comprehensively understand EEGs. Furthermore, the feature distribution is visualized in a 2-D space using T-SNE in Fig. 6. It can be seen that the shared features of the two tasks can be better distinguished after adopting the channel weight update module. The improvement is more obvious in the ESC-Task. The visualizations validate that using the weight update module can effectively improve the performance, which is of great significance for the subsequent analysis of the brain network.

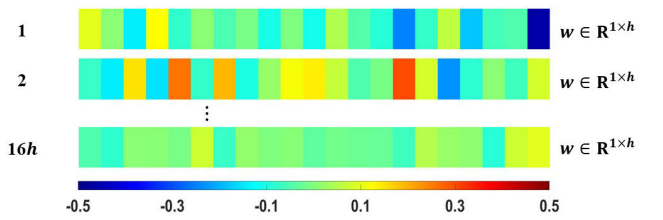


Fig. 5. Visualization of the weight matrix of the second full connection layer.

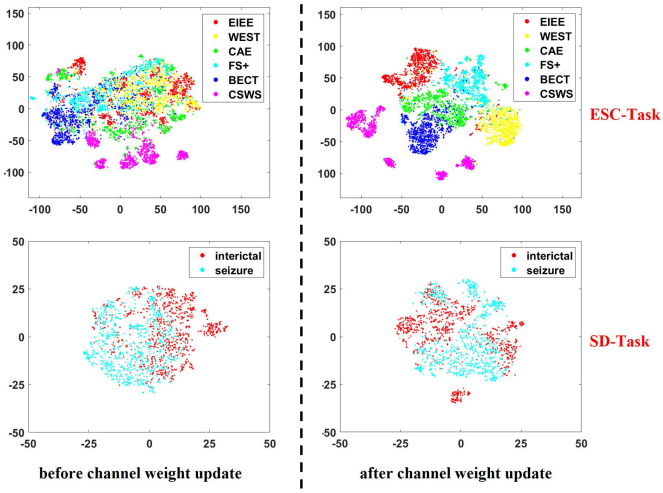


Fig. 6. Visualization of network features before and after using the channel weight update module.

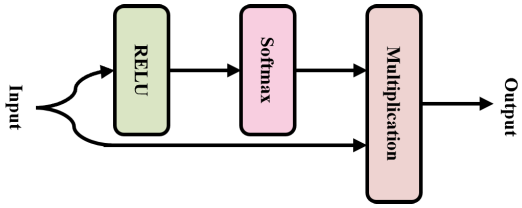


Fig. 7. Information sharing gate module.

In multitask learning, the information flow between different tasks may benefit each task [49]. Considering that it is useful to share information between ESC and SD, we design an information-sharing gate in the proposed DSMN-ESS. In general, the connectivity between two tasks should be appropriate, neither too strong nor too weak. Too weak connectivity means no information flow between two tasks. Although a strong connectivity can result in high structural similarity of all tasks, it may easily lead to optimization conflicts. In [50], three gate mechanisms are discussed, whose function is to control the information flow between tasks.

Fig. 7 shows the structure of the information-sharing gate designed in this article. There includes an activation layer, a softmax layer, and a multiplication operation. The ReLU function δ is to suppress useless information, and the softmax function is adopted to construct a selection gate as

$$\begin{aligned} m_{t1} &= \text{softmax}(\sigma(x_{t1})) \cdot x_{t1} \\ h_{t2} &= x_{t2} + m_{t1} \end{aligned} \tag{13}$$

where x_{t1} and x_{t2} are the features of the ESC-Task and SD-Task, respectively, m_{t1} is the useful information passed from ESC-Task to SD-Task. h_{t2} is the fused information, used for the input of the next full connection layer of the SD-Task.

F. Loss Function

Due to the short duration of the ictal period, an imbalance issue always exists in epilepsy SD [44], which seriously affects the minority classification in the SD-Task. To address this issue, we recur to using novel loss functions for the task. Cross entropy is a commonly used classification loss function, but it is not enough to solve the data imbalance problem. Alternatively, compared with cross entropy, AUC is a robust evaluation indicator for balanced and imbalanced data. In this article, we aim to maximize AUC to deal with the imbalance issue in SD-Task. The normalized Wilcoxon–Mann–Whitney (WMW) statistic [51], which is proved to be exactly equal to AUC, is calculated as follows:

$$\begin{aligned} \text{WMW} &= \frac{\sum_{i=0}^{P-1}\sum_{j=0}^{N-1} I(x_i, x_j)}{PN} \\ I(x_i, x_j) &= \begin{cases} 1: & x_i > x_j \\ 0: & \text{otherwise} \end{cases} \end{aligned} \tag{14}$$

where P and N are the number of positive and negative samples, x_i is the classifier outputs for the ith positive sample, and x_j is the classifier outputs for jth negative sample. But the normalized WMW statistic has a defect, that is, it is nondifferentiable due to the discrete calculation. Inspired by [51], maximizing AUC can be approximated by minimizing the following loss function:

$$\begin{aligned} \mathcal{L}_{\text{AUC}}(x, y) &= \frac{\sum_{i=0}^{P-1}\sum_{j=0}^{N-1} R(x_i, x_j)}{PN} \\ R(x_i, x_j) &= \begin{cases} -\tanh(x_i - x_j - \gamma), & x_i - x_j < \gamma \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{15}$$

where γ is a hyperparameter, and y is the label. As shown in Fig. 8, for a positive sample, when its output is higher than the one of a negative sample over a margin γ in the process of minimizing $\mathcal{L}_{\text{AUC}}$, this pair of positive and negative samples will not affect the objective function with a loss of 0. It is worth mentioning a positive margin γ is required for better generalization performance.

Then, the ESC-Task loss $\mathcal{L}_{\text{ESC}}$ and SD-Task loss $\mathcal{L}_{\text{SD}}$ functions are designed as follows:

$$\begin{aligned} \mathcal{L}_{\text{ESC}} &= \mathcal{L}_{\text{cross}}(X_{\text{ESC}}, Y_{\text{ESC}}) \\ \mathcal{L}_{\text{SD}} &= \mathcal{L}_{\text{AUC}}(X_{\text{SD}}, Y_{\text{SD}}) \\ L &= \alpha\mathcal{L}_{\text{ESC}} + (1-\alpha)\mathcal{L}_{\text{SD}} \end{aligned} \tag{16}$$

where ESC-Task uses the cross entropy loss $\mathcal{L}_{\text{cross}}$ as there generally has no imbalance issue in ESC. X_{ESC} and Y_{ESC} are the output and true label of the ESC-Task, α is a hyperparameter to adjust the ESC-Task and SD-Task weights.

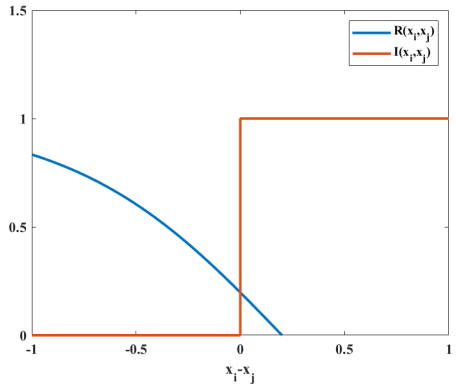


Fig. 8. $R(x_i, x_j)$ compare with $I(x_i, x_j)$, where $\gamma = 0.2$ in $R(x_i, x_j)$.

G. Setups of DSMN-ESS

In DSMN-ESS, the filtered multichannel EEGs and the corresponding STFT spectrogram are adopted as inputs. We split the data in training, testing, and validation dataset at a ratio of 6:2:2. For the hyperparameter in (16), we fix α to 0.5 to achieve the balanced performance between ESC-Task and SD-Task. To explore the effect of γ, we adjust it from 0.1 to 0.85 with a step size of 0.05. The kernel size of each 2-D convolution is 3×3 and each 3-D convolution is $3 \times 3 \times 3$. In the training, the batch size is set to be 32. The ADAM optimization technique is used with a weight decay of 0.0005, an initial learning rate of 0.001, and eps 1e-8. The implementation procedure of the proposed network DSMN-ESS is summarized in Algorithm 1.

Algorithm 1 DSMN-ESS

Input:
Filtered EEG signal and STFT spectrogram database $\mathbf{B} = \{(M_i, N_i)\}$, input $M_i \in \mathbb{R}^{h\times w}$ and $N_i \in \mathbb{R}^{D\times H\times W}$, hyperparameters $\alpha = 0.5$, γ, and maximal iterative number K.
Output:
The training model of DSMN.
Processing stage:

1) while $k \leq K$ do
   a) Randomly select a batch Filtered EEG signal and STFT spectrogram (M_i, N_i) from $\mathbf{B}$
   b) Perform feature extraction on (M_i, N_i) through forward propagation
   c) $k \leftarrow k + 1$
2) Calculate the loss $L = \alpha\mathcal{L}_{ESC} + (1-\alpha)\mathcal{L}_{SD}$
3) Compute the $loss$ by backpropagation.
4) Update the parameters of network
5) end while

III. Results and Discussion

A. Database

In this study, we evaluate the performance on real recorded EEGs from children suffering from epilepsy syndromes. The EEGs are collected from the Children's Hospital, Zhejiang

TABLE I
DATABASE INFORMATION

Epilepsy syndrome	Patient No.	Average age	Recording time (m)
EIEE	7	1m17d	1050
WEST	12	11m22d	2243
CAE	12	7y9m	4384
FS+	7	4y1m	1780
BECT	6	9y6m	4938
CSWS	5	7y11m	2930

y: year, m: month, d: day, m: minute

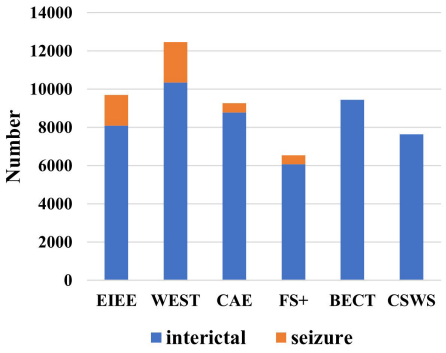


Fig. 9. Number of samples.

University School of Medicine (CHZU) with the informed consent of the patient's legal guardians. The EEGs are recorded according to the 10–20 International system, consisting of 21 channels: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, A1, A2, Fz, Cz, and Pz, with a sampling rate of 1000 Hz. In particular, the CHZU database uses a monopolar lead with the reference electrode (REF) being placed near the electrode Cz. The experimental data consist of six typical epilepsy syndromes, which are the EIEE, infantile spasms (West), childhood absence epilepsy (CAE), febrile seizures plus (FS+), benign childhood epilepsy with centrotemporal spikes (BECT), and epileptic encephalopathy with continuous spike and waves during slow wave sleep (CSWS). Table I shows the number of patients, average age, and total duration of EEG recordings for each epilepsy syndrome. With the help of neurologists in CHZU, we labeled the EEG data and divided it into seizure and interictal periods, respectively. In particular, there is no seizure data for BECT and CSWS. In this study, the EEGs are first segmented into 4 s long segments with an overlap rate of 50%. The number of seizure and interictal samples for each epilepsy syndrome used for analysis are shown in Fig. 9. Finally, each sample is labeled for subsequent neural network training.

B. Results on Multitask Learning

In this part, the first experiment is to test the effectiveness of using the AUC in SD-Task for imbalance data learning. SD-Task without using the AUC loss is denoted as $L1$, where the cross entropy is used to replace the AUC loss. The proposed loss function using AUC loss in SD-Task is denoted as $L2$. Fig. 10 shows the macroaverage AUC score obtained from the CHZU database, where the macroaveraging will cause

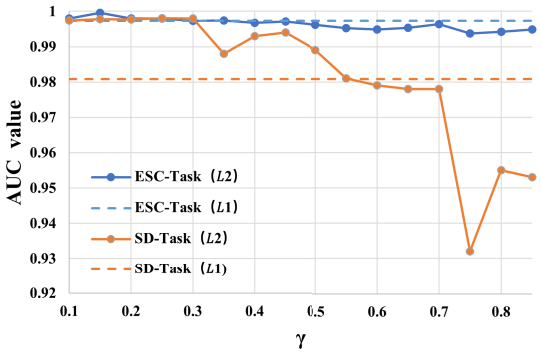


Fig. 10. Dependence of the AUC score on γ for both ESC-Task and SD-Task. $L1 = 0.5\mathcal{L}_{\text{cross}}(X_{\text{ESC}}, Y_{\text{ESC}}) + 0.5\mathcal{L}_{\text{cross}}(X_{\text{SD}}, Y_{\text{SD}})$, $L2 = 0.5\mathcal{L}_{\text{cross}}(X_{\text{ESC}}, Y_{\text{ESC}}) + 0.5\mathcal{L}_{\text{AUC}}(X_{\text{SD}}, Y_{\text{SD}})$.

TABLE II
RESULTS COMPARISON (WUM: WEIGHT UPDATE MODULE, ISG: INFORMATION SHARING GATE, N: NO, Y: YES) (a) RESULTS OF ESC-TASK. (b) RESULTS OF SD-TASK

(a)

WUM	ISG	Accuracy	Precision	Recall	Specificity	F_1	AUC
N	Y	98.25	98.24	98.23	99.65	98.23	99.45
Y	N	98.92	98.92	98.87	99.78	98.89	99.78
Y	Y	**99.27**	**99.23**	**99.30**	**99.86**	**99.52**	**99.95**

(b)

WUM	ISG	Accuracy	Precision	Recall	Specificity	F_1	AUC
N	Y	98.82	96.35	95.89	95.89	96.12	99.17
Y	N	98.67	95.50	95.86	95.86	95.68	98.83
Y	Y	**99.07**	**97.05**	**96.95**	**96.95**	**96.93**	**99.78**

calculation results to be affected by uncommon categories. In the figure, the performance of using the proposed loss $L2$ on different margin values γ is also presented. As observed as follows.

1) For the SD-Task, when a small margin value γ is used ($\gamma < 0.5$), using the AUC loss can perform consistently better than using the cross entropy loss. It indicates that AUC loss is effective in dealing with the imbalance issue in epilepsy SD. The detection accuracy reaches the optimal performance (0.9978) when $\gamma = 0.15$.
2) Using $L2$ loss, the AUC scores of both ESC-Task and SD-Task generally show a downward trend when the margin value γ is increasing, more obvious in the SD-Task. This phenomenon can be explained that ESC-Task is relatively balanced, while the data of SD-Task is extremely unbalanced, $L2$ aims to change the loss of SD-Task from cross-entropy to AUC.
3) Compared with SD-Task, the performance of ESC-Task is close on either using the AUC loss $L2$ or the cross entropy loss $L1$. Overall, using the AUC loss of $L2$ with $\gamma = 0.15$, the optimal performance can be achieved for multitask learning.

The second experiment is conducted to test the effectiveness of the weight update module and the information sharing gate. The results are shown in Table II. It is clearly observed that using the weight update module and the information sharing gate both are able to enhance the model performance.

The third experiment is conducted to compare the effectiveness of multitask learning with respect to the single task

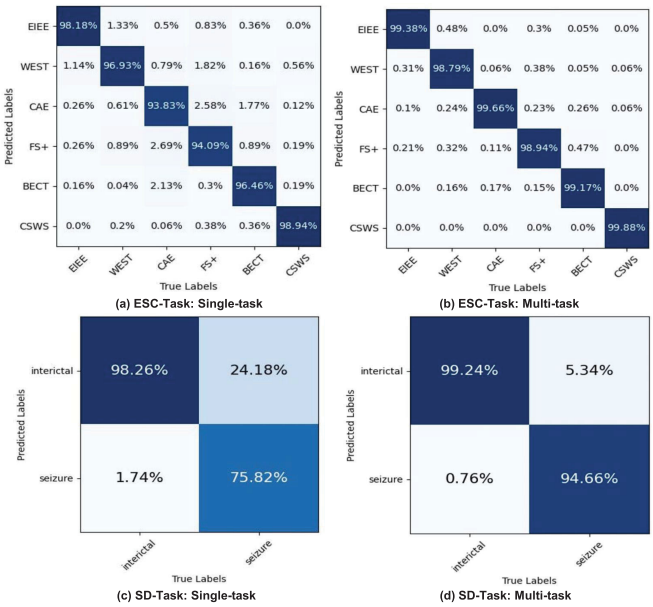


Fig. 11. Confusion matrices of single-task network and multitask DSMN network. (a) ESC-Task: Single-task. (b) ESC-Task: Multitask. (c) SD-Task: Single-task. (d) SD-T ask: Multitask.

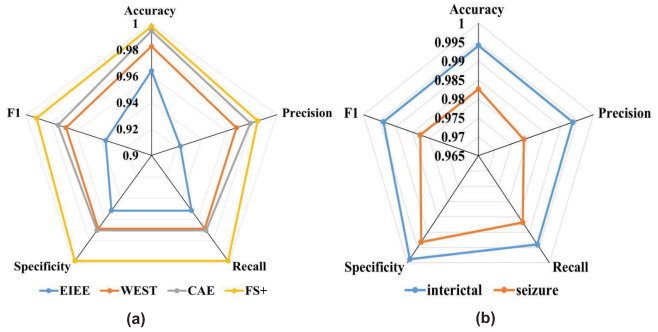


Fig. 12. Results of (a) SD for each epileptic syndrome and (b) ESC in each period.

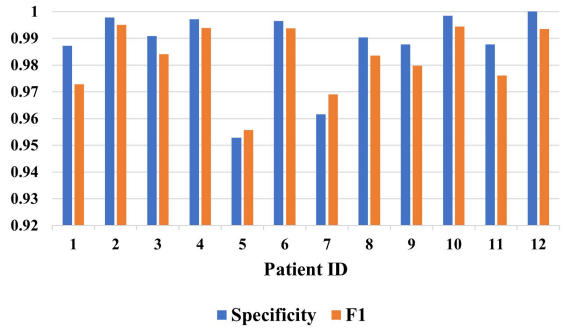


Fig. 13. Patient-specific performance based on specificity and F1 score.

model. Here, the single-task model is designed by changing the full connection module of the DSMN framework into one output. Then, we use the single-task model to train ESC-Task and SD-Task, respectively, to compare the performance with the multitask DSMN-ESS model. The confusion matrices of ESC-Task and SD-Task based on single-task model can be seen in Fig. 11(a) and (c), respectively, while Fig. 11(b) and (d) shows the corresponding confusion matrices of ESC-Task and SD-Task obtained by the multitask model. The result is derived using the optimal γ (0.15) of the two models on the SD-task. Through comparisons, we find that the single-task model performs worse than the multitask model, on the ESC-Task and SD-Task. Specifically, the confusion matrix [Fig. 11(c) and (d)] shows an increase in the recognition accuracy of all classes in both two tasks. Thus, by sharing the feature representation between related tasks, multitask learning can achieve better generalization performance than a single task.

DSMN not only can detect seizures for a specific type of epilepsy syndrome, but also classify epilepsy syndrome for a specific period, which has a strong practical significance. Fig. 12(a) shows the results of SD for each epileptic syndrome. It can be seen that the effect of SD on different epileptic syndromes is different, Fs+ is the best, and EIEE is the worst. The brain development of patients with EIEE is generally incomplete and there are many disturbances in EEGs, which may be the reason for the worst result. Fig. 12(b) shows the results of ESC in each period. It can be seen that the effect of ESC with interictal data is better than that with seizure data. This may be caused by the similarity of EEG patterns of some syndromes during the seizure. For instance, EIEE shows a high amplitude slow wave burst, followed by diffuse low voltage, similar to the pattern of infantile spasm (WEST). The EEG pattern of WEST and CAE is often accompanied by spikes. Cao et al. [31] used interictal data to classify epilepsy syndrome and achieved good results. But in [31], only ESC exploring on interictal EEGs is presented. Seizure onset detection as well as ESC on ictal EEGs is not studied.

C. Patient-Specific Analysis

The patient-specific performance of the proposed DSMN on ESC and SD is studied in this section. In particular, only 12 WEST patients are analyzed. Patient-specific analysis of other epileptic syndromes is omitted here as similar performance can be observed. The specificity and F1 score of each subject is shown in Fig. 13. It can be seen that the specificity and F1 score of patients 5 and 7 are generally lower than those of other patients. Even though, they also have excellent performance, the specificity and F1 score are greater than 0.95. The average patient-specific performance specificity and F1 score are high to 0.987 and 0.983 obtained by DSMN, respectively. For subjects 12, the specificity can even reach to 1.

D. Weight Update Module Analysis

In this section, we will discuss how to determine key brain regions related to epilepsy recognition through the proposed weight update module. Some EEG channels are irrelevant to epilepsy recognition, and the performance of the trained model will be degraded because irrelevant channels may introduce noise. Meanwhile, more computational cost is also required. Electrode set reduction can not only improve model performance but also reduce computational complexity. It is

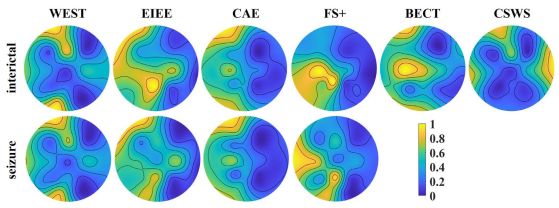


Fig. 14. Average weight distribution of different brain regions.

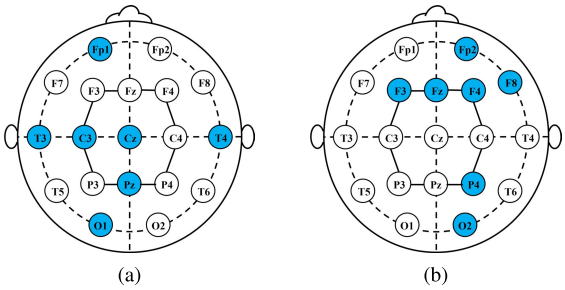
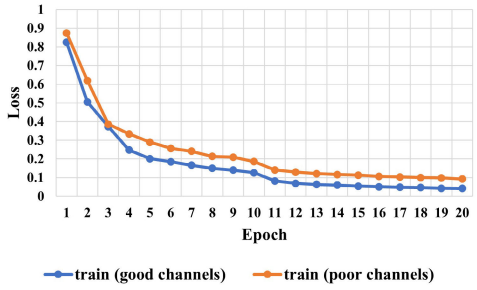
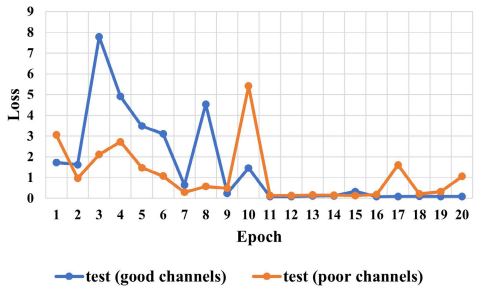


Fig. 15. Two different channel placements according to average weight distribution of different brain regions. (a) Seven channels: Fp1, C3, C4, O1, T3, T4, Cz, and Pz. (b) Seven channels: Fp2, F3, F4, P4, O2, F8, and Fz.



(a)



(b)

Fig. 16. Comparisons on the training, testing loss with different EEG channels. (a) Train loss on different epochs. (b) Test loss on different epochs.

of practical significance to study the channel weight of EEG epilepsy recognition.

In weight update module analysis, we confirmed its effectiveness by visualizing network features before and after channel weight update using T-SNE. We map the channel weight distribution of the weight update module in the first convolutional block to the brain scalp. The average channel weight of each type of sample is calculated. Fig. 14 shows the average weight distribution of different brain regions in different periods of different syndromes. As observed, the epileptic status is highly correlated with brain regions. For example, in CAE, the network pays more attention to the prefrontal region, the central temporal region, and the occipital region, especially in the prefrontal region. This is consistent with the results of [5], and [52], stating that frequent posterior cephalic rhythm δ activity and focal spikes can be found in the interical EEG of CAE (the most obvious area is the frontal area, which can also be located in the central temporal area or the occipital area). For BECT, unlike absence epilepsy, the network pays more attention to the central temporal region. This result has also been confirmed in [53], [54], and [55] that the epileptiform discharge of BECT usually has the highest amplitude in the central temporal region and/or the central region. It can also reach the parietal region and the posterior temporal region. Sometimes the discharge in the frontal or occipital region can also be observed.

According to the average weight distribution of different brain regions in Fig. 14, we further conducted analysis on using two different electrode placements. Fig. 15 shows the two different channel placements evaluated in this experiment (here, according to the above weight distribution, good channel refers to that both syndrome and period are paid more attention, and poor channel means that both syndrome and period are not paid much attention). We replace 21 EEG channels with these two different channel placements as the network inputs and compare their performance. For the network, the dimension of the full connection layer is adjusted to adapt to the new inputs and the weight update module has been removed. Using good channels, our model can achieve the AUC of 99.42% and 98.91% in ESC-Task and SD-Task, respectively, only slightly lower than the AUC of 99.95% and 99.78% when using the full 21 channels. Fig. 16 shows the training and testing loss using good channels and poor channels, with different learning epochs. It is observed that the training loss of good channels is lower than using poor channels in all epochs. When the training model tends to converge, the testing loss of good channels is lower than using the poor channels. In addition, the model appears overfitting phenomenon when using poor channels in training. These poor channels may have some noise features, leading to a degraded performance. These results confirm the effectiveness of channel selection.

E. Comparisons With State-of-the-Art (SOTA) Methods

In this section, the performance of DSMN-ESS is compared with several SOTA multitask methods on ESC and SD. The SOTA methods include the multitask fully convolutional neural network (FCN) architecture [56], the multitask framework GEME [34], EEGNet [57], LightSeizureNet [58], and Transformer [59]. To have a fair comparison, we adopted the dual flow mode in all compared methods (copy the backbone of the network to become another flow), the dimension of each stream input determines whether the convolution operation in the network is extended to 2-D or 3-D operations. Moreover, the multicategory classification in multitask FCN is changed to two-category classification, and the single output in EEGNet, LightSeizureNet, and Transformer are extended to two outputs to change single-task model to multitask model. For all

TABLE III
RESULTS COMPARISONS WITH RESPECT TO SOTA MODELS. (a) RESULTS OF ESC-TASK. (b) RESULTS OF SD-TASK

(a)

LOSS	Network	ACC	P	R	S	F1	AUC
$L1$	Capsule network	79.73	80.44	78.62	95.88	79.15	96.11
	FCN[56]	91.58	91.48	91.34	98.31	91.38	95.02
	GEME[34]	96.78	96.77	96.68	99.35	96.72	97.80
	EEGNet[57]	93.01	92.91	92.64	98.59	92.76	96.72
	LightSeizureNet[58]	87.98	87.90	87.71	97.58	87.69	92.67
	Transformer[59]	93.07	93.03	93.00	98.60	93.01	97.94
	Ours	**99.27**	**99.26**	**99.29**	**99.85**	**99.54**	**99.73**
$L2$	FCN[56]	92.45	92.32	92.07	98.49	92.18	95.43
	GEME[34]	98.17	98.26	98.04	99.63	98.15	97.96
	EEGNet[57]	85.38	85.69	85.71	97.13	84.94	95.32
	LightSeizureNet[58]	88.83	88.99	88.36	97.74	88.58	92.79
	Transformer[59]	94.94	94.98	94.73	98.97	94.85	98.94
	Ours	**99.27**	**99.23**	**99.30**	**99.86**	**99.52**	**99.95**

(b)

LOSS	Network	ACC	P	R	S	F1	AUC
$L1$	Capsule network	95.30	90.10	76.39	76.39	81.54	93.81
	FCN[56]	96.29	87.61	88.27	88.27	87.95	94.60
	GEME[34]	97.95	93.82	92.60	92.60	93.19	97.60
	EEGNet[57]	97.16	90.44	91.12	91.12	90.78	97.02
	LightSeizureNet[58]	95.05	84.26	77.83	77.83	81.54	88.84
	Transformer[59]	97.60	93.75	90.08	90.08	91.85	97.69
	Ours	**98.86**	**95.69**	**96.82**	**96.82**	**96.31**	**98.08**
$L2$	FCN[56]	97.68	92.61	92.10	92.10	92.35	96.21
	GEME[34]	98.54	95.95	94.35	94.35	95.13	98.57
	EEGNet[57]	96.56	86.98	93.13	93.13	89.55	97.41
	LightSeizureNet[58]	95.03	86.76	82.60	82.60	83.41	90.27
	Transformer[59]	97.30	90.61	92.09	92.09	91.33	98.28
	Ours	**99.07**	**97.05**	**96.95**	**96.95**	**96.93**	**99.78**

$L1$: $0.5\mathcal{L}_{\text{cross}}\left(X_{ESC}, Y_{ESC}\right) + 0.5\mathcal{L}_{\text{cross}}\left(X_{SD}, Y_{SD}\right)$
$L2$: $0.5\mathcal{L}_{\text{cross}}\left(X_{ESC}, Y_{ESC}\right) + 0.5\mathcal{L}_{\text{AUC}}\left(X_{SD}, Y_{SD}\right)$
ACC: Accuracy, P: Precision, R: Recall, S: Specificity

methods, both the $L1$ and $L2$ loss functions are evaluated. The parameters of the FCN, GEME, EEGNet, LightSeizureNet, Transformer, and our proposed DSMN-ESS are 16.76, 1.14, 0.02, 2.06, 13.89, and 2.20 M, respectively. The GFLOPs of FCN, GEME, EEGNet, LightSeizureNet, Transformer, and our proposed DSMN-ESS are 10.22, 0.14, 9.05, 0.635, 11.49, and 2.06, respectively. Although the complexity of our model is larger than GEME and LightSeizureNet, it is still acceptable compared to FCN, EEGNET, and Transformer. Our model brings a performance improvement at the expense of a slightly computational complexity.

Table III shows the results listed on the ESC-Task and SD-Task, respectively. For both $L1$ and $L2$ losses, the model with the best result is highlighted by bold font in the table. As presented: 1) for both the ESC-Task and SD-Task, our DSMN-ESS model obtains the best results on all evaluation metrics under both the $L1$ and $L2$ loss function; 2) for FCN, GEME and LightSeizureNet models with $L2$, the performance of SD-Task and ESC-Task are generally improved when comparing with using the $L1$ loss function; and 3) although the AUC of SD-Task is slightly improved by EEGNet with the $L2$ loss function, most other performance metrics of ESC-Task are reduced. It is speculated that EEGNet cannot be adjusted well due to using a simple structure with too few parameters.

IV. CONCLUSION

To simultaneously achieve childhood ESC and SD, a dual-stream multitask learning network DSMN-ESS has been developed in this article. To exploit the correlation between epilepsy syndrome and SD, a weight update module and an information-sharing gate were designed in DSMN-ESS. To further address the imbalance issue in the SD-Task, a loss function based on the AUC optimization is designed in DSMN-ESS learning. The study was evaluated on the long-term EEGs of 49 childhood epilepsy patients recorded from CHZU, and the ablation study and comparison experiment validated that as follows.

1) The proposed DSMN-ESS model can achieve good results in both ESC-Task and SD-Task with 99.95% and 99.78% AUCs, respectively. The adopted AUC loss can well address the data imbalance issue in epilepsy detection.
2) The proposed multitask learning can achieve better performance than single-task learning, either on ESC and SD.
3) For different epileptic syndromes, the active and preferable brain region is different. The weight update module can achieve reasonable EEG channel and brain region selection, which can further improve the model performance.

Future research will consider improving EEG signal preprocessing and feature extraction stages to embed them in the network to form an end-to-end learning network.

REFERENCES

[1] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowl.-Based Syst.*, vol. 45, pp. 147–165, Jun. 2013.

[2] P. Jain, S. Sharma, and M. Tripathi, "Diagnosis and management of epileptic encephalopathies in children," *Epilepsy Res. Treatment*, vol. 2013, pp. 1–9, Jul. 2013.

[3] J. De Jong et al., "Agenesis of the corpus callosum, infantile spasms, ocular anomalies (Aicardi's syndrome). Clinical and pathologic findings," *Neurology*, vol. 26, no. 12, pp. 1152–1158, 1976.

[4] R. Zheng et al., "Scalp EEG functional connection and brain network in infants with west syndrome," *Neural Netw.*, vol. 153, pp. 76–86, Sep. 2022.

[5] J. T. Park, A. M. Shahid, and A. Jammoul, "Common pediatric epilepsy syndromes," *Pediatric Ann.*, vol. 44, no. 2, pp. 30–35, Feb. 2015.

[6] N. Sofijanov et al., "Febrile seizures: Clinical characteristics and initial EEG," *Epilepsia*, vol. 33, no. 1, pp. 52–57, Jan. 1992.

[7] B. D. Bernardina et al., "Sleep and benign partial epilepsies of childhood: EEG and evoked potentials study," *Epilepsy Res. Suppl.*, vol. 2, pp. 83–96, Jan. 1991.

[8] E. R. Perez, V. Davidoff, P.-A. Desplartd, and T. Deonna, "Mental and behavioural deterioration of children with epilepsy and CSWS: Acquried epileptic frontal syndrome," *Develop. Med. Child Neurol.*, vol. 35, no. 8, pp. 661–674, Nov. 2008.

[9] M. J. Friedman and G. Q. Sharieff, "Seizures in children," *Pediatric Clinics*, vol. 53, no. 2, pp. 257–277, 2006.

[10] J. Wang, L. Shi, W. Wang, and Z.-G. Hou, "Efficient brain decoding based on adaptive EEG channel selection and transformation," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 6, pp. 1314–1323, Dec. 2022.

[11] A. M. Alvi, S. Siuly, and H. Wang, "A long short-term memory based framework for early detection of mild cognitive impairment from EEG signals," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 2, pp. 375–388, Apr. 2023.

[12] E. Eldele et al., "ADAST: Attentive cross-domain EEG-based sleep staging framework with iterative self-training," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 1, pp. 210–221, Feb. 2023.

[13] Q. She, C. Zhang, F. Fang, Y. Ma, and Y. Zhang, "Multisource associate domain adaptation for cross-subject and cross-session EEG emotion recognition," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[14] Y. Zhou et al., "Cross-subject cognitive workload recognition based on EEG and deep domain adaptation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[15] H. Gao et al., "Basic taste sensation recognition from EEG based on multiscale convolutional neural network with residual learning," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.

[16] R. S. Selvakumari, M. Mahalakshmi, and P. Prashalee, "Patient-specific seizure detection method using hybrid classifier with optimized electrodes," *J. Med. Syst.*, vol. 43, no. 5, pp. 1–7, May 2019.

[17] M. Zabihi, S. Kiranyaz, T. Ince, and M. Gabbouj, "Patient-specific epileptic seizure detection in long-term EEG recording in paediatric patients with intractable seizures," in *Proc. IET Intell. Signal Process. Conf.*, 2013, pp. 1–7.

[18] R. Zheng, J. Cao, Y. Feng, X. Zhao, T. Jiang, and F. Gao, "Seizure prediction analysis of infantile spasms," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 366–376, 2023.

[19] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4489–4497.

[20] D. Tran, J. Ray, Z. Shou, S.-F. Chang, and M. Paluri, "ConvNet architecture search for spatiotemporal feature learning," 2017, *arXiv:1708.05038*.

[21] C. Feichtenhofer, "X3D: Expanding architectures for efficient video recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 200–210.

[22] J. Cao, D. Hu, Y. Wang, J. Wang, and B. Lei, "Epileptic classification with deep-transfer-learning-based feature fusion algorithm," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 2, pp. 684–695, Jun. 2022.

[23] X. Cui, J. Cao, X. Lai, T. Jiang, and F. Gao, "Cluster embedding joint-probability-discrepancy transfer for cross-subject seizure detection," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 593–605, 2023.

[24] T. Tao, L. Guo, Q. He, H. Zhang, and L. Xu, "Seizure detection by brain-connectivity analysis using dynamic graph isomorphism network," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 2302–2305.

[25] Y. Li, Y. Liu, Y.-Z. Guo, X.-F. Liao, B. Hu, and T. Yu, "Spatio-temporal–spectral hierarchical graph convolutional network with semisupervised active learning for patient-specific seizure prediction," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 12189–12204, Nov. 2022.

[26] C. Li, Y. Zhao, R. Song, X. Liu, R. Qian, and X. Chen, "Patient-specific seizure prediction from electroencephalogram signal via multi-channel feedback capsule network," *IEEE Trans. Cognit. Develop. Syst.*, early access, Oct. 5, 2022, doi: 10.1109/TCDS.2022.3212019.

[27] Y. Sun et al., "Continuous seizure detection based on transformer and long-term iEEG," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 11, pp. 5418–5427, Nov. 2022.

[28] X. Cui, J. Cao, D. Hu, T. Wang, T. Jiang, and F. Gao, "Regional scalp EEGs analysis and classification on typical childhood epilepsy syndromes," *IEEE Trans. Cognit. Develop. Syst.*, vol. 52, no. 2, pp. 662–674, Jun. 2023.

[29] X. Cui et al., "Deep feature fusion based childhood epilepsy syndrome classification from electroencephalogram," *Neural Netw.*, vol. 150, pp. 313–325, Jun. 2022.

[30] Y. Feng et al., "3D residual-attention-deep-network-based childhood epilepsy syndrome classification," *Knowl.-Based Syst.*, vol. 248, Jul. 2022, Art. no. 108856.

[31] J. Cao et al., "Two-stream attention 3D deep network based childhood epilepsy syndrome classification," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2022.

[32] D. Kollias, V. Sharmanska, and S. Zafeiriou, "Face behavior a la carte: Expressions, affect and action units in a single network," 2019, *arXiv:1910.11111*.

[33] A. Ruiz, J. Van De Weijer, and X. Binefa, "From emotions to action units with hidden and semi-hidden-task learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3703–3711.

[34] X. Nie, M. A. Takalkar, M. Duan, H. Zhang, and M. Xu, "GEME: Dual-stream multi-task gender-based micro-expression recognition," *Neurocomputing*, vol. 427, pp. 13–28, Feb. 2021.

[35] J. A. Miranda-Correa and I. Patras, "A multi-task cascaded network for prediction of affect, personality, mood and social context using EEG signals," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 373–380.

[36] X. Ma, S. Qiu, Y. Zhang, X. Lian, and H. He, "Predicting epileptic seizures from intracranial EEG using LSTM-based multi-task learning," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*. Cham, Switzerland: Springer, 2018, pp. 157–167.

[37] T. Loddenkemper, "Proposal for Revised Classification of Epilepsies and Epileptic Syndromes, Commission on Classification and Terminology of the International League Against Epilepsy," *Epilepsia*, vol. 30, no. 4, pp. 389–399, 1989.

[38] L. Fusco and F. Vigevano, "Ictal clinical electroencephalographic findings of spasms in west syndrome," *Epilepsia*, vol. 34, no. 4, pp. 671–678, Jul. 1993.

[39] E. Beghi, "The concept of the epilepsy syndrome: How useful is it in clinical practice?" *Epilepsia*, vol. 50, pp. 4–10, May 2009.

[40] 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.

[41] 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.

[42] D. Hu, J. Cao, X. Lai, Y. Wang, S. Wang, and Y. Ding, "Epileptic state classification by fusing hand-crafted and deep learning EEG features," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 4, pp. 1542–1546, Apr. 2021.

[43] Z. Wang, D. Wu, F. Dong, J. Cao, T. Jiang, and J. Liu, "A novel spike detection algorithm based on multi-channel of BECT EEG signals," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 12, pp. 3592–3596, Dec. 2020.

[44] D. Hu, J. Cao, X. Lai, J. Liu, S. Wang, and Y. Ding, "Epileptic signal classification based on synthetic minority oversampling and blending algorithm," *IEEE Trans. Cognit. Develop. Syst.*, vol. 13, no. 2, pp. 368–382, Jun. 2021.

[45] Z. Xu, T. Wang, J. Cao, Z. Bao, T. Jiang, and F. Gao, "BECT spike detection based on novel EEG sequence features and LSTM algorithms," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1734–1743, 2021.

[46] 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.

[47] J. Cao et al., "Unsupervised eye blink artifact detection from EEG with Gaussian mixture model," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 8, pp. 2895–2905, Aug. 2021.

[48] R. Rosch, T. Baldeweg, F. Moeller, and G. Baier, "Network dynamics in the healthy and epileptic developing brain," *Netw. Neurosci.*, vol. 2, no. 1, pp. 41–59, Mar. 2018.

[49] C. Ehmann and W. Samek, "Transferring information between neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2361–2365.

[50] Y. Li, A. Kazemeini, Y. Mehta, and E. Cambria, "Multitask learning for emotion and personality traits detection," *Neurocomputing*, vol. 493, pp. 340–350, Jul. 2022.

[51] L. Yan, R. H. Dodier, M. Mozer, and R. H. Wolniewicz, "Optimizing classifier performance via an approximation to the Wilcoxon–Mann–Whitney statistic," in *Proc. 20th Int. Conf. Mach. Learn.*, 2003, pp. 848–855.

[52] L. G. Sadleir, K. Farrell, S. Smith, M. B. Connolly, and I. E. Scheffer, "Electroclinical features of absence seizures in childhood absence epilepsy," *Neurology*, vol. 67, no. 3, pp. 413–418, Aug. 2006.

[53] J. Nicolai, A. P. Aldenkamp, J. Arends, J. W. Weber, and J. S. H. Vles, "Cognitive and behavioral effects of nocturnal epileptiform discharges in children with benign childhood epilepsy with centrotemporal spikes," *Epilepsy Behav.*, vol. 8, no. 1, pp. 56–70, Feb. 2006.

[54] N. Specchio et al., "International league against epilepsy classification and definition of epilepsy syndromes with onset in childhood: Position paper by the ILAE task force on nosology and definitions," *Epilepsia*, vol. 63, no. 6, pp. 1398–1442, Jun. 2022.

[55] J. Nicolai et al., "EEG characteristics related to educational impairments in children with benign childhood epilepsy with centrotemporal spikes," *Epilepsia*, vol. 48, no. 11, pp. 2093–2100, Nov. 2007.

[56] A. I. L. Namburete, W. Xie, M. Yaqub, A. Zisserman, and J. A. Noble, "Fully-automated alignment of 3D fetal brain ultrasound to a canonical reference space using multi-task learning," *Med. Image Anal.*, vol. 46, pp. 1–14, May 2018.

[57] 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.

[58] S. Qiu, W. Wang, and H. Jiao, "LightSeizureNet: A lightweight deep learning model for real-time epileptic seizure detection," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 4, pp. 1845–1856, Apr. 2023.

[59] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

Jiuwen Cao (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the School of Applied Mathematics, University of Electronic Science and Technology of China, Chengdu, China, in 2005 and 2008, respectively, and the Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore, in 2013.

From 2012 to 2013, he was a Research Fellow with NTU. He is currently a Professor with Hangzhou Dianzi University, Hangzhou, China, where he is also the Dean of the School of Automation. He has authored more than 180 papers in top-tier journals and conferences. His main research interests include machine learning, neural networks, and medical signal processing.

Dr. Cao has been awarded the Best Conference Paper of ICCSIP 2020 and the Best Conference Paper Finalist of ICCSIP2021. He served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPER, *Journal of the Franklin Institute*, *Neurocomputing*, *Military Medical Research*, and *Multidimensional Systems and Signal Processing*.

Yaohui Chen received the B.S. degree from Zhejiang Ocean University, Zhejiang, China, in 2021. He is currently pursuing the M.S. degree with the Institute of Information and Control, Hangzhou Dianzi University, Hangzhou, China.

His current research interests include machine learning, neural networks, and signal processing.

Runze Zheng received the B.S. degree from the Shaoyang College of Electrical Engineering, Shaoyang, China, in 2018, and the M.S. degree from the Hangzhou Dianzi University of Automation, Hangzhou, China, in 2021, where he is currently pursuing the Ph.D. degree with the School of Automation.

His current research interests include childhood epilepsy, brain networks, and intelligent signal processing and analysis.

Xiaonan Cui is currently pursuing the Ph.D. degree with the School of Automation, Hangzhou Dianzi University, Hangzhou, China.

Her current research interests include pattern recognition, transfer learning, and medical image processing.

Tiejia Jiang received the B.S. degree from the Department of Clinical Medicine, Wenzhou Medical University, Wenzhou, China, in 2008, and the M.S. degree in pediatrics from Zhejiang University, Hangzhou, China, in 2017.

He is currently the Deputy Chief Physician with the Children's Hospital affiliated to Medical College, Zhejiang University, mainly engaged in electrophysiological signal analysis and research of various children's neurological diseases.

Feng Gao received the B.S. degree from the Department of Clinical Medicine, Zhejiang University, Hangzhou, China, in 1989, and the M.S. degree in pediatrics from Zhejiang University, in 1998.

He is currently the Chief Physician with the Children's Hospital affiliated to Medical College, Zhejiang University, engaged in clinical and scientific research of children's neurological diseases. His main interest in research is the electrophysiological characteristics and etiology of epilepsy.