A Spatiotemporal Channel Attention Residual Network With Extended Series Mean Amplitude Spectrum for Epilepsy Detection

Qi Wang, Chenxi Huang, Qingpeng Zeng, Chunquan Li, *Member, IEEE*, and Ting Shu

***Abstract*—Accurately detecting electroencephalogram (EEG) signals in a specific period before the epileptic seizure can effectively predict epilepsy and reduce the harm caused by epilepsy to the patient. However, the current research rarely pays attention to the influence of spatiotemporal features on the detection of EEG signals before seizures. To accurately predict epilepsy before seizures, this article proposes a spatiotemporal channel attention residual network (STCARN) with extended series mean amplitude spectrum (MAS) of EEG signals. Specifically, the extended series MAS feature representation is first developed to rationally combine the temporal relevance of multiple MAS and the spatial relevance of EEG channels, fully representing the related activities of the brain. Furthermore, STCARN is proposed to extract the spatiotemporal information of extended series MASs by rationally fusing residual convolutional structure, channel attention mechanism, and recurrent network structure, which significantly improves the performance of epilepsy detection. STCARN is evaluated in the three-classification and five-classification tasks for epilepsy detection. The results indicate that STCARN achieves 99.98% and 99.51% accuracy in the two tasks, respectively.**

***Index Terms*—Convolution neural network (CNN), LSTM, mean amplitude spectrum (MAS), preictal prediction, seizure prediction.**

I. Introduction

EPILEPSY is a common chronic noncommunicable brain disease in humans, distributed in all age groups. Currently, there are approximately 50 million patients with epilepsy in the world. People with epilepsy are three times more likely to die prematurely than the general population [1]. It is very effective to detect epilepsy through electroencephalography (EEG) signals from the brain–computer interface. Because EEG signals contain a wealth of EEG information by recording and amplifying the spontaneous biopotential of the brain on the scalp with a sophisticated electronic instrument, which can reflect the spontaneous and rhythmic electrical activities of the brain cell group [2]. The clinical analysis of EEG mainly relies on the observation and manual annotation of medical staff, which is also the current gold standard for detecting epilepsy [3]. However, due to the different features of the initial position and propagation mode of brain activity, there is great uncertainty in epileptic seizures. Furthermore, due to the randomness of epileptic seizures and the artificial judgment of subjective factors, the effect of manual feature extraction and classification is often not ideal, with great limitations, especially in the face of tens of thousands of EEG data. Therefore, it is of great significance to develop epilepsy detection and prediction algorithms based on EEG signals [4].

Since Gotman [5] proposed an automatic epilepsy recognition method based on EEG signals in 1982, the EEG-based automatic epilepsy detection method has received widespread attention. Guo et al. [6] used an artificial neural network (ANN) to classify EEG signals regarding the presence or absence of seizures. Li et al. [7] and Ouelha and Boashash [8] applied a support vector machine (SVM) to perform automatic detection in epilepsy recognition. In addition, some K-nearest neighbors [9] and random forest LR [10] methods have also been applied to epilepsy detection. Song and Zhang [11] used an ANN-based extreme learning machine (ELM) to provide epilepsy automatic detection.

Because deep-learning technology can provide efficient feature extraction and autonomous learning capabilities, it has been widely used in many fields, such as computer vision and brain–computer interface [12]. Deep-learning methods are also frequently used in epilepsy detection. Acharya et al. [13] proposed a 13-layer deep neural network (DNN) for epilepsy detection on the original EEG signal. Zhang et al. [14] proposed a new complex DNN model with an attention mechanism to learn specific epilepsy features from EEG signals. Avcu et al. [15] proposed a seizureNet convolutional neural network, which can realize the automatic detection of epilepsy by converting the original signal into a time-spectrogram through the Fourier transform. Li et al. [16] proposed a convolution neural network (CNN) model to extract

Manuscript received 6 December 2021; revised 31 July 2022 and 30 September 2022; accepted 14 November 2022. Date of publication 26 December 2022; date of current version 11 December 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62173176, Grant 61863028, Grant 81660299, and Grant 61503177; and in part by the Science and Technology Department of Jiangxi Province of China under Grant 2020ABC03A39 and Grant 20212ABC03A06. *(Qi Wang, Chenxi Huang, and Qingpeng Zeng are contributed equally to this work.) (Corresponding author: Chunquan Li.)*

Qi Wang, Chenxi Huang, and Qingpeng Zeng are with the College of Mathematics and Computer Sciences, Nanchang University, Nanchang 330031, China (e-mail: 198608502@qq.com; hcxncu@outlook.com; zengqingpeng@ncu.edu.cn).

Chunquan Li is with the School of Information Engineering, Nanchang University, Nanchang 330031, China (e-mail: lichunquan@ncu.edu.cn).

Ting Shu is with the Second Affiliated Hospital, Nanchang University, Nanchang 330031, China (e-mail: shuting1020@163.com).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TCDS.2022.3232121.

Digital Object Identifier 10.1109/TCDS.2022.3232121

the time–frequency features of EEG signals by embedding a squeeze and excitation pattern [17]. Bizopoulos et al. [18] performed feature learning on EEG images for detection via AlexNet [19], VGGNet [20], ResNet [21], and DenseNet [22]. Besides convolutional neural networks, recurrent neural network (RNN) learning models have also been used to perform epilepsy detection. Chen et al. [23] used a three-layer LSTM [24] network for feature extraction and then used the sigmoid activation function to perform epilepsy detection. In addition, Roy et al. [25] designed an RNN called ChronoNet with softmax for epilepsy classification.

With the development of some portable EEG detection equipment, it is possible to realize preseizure detection of epilepsy, namely, real-time detection before epilepsy. Studies in [26], [27], and [28] pointed out that, due to the sudden and repetitive nature of epileptic seizures, detecting preseizure EEG signals with a time scale of less than 1 h as accurately and reasonably as possible is the key to the treatment and prevention of epileptic seizures. Hu et al. [26] proposed a CNN+SVM classification method based on the mean amplitude spectrum (MAS) feature of the EEG signals, which can detect the preseizure state, seizure state, and interseizure period of epilepsy at different time scales of EEG. Cao et al. [28] used Inception-v3, Resnet152, and Inception-Resnet-v2 networks with the MAS feature extraction method to perform epilepsy classification detection. Cao et al. [27] proposed a stacked convolutional neural network, which uses multiple subnetworks with AWF scheme to achieve epilepsy detection. Cao et al. [29] used AlexNet, VGG-19, ResNet152, etc., with migration learning to achieve the best preseizure detection accuracy rate.

Particularly, effective feature representation is also very important for epilepsy detection based on EEG signals. Specifically, typical epilepsy feature representation from EEG signals can effectively describe the difference between EEG signals in various states, such as seizures and normal (the difference between spikes and background signals), thereby helping the classification model to effectively perform epileptic seizure screening [30]. The common EEG features representative methods include autocorrelation and correlation density [31], maximum Lyapunov exponent [32], power spectral density and entropy [33], [34], wavelet transform [35], [36], nonlinear analysis [37], feature rhythm sub-band statistics [38], fractional linear prediction modeling error [39], multiscale radial basis functions (RBFs) and improved particle swarm optimization (MRBF-MPSO) time–frequency representation [40], regular multivariate decomposition (CPD) and block item decomposition (BTD) [41], the spatial pattern of principal component analysis [42], etc.

Although the above studies have achieved promising epilepsy detection performance, further improvements are still needed.

1) The above most investigations on epilepsy automatic detection failed to involve the detection of preseizure. But in fact, preseizure detection is very important. Correct detection of preseizure signals can help to take timely measures to reduce the harm of epilepsy to patients.
2) Only a small number of investigations in [26], [27], [28], and [29] conducted preseizure detection on epilepsy through convolutional neural network detection or ensemble learning. However, these investigations did not fully consider how to effectively establish the relationship between the spatiotemporal characteristics of EEG signals for epilepsy detection, degenerating the detection accuracy.
3) Most of the existing feature representations of EEG failed to consider the temporal relevance of features for the detection of preseizure, reducing the consistency of feature expression of EEG signals.

To address these problems, inspired by the cascaded convolutional recurrent network (CRNN) [43] and the method of extracting temporal features [44], we propose a spatiotemporal channel attention residual network (STCARN) with extended series MAS of EEG signals for the detection of preseizure. Specifically, an extended MAS feature representation is first proposed to effectively establish the spatiotemporal relevance of EEG signals by combining multiple MAS in time continuity with the expansion of the number of sub-bands and channels of EEG signals. More importantly, STCARN is proposed to effectively extract the spatiotemporal relationship of extended series MAS at low, medium, and high levels by rationally combining residual convolutional structure, channel attention mechanism, and recurrent network structure. Via the channel attention mechanism, STCARN automatically obtains the important channel features and suppresses the unimportant channel features. Through the operation of residual connection, STCARN solves the problems of gradient disappearance, gradient explosion, and model degradation in network model training. Particularly, unlike CRNN combining the convolution and recurrent network for motor imagination [43], STCARN combines the residual convolution, channel attention mechanism, and recurrent network structure for epilepsy detection. Unlike CRNN using position electrode diagrams as feature information for motor imagination recognition, the proposed STCARN proposes a new extended series MAS for epilepsy detection.

The contribution of this article can be summarized as follows.

1) A new extended series MAS is proposed to fully represent the spatiotemporal relevance of EEG signals, which can more effectively reflect the related activities of the brain in the detection of epilepsy preseizure.
2) A data-driven STCARN model is proposed to effectively extract the spatiotemporal relationship of extended series MAS of EEG signals, which can significantly improve epilepsy detection accuracy.
3) A comprehensive evaluation involving the classified three states (interictal, preictal, and seizure) and five states (interictal, preictal three states, and seizure) is conducted with related experiments on the CHB-MIT [45] data set. The experimental results indicate that the

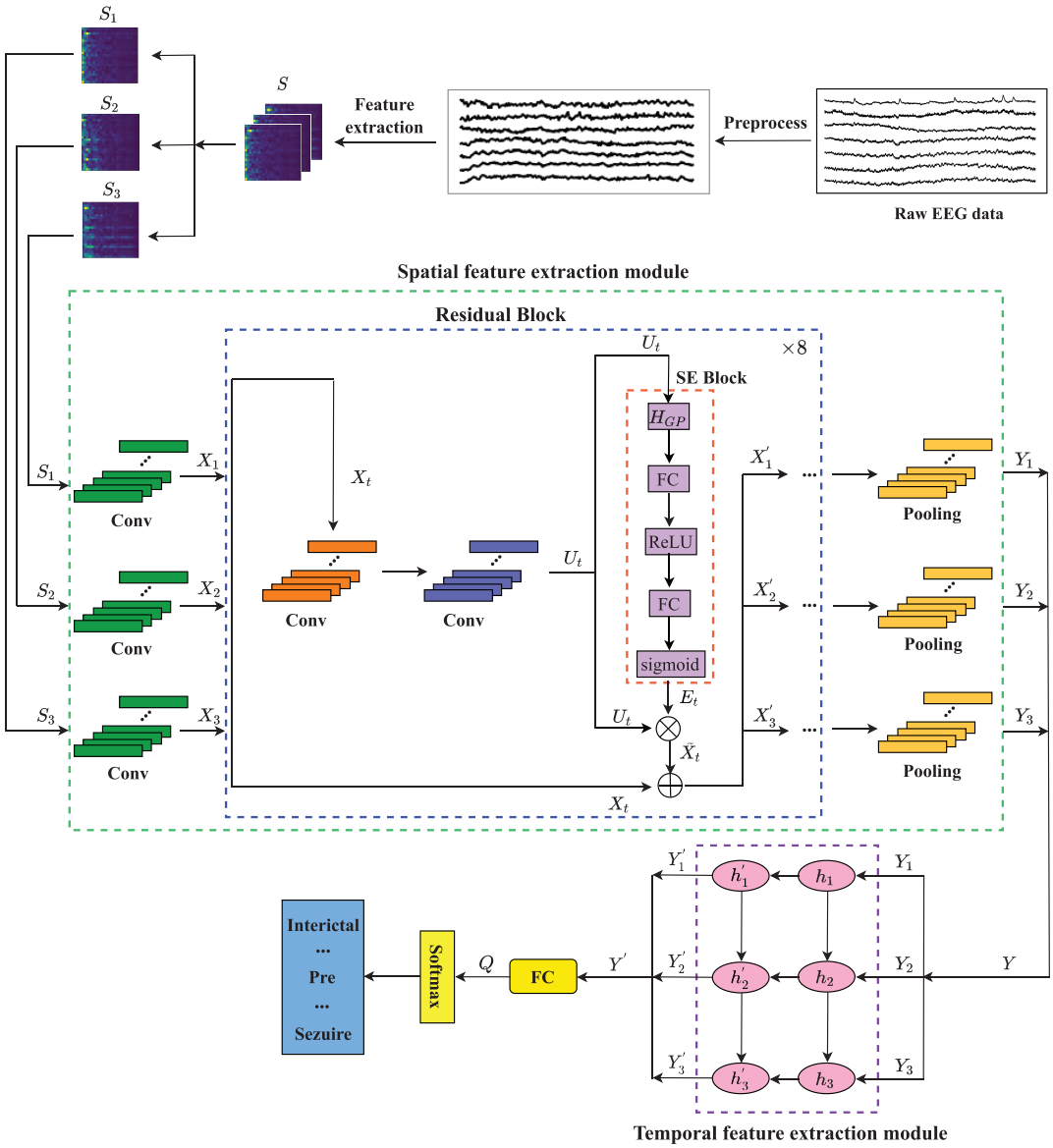


Fig. 1. Flowchart of epilepsy detection based on STCARN.

proposed method can provide better performance than existing state-of-the-art methods in epilepsy detection.

II. Proposed Algorithm

A. Overview

Fig. 1 shows the overall flow of the proposed epilepsy detection method, which is roughly divided into two parts: 1) extended series MAS and 2) STCARN. The extended series MAS is extracted to obtain the spatiotemporal relevance of EEG signals. Each extended series MAS includes three MAS features (S_1, S_2, and S_3) in time continuity and can effectively reflect the related spatiotemporal activities of the brain in the detection of preseizure.

STCARN covers the spatiotemporal feature extraction module. The spatial feature extraction module covers three residual blocks, which are used to deal with the extended series MAS features (S_1, S_2, and S_3), respectively. Note that the three residual blocks' weights in the spatial feature extraction module are shared. For each residual block, the residual convolutional network with the channel attention mechanism is developed to extract spatial features from one MAS feature. Therefore, the spatial feature extraction module can effectively extract the spatial features of extended series MASs, namely, three MAS features: S_1, S_2, and S_3 in time continuity, instead of only one MAS feature such as S_1 when performing the epilepsy detection of preseizure.

The temporal feature extraction module introduces LSTM to extract features from the extended series MAS. Specifically, the output of the spatial feature extraction module is the mapped features of the extended series MAS features, namely, (X_1', X_2', and X_3'). They are entered into pooling layers to extract Y_1, Y_2, and Y_3. LSTM uses the extracted Y_1, Y_2, and Y_3 to obtain the effective temporal relationship features between different MAS features. Finally, the features are stacked and detected by the softmax of STCARN.

Note that STCARN is developed to map the extended series MAS features for extracting the spatiotemporal feature information of EEG signals. Therefore, STCARN is regarded as a data-driven spatiotemporal network model. Through the data driven of extended series MAS features, STCARN can rationally combine residual convolutional structure, recurrent network structure, and channel attention mechanism to effectively extract the spatiotemporal information of extended series MAS for the epilepsy detection of preseizure. The detailed description involving extended series MAS and STCARN is given as follows.

B. Extended Series Mean Amplitude Spectrum

Although epilepsy is caused by the abnormal firing of neurons in the brain, the correlation between the abnormal firing of neurons and epilepsy has not been fully studied. EEG can record abnormal spontaneous biological potential of the brain, called epileptic discharges involving EEG sharp wave, spike wave, spike slow complex wave, and other sudden abnormalities. Therefore, it is a very effective method to detect epilepsy by extracting and analyzing the activity state of EEG in various bands.

The frequency range of human EEG signals mainly concentrates on 0.3–70 Hz, namely, δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–70 Hz), which can reflect most of the brain activities. Adults generally do not have δ band EEG signals when they are awake. However, the presence of β band brain waves will be more obvious in the cerebral cortex, once they are excited.

MAS [26] is a feature representation method to detect spectral information of various frequencies in the brain. As shown in Fig. 2(a), [26], [27], [28], and [29] extracted the original MAS of the EEG signals within 2 s. Each original MAS is a feature matrix of $18 \times 19 \times 1$. Here, 18 means the number of channels of EEG signals; 19 means the number of sub-bands of the EEG signal in each channel, which indicates the 0–70 Hz frequency band of the EEG signal of each channel is divided into 19 sub-bands [26]. Specifically, each of δ, θ, and α is divided into three sub-bands; each of β and γ is divided into five sub-bands; thus, they are a total of 19 sub-bands. The sub-bands of the original MAS extracted from these rhythms show promising results in the EEG representation of epilepsy.

In Fig. 2(b), we propose an improved MAS called extended series MAS. Different from the original MAS [26], the proposed extended series MAS is a MAS series in time continuity, defined as $S = [S_1, S_2, S_3]$. For $t = 1$, 2, and 3, S_t is an extracted MAS from the EEG signals in the tth second. Therefore, each S is extracted from the EEG signals within 3 s. Such extended series MAS can effectively fuse the temporal relevance of multiple MAS features to improve epilepsy detection.

Different from the above original MAS [26], each S_t, $t = 1$, 2, and 3, expands the number of channels of EEG signals from 18 to 23, namely, P1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10, FT10-T8, and T8-P8 [46]. Such channel extension

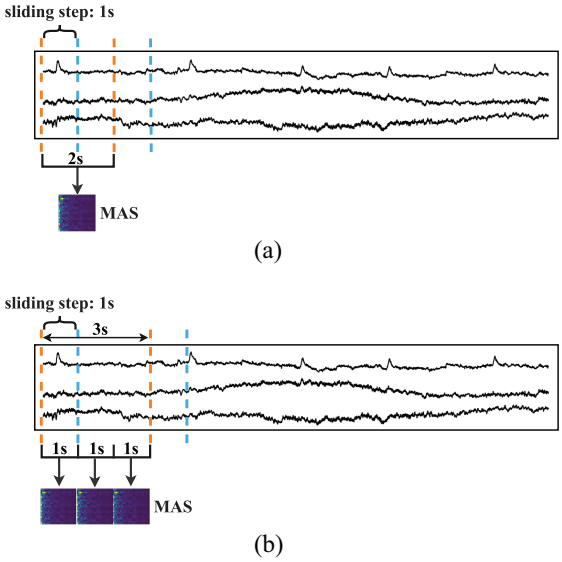


Fig. 2. MAS extraction method. (a) Traditional methods were used to extract MAS. (b) Our proposed MAS extraction with a time-series method.

aims to obtain more spatial feature information for improving epilepsy detection.

Unlike the above original MAS [26], each S_t, $t = 1$, 2, and 3, extends the number of sub-bands of the EEG signal in each channel to 25 from 19. Note that the low-frequency bands of EEG signals, such as δ, θ, and α contain abundant epileptic characteristic information. Therefore, we divide each of the low-frequency bands into five sub-bands instead of three sub-bands in [26], which fully refines the features of these low-frequency bands. On the other hand, the 60-Hz signal in the γ band is considered to contain power frequency interference. Therefore, the 57–63-Hz signal is filtered out during processing.

For the above reasons, each S_i is a feature matrix of $23 \times 25 \times 1$. Therefore, each $S = [S_1, S_2, S_3]$ is a feature matrix of $23 \times 25 \times 3$, which is a 3-D continuous-time MAS feature. To obtain each S_t, $t = 1$, 2, and 3, discrete Fourier transform (DFT) with 256 points is first performed on the EEG signal of each channel within 1 s as follows:

$$\begin{aligned} X(k) &= \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi j}{N}kn} \\ k &= 0, 1, 2, \ldots, N \end{aligned} \tag{1}$$

where $x(n)$ represents the sampled signal from one EEG channel before DFT transformation; $X(k)$ represents the data after DFT transformation; N represents the length of each frame sampled; e is the base of the natural logarithm; and j is an imaginary unit.

Here, define $S_t = [S_{t,1}; S_{t,2}; \ldots; S_{t,23}]$, $t = 1$, 2, 3 and S_t, $u = [s_{t,u,1}, s_{t,u,2}, \ldots, s_{t,u,25}]$, $u = 1, 2, \ldots, 23$. Each $s_{t,u,v}$, $t = 1$, 2, 3, $u = 1$, 2 , …, 23, and $v = 1, 2, \ldots, 25$ can be calculated as

$$\begin{aligned} s_{t,u,v} &= \text{avg}(P(k)), k \in K_v \\ P(k) &= |X_k| \end{aligned} \tag{2}$$

TABLE I
STRUCTURE OF SPATIAL FEATURE EXTRACTION MODULE

layer name	output size	Convolution kernel parameter
conv 1	23×25	$Conv(3\times3,64)$
ResBlock1	23×25	$\begin{bmatrix} Conv(3\times3,64) \\ Conv(3\times3,64) \\ SEBlock \end{bmatrix}\times 2$
ResBlock2	12×13	$\begin{bmatrix} Conv(3\times3,128) \\ Conv(3\times3,128) \\ SEBlock \end{bmatrix}\times 2$
ResBlock3	6×7	$\begin{bmatrix} Conv(3\times3,256) \\ Conv(3\times3,256) \\ SEBlock \end{bmatrix}\times 2$
ResBlock4	3×4	$\begin{bmatrix} Conv(3\times3,512) \\ Conv(3\times3,512) \\ SEBlock \end{bmatrix}\times 2$
Average pool	1×1	$Pooling(3\times4)$

where t, u, and v represent the tth MAS feature of each extended series MAS S at the tth second, the uth EEG channel, and the vth EEG sub-band, respectively. K_v denotes the set of values of k of the uth sub-band. $P(k)$ means the amplitude spectrum of X_k.

In general, the proposed extended series MAS uses three extended MASs in time continuity; the number of sub-bands and channels of each extended MAS are expanded to 25 and 23, respectively. Therefore, the proposed extended series MAS can effectively fuse the temporal relevance of multiple MAS features and the spatial relevance of multiple EEG channels, fully representing the related activities of the brain in the detection of preseizure.

C. Proposed STCARN

There are temporal and spatial relationships of MAS features between different channels. For each channel, there are also temporal relationships of MAS features. Therefore, STCARN is developed to obtain the spatiotemporal relationships of the above MAS features via combining spatiotemporal feature extraction modules as follows.

1) Spatial Feature Extraction Module: To effectively extract the spatial correlation of MAS features between different EEG channels, the spatial feature extraction module of STCARN is developed. Fig. 1 shows that the spatial feature extraction module involves three residual blocks in parallel. The network structure of each residual block is listed in Table I, involving one convolution layer (conv 1), eight ResBlocks, and one average pool layer in cascade. The eight ResBlocks include two ResBlock1, two ResBlock2, two ResBlock3, and two ResBlock4, which are connected in cascade. Due to the small feature dimension of the extended series MAS, more cascading ResBlocks may lead to overfitting. Therefore, only eight ResBlocks are used to extract spatial features.

Table I lists the network structures of eight ResBlocks, each of which consists of two convolutional layers with a residual connection and SEBlock. Taking ResBlock1 as an example, its Conv(3 × 3, 64) represents one convolutional layer with 64 convolution kernels and 3 × 3 kernel size; its SEBlock means a channel attention mechanism called squeeze–excitation (SE) module. More details involving the entire spatial feature extraction module are as follows.

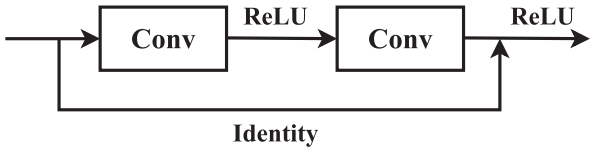


Fig. 3. Residual connection of the two convolutional layers of ResBlock1.

First, the extended series MAS $S \in \mathbb{R}^{23\times25\times3}$ is decomposed into $S_t(t=1,2,3) \in \mathbb{R}^{23\times25}$, each of which is entered into the corresponding residual block in Table I. Without loss of generality, for each residual block, S_t is first sent to one convolution layer (conv 1) with a size of $H' \times W'$ (3 × 3) and the number of convolution kernels of C' (64). After passing through a convolution layer, S_t is mapped to $X_t \in \mathbb{R}^{H'\times W'\times C'}$ as

$$X_t = \delta(W_t S_t), t = 1, 2, 3 \tag{3}$$

where W_t represents the learnable convolution parameter (weight) that maps the MAS feature sample S_t of the tth second to the output X_t; and δ represents the ReLU activation function.

Then, $X_t \in \mathbb{R}^{H'\times W'\times C'}$ is sent to eight ResBlocks, which aims to further extract the spatial information of MAS features. Table I lists the network structure of eight ResBlocks, each of which consists of two convolutional layers with a residual connection and SEBlock. Although the eight ResBlocks have different numbers and sizes of convolution kernels, they have the same network structure. Therefore, without loss of generality, we only provide the computational process of a ResBlock (ResBlock1) as follows.

For ResBlock1, X_t is first sent to two convolutional layers with a residual connection. Fig. 3 shows the residual connection of the two convolutional layers of ResBlock1. Here, the number of filters of the two convolution layers is set as C. BatchNorm is performed after each convolution layer. After passing through two convolutional layers, the input feature $X_t(t=1,2,3)$ can be mapped to $U_t(t=1,2,3)$ as follows:

$$U_t = F_{tr}(X_t) = W_{t2}\delta(W_{t1}X_t), t = 1, 2, 3 \tag{4}$$

where δ represents the ReLU activation function, and W_{t1} and W_{t2} represent the weights in the two-layer convolution kernel connecting the input X_t to the output U_t, respectively.

Next, the self-attention mechanism (SEBlock) is performed to obtain the important feature channels of EEG channels. Fig. 4 shows the network structure of SEBlock, which consists of squeeze and excitation. First, the squeeze operation for U_t compresses the feature map on the space dimension. The squeeze operation compresses the global space information into the channel. $Z \in \mathbb{R}^C$ is achieved by reducing the spatial dimension $H \times W$ of $U_t = [u_1, u_2, \ldots, u_C]$. The formula for calculating z of channel c is

$$z_c = F_{sq}(u_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j),\ c \in [1,2,\ldots,C] \tag{5}$$

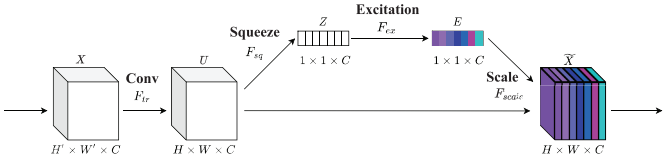


Fig. 4. SE module.

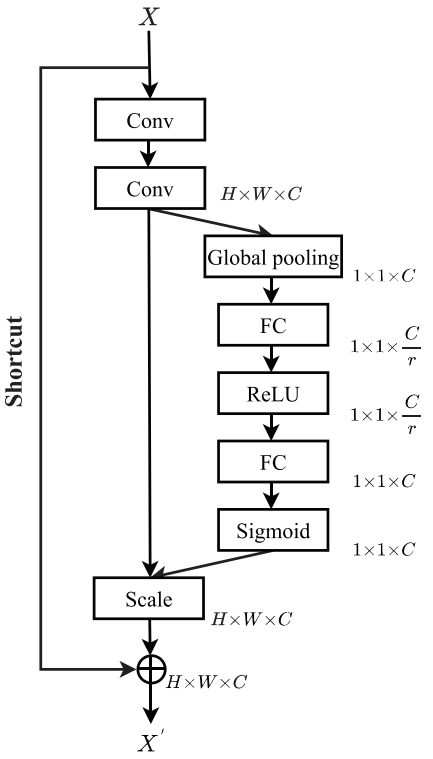


Fig. 5. Entire network structure of ResBlock1.

where u_c is the feature mapping of channel c, and $Z = [z_1, z_2, \ldots, z_C]$.

Fig. 5 shows the entire network structure of ResBlock1. It can be seen that after average global pooling, a gating mechanism composed of the sigmoid activation function is used to fully extract the dependencies between channels, reduce the model complexity, and improve the generalization. To parameterize the gating mechanism, two fully connected (FC) layers are constructed around the nonlinearity, as shown in Fig. 5. Specifically, the first FC layer is called a dimension reduction layer with parameter W_1, reduction rate r, and ReLU activation function. The second is a dimension increase layer with parameter W_2. Therefore, the gating mechanism is expressed as

$$E = F_{ex}(Z, W) = \sigma(g(Z, W)) = \sigma(W_2\delta(W_1 Z)) \tag{6}$$

where $W_1 \in \mathbb{R}^{(C/r)\times C}$, $W_2 \in \mathbb{R}^{C\times(C/r)}$, δ represents the ReLU activation function, and $E = [e_1, e_2, \ldots, e_C]$.

In the ResBlock1, the final output of SEBlock is obtained by rescaling u using active E

$$\widetilde{x}_c = F_{\text{scale}}(u_c, e_c) = u_c * e_c, \ c \in [1, 2, \ldots, C] \tag{7}$$

where $u_c \in \mathbb{R}^{H\times W}$ and $\widetilde{X}_t = [\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_C](t = 1, 2, 3)$. $\widetilde{X}_t$ refers to the extracted features of attention, then $\widetilde{X}_t$ and X_t

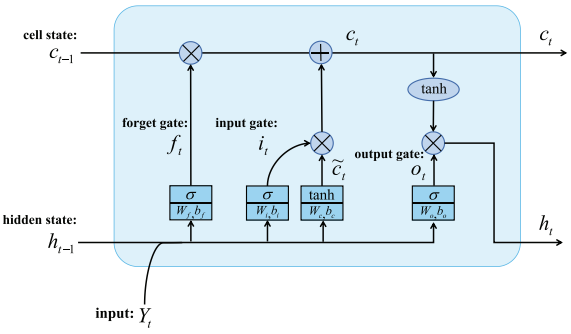


Fig. 6. Structure of the LSTM cell.

are residually connected to obtain the final output $X_t{}'$ of the ResBlocks1, as follows:

$$X_t{}' = \widetilde{X}_t + X_t. \tag{8}$$

Note that each residual block in Table I has eight ResBlocks. When $X_t(t = 1, 2, 3) \in \mathbb{R}^{H'\times W'\times C'}$ passes through the cascading eight ResBlocks, the final output feature map $X'_t(t = 1, 2, 3) \in \mathbb{R}^{H\times W\times C}$ can be obtained. Then, $X'_t(t = 1, 2, 3)$ is sent to an average pooling layer to reduce the number of parameters and further extract features and set the filter core size of the average pooling layer to $H \times W$. The tth coded representation after pooling is $\{Y_t | Y_t = \text{MaxPool}(X_t{}'), t = 1, 2, 3\}$, $Y_t \in \mathbb{R}^C$.

2) Temporal Feature Extraction Module: To capture this continuous-time relationship of MAS features, the temporal feature extraction module is developed using LSTM [47], [48]. Fig. 6 shows the structure of the temporal feature extraction module (LSTM), which is based on the recurrent network. Each LSTM cell receives three inputs, namely, Y_{t-1} (previous cell state), h_{t-1} (previous hidden layer state), and Y_t (current input). The corresponding outputs are c_t (current cell state) and h_t (current hidden layer state). LSTM includes the forget gate, input gate, and output gate. The function of the forget gate is to decide what information should be discarded or retained as follows:

$$f_t = \sigma\left(w_f \cdot \left[h_{t-1}, Y_t\right] + b_f\right). \tag{9}$$

The input gate is used to update the cell state as follows:

$$i_t = \sigma\left(w_i \cdot \left[h_{t-1}, Y_t\right] + b_i\right) \tag{10}$$

$$\widetilde{c}_t = \tanh\left(w_c \cdot \left[h_{t-1}, Y_t\right] + b_c\right). \tag{11}$$

The calculation method of the unit state is

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c}_t. \tag{12}$$

The function of the output gate is to calculate the state of the hidden layer, which contains the previously entered information. Its calculation can be expressed as

$$o_t = \sigma\left(w_o \cdot \left[h_{t-1}, Y_t\right] + b_o\right)$$
$$h_t = o_t * \tanh(c_t) \tag{13}$$

where the number of LSTM units of each layer is set to 3. This indicates that the time step is set to 3. In other words, the LSTM unit receives the features of each second within 3 s.

The output of each time step can be regarded as the time information extracted from each sample. We use two stacked layers to memorize and encode all scanned spatiotemporal regions [49].

$Y_t(t = 1, 2, 3) \in \mathbb{R}^C$ after the maximum pooling layer is spliced into $Y = [Y_1, Y_2, Y_3]$ as the input of the LSTM cell. The feature length of $Y_t(t = 1, 2, 3)$ is C, and the output after two LSTM layers is $Y' = [Y'_1, Y'_2, Y'_3]$, where $[Y'_1, Y'_2, Y'_3]$ represents the output of each hidden layer of the outermost LSTM.

3) Result Layer: After extracting time features from the LSTM of two layers, the feature Y' is mapped to Q through an FC layer. Then, Q is applied to the softmax layer to learn the probability belonging to each category. The output of this layer is a probability vector, whose dimension is equal to the number of classes as follows:

$$p(j) = \frac{e^{q_j}}{\sum_{j=1}^{l} e^{q_j}} \tag{14}$$

where $l(l = 1, \ldots, L)$, q_j, and $p(j)$ represent the total number of categories, the input of the jth neuron in the softmax layer, and the probability that the sample belongs to the jth category, respectively.

The loss function f of STCARN in epilepsy classification is shown as follows:

$$f = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{l}\widehat{y}_{ij}(i,j)\log\left(p_{ij}\right) \tag{15}$$

where $\widehat{y}_{ij}$ is the actual label of the EEG sample; p_{ij} is the predictive label of EEG samples; n is the number of samples; and l is defined as the number of types of epilepsy recognition output results. We use backpropagation to adjust network parameters and structure.

III. Experiments and Discussion

Our proposed STCARN was evaluated with the CHB-MIT data set [45]. Such data set involves the EEG recordings of pediatric patients with refractory epileptic seizures from Children's Hospital Boston, including 23 medical EEG records of 22 subjects. All signals were sampled at a rate of 256 samples per second with 16-bit resolution. In the CHB-MIT data set, most data contains EEG signals of 23 channels, however, a few data include EEG signals of 24 or 26 channels. These EEG signals are recorded using an international 10–20 EEG electrode location and naming system.

Three types of different MAS features were extracted to make a comparison. First, [26], [27], and [28] extracted MAS maps with EEG signals in a 2-s time window, and the sliding step of the window was 1 s. The number of channels and sub-bands of each extracted MAS map is 19 and 18, respectively. Second, we also adopted a similar method [26], [27], [28] to extract MAS images with EEG signals in a 2-s time window, and the sliding step of the window was 1 s. However, the number of sub-bands is expanded from 19 to 25, and the number of channels of data is expanded from 18 to 23. Therefore, the size of the MAS map is set to $23 \times 25 \times 1$ dimensions. Finally, we select the EEG signal with a time window of 3 s and extract MAS images for every second within 3 s. The sliding step of the window is also 1 s. Therefore, the MAS feature image with a dimension of $23 \times 25 \times 3$ is obtained.

STCARN is compared with some current state-of-the-art algorithms [26], [27], [28]. All network models use the SGD algorithm with learning momentum (SGDM) for parameter training. Momentum (Momentum factor) is 0.9, weight attenuation is 0.0004, maximum learning epoch is 240, and cosine annealing of the learning rate makes the learning rate change according to the cosine period.

Note that because of the space limitation, the number of samples for each compared model and the evaluation of time and space complexity for STCARN are given in Sections S-I and S-II of the supplementary file, respectively.

A. Three Categories

Similar to [26], the data set is divided into three categories: interictal, preictal, and seizure data. Interictal data must be at least 4 h apart from seizure data; preictal data must be within 1 h of the onset. For the three classifications, we generate sample data of the interseizure state, preseizure state, and seizure state.

From the experiments of MAS $25 \times 23 \times 1$ and MAS $25 \times 23 \times 3$ feature data, we extracted interictal, preictal, and seizure data of 10 791, 11 111, and 8124, respectively.

The fivefold cross-validation method was used to divide data and test performance. In order to ensure a balanced distribution of epileptic signals between different states, the data division within each category also roughly follows the above cross-validation ratio.

In order to prove the effectiveness of STCARN with the extended series MAS, the following methods are used to perform the test.

1) MAS $19 \times 18 \times 1$ + KNN: This is the method proposed in [27]. The MAS feature of 19 sub-bands from 18 channels was extracted in 2 s. Since there is no information in the time dimension, the classifier uses KNN for classification.

2) MAS $19 \times 18 \times 1$ + SCNN: This method is proposed by [27]. The MAS features of 19 sub-bands from 18 bands were extracted in 2 s, but the stacked CNN network structure is adopted for feature learning, and the features of different networks are fused. Finally, softmax is used for classification.

3) MAS $25 \times 23 \times 1$ + KNN: The proposed MAS extraction method of extended channels and sub-bands is used for MAS of 25 sub-bands of 23 channels that were extracted in 2-s time. The classifier uses KNN for the classification.

4) MAS $25 \times 23 \times 1$ + ResNet18: The proposed MAS extraction method of extended channels and sub-bands is used for the MAS of 25 sub-bands of 23 channels that were extracted in 2-s time. But, ResNet18 is used for the network model and softmax is used for the classification.

5) MAS $25 \times 23 \times 3$ + STCARN: The extended series MAS feature representation covers 25 sub-bands features from each of 23 channels in 3-s time. STCARN is used to extract spatiotemporal features with the softmax classifier.

Table II shows the average recognition accuracy of the proposed method and the compared four methods in epilepsy

TABLE II
EPILEPSY THREE CLASSIFICATION ACCURACY

Feature	Models	accuracy
MAS 19×18×1	KNN	97.52±0.13%
MAS 19×18×1	SCNN	98.62±0.01%
MAS 25×23×1	KNN	99.21±0.0006%
MAS 25×23×1	ResNet18	99.77±0.01%
MAS 25×23×3	STCARN	**99.98±0.01%**

three classification task. It can be seen that KNN with MAS features of 25 × 23 × 1 dimension obtains higher average accuracy than KNN and SCNN with MAS features of 19 × 18 × 1 dimension. This implies that the method of expanding the number of channels and sub-bands is effective when extracting MAS features. In addition, when all MAS images with a dimension of 25 × 23 × 1 are used as input data, the accuracy of ResNet18 is about 0.56% higher than that of the KNN algorithm. Finally, the recognition accuracy of the STCARN algorithm on the extracted MAS image with a dimension of 25 × 23 × 3 is as high as 99.98%, reaching the current best performance. This also shows that the STCARN model is beneficial for learning the temporal continuity of MAS 25 × 23 × 3 images.

B. Five Categories

The EEG feature signals of 1 h before epileptic seizure were subdivided into five subcategories, namely, epileptic seizure, interictal, and preictal (PreI, PreII, and PreIII). In our proposed STCARN experiment, the number of samples of five categories of epilepsy was 8111, 10 791, 9497, 9239, and 10 597, respectively. PreI, PreII, and PreIII states indicate EEG signals 40–60, 20–40, and 0–20 min before the epileptic seizure, respectively. In order to prove the effectiveness of the proposed algorithm with the extended series MAS, we choose the following test methods.

1) MAS 19 × 18 × 1 + ResNet152: In [28], MAS features were extracted from 19 sub-bands of 18 channels in 2 and 10 s time windows. ResNet152 was used for feature learning. Softmax was used for state classification prediction and epilepsy detection.

2) MAS 19 × 18 × 1 + TFCNN: This method is proposed by [29]. The MAS features of 19 sub-bands from 18 bands were extracted in 2 s. But, the stacked CNN model is changed into AlexNet, VGG19, Inception-V3, ResNet152, and Inception-resnet-v2. The classifier adopts a hierarchical neural network for classification.

3) MAS 25 × 23 × 1 + ResNet152: The proposed MAS extraction method of extended channels and sub-bands is used for the MAS of 25 sub-bands of 23 channels that were extracted in 2-s time. ResNet152 was used for feature learning with the softmax classifier.

4) MAS 25 × 23 × 1 + ResNet18: The proposed MAS extraction method of extended channels and sub-bands is adopted to extract MAS of 25 sub-bands of 23 channels in 2-s time. The learning network structure is ResNet18, which is relatively simple and avoids overfitting. The classifier is softmax.

5) MAS 25 × 23 × 3 + STCARN: The extended series MAS feature representation covers 25 sub-bands features from each of 23 channels in 3-s time. STCARN is used to extract spatiotemporal features with the softmax classifier.

Table III shows the average recognition accuracy of the proposed method and the compared methods in the five classification tasks of epilepsy. In [26], the MAS features were extracted in 2 and 10 s. The corresponding classification accuracy of ResNet152 reached 87.82% and 92.2%, in 2 and 10 s, respectively. MAS 19 × 18 × 1 + SCNN [27] obtains an accuracy rate of 87.95%. However, MAS 25 × 23 × 1 + ResNet152 obtains the average recognition accuracy of 93.13% for detecting epileptic state, which is significantly better than MAS 19 × 18 × 1 + ResNet152. This further shows that it is effective to expand the number of channels and sub-bands of MAS features.

In addition, it can be seen that ResNet18 on MAS 25 × 23 × 1 obtains an accuracy of 93.41%, which is about 0.28% higher than ResNet152. This indicates that a shallower Resnet network can be sufficient to extract useful feature information. Finally, when using the extended series MAS of 25 × 23 × 3, STCARN reaches an accuracy of 99.51%, which is the current best performance among all compared models. This proves that STCARN has certain advantages in learning the feature of MAS maps with time dimensions.

In Fig. 7, for each compared model, the diagonal of each corresponding confusion matrix represents the probability of each category being correctly classified; each off-diagonal element (i, j) represents the percentage of samples of type i that are misclassified as type j (misclassification rate). Seizure and interictal represent the seizure and intermittent periods of epilepsy, respectively. Each of PreI, PreII, and PreIII means the preseizure period. PreIII, PreII, and PreI are the first, second, and third longest time before the seizure. As can be seen in Fig. 7, STCARN achieves the highest accuracy in the seizure, the interictal, the PreII, and the PreIII.

For the detection of PreI, the accuracy of STCARN is worse than that of MAS 25 × 23 × 1 + ResNet152 and MAS 25 × 23 × 1 + ResNet18. However, the difference between them is very small and almost negligible. Especially in the prediction of PreI and PreII, the accuracy of STCARN is 99.07% and 99.67%, respectively, indicating that STCARN has a significant advantage over the other five models. Therefore, STCARN is more suitable for preseizure detection than other models.

To show the strong feature learning ability of the proposed STCARN, we compare the probability vector obtained by the softmax layer of five different methods. The experiment was performed on the CHB-MIT data set. The samples of five epileptic states were selected for evaluation, namely, seizure, interictal, PreI, PreII, and PreIII. Fig. 8 is the output diagram of the softmax layer of the compared five models: 1) STCARN; 2) MAS 25 × 23 × 1 + ResNet18; 3) MAS 25 × 23 × 1 + ResNet152; 4) MAS 19 × 18 × 1 + SCNN; and 5) MAS 19 × 18 × 1 + TFCNN. It can be seen that our proposed STCARN has obtained the highest prediction probability in each category of seizure, interictal, PreI, PreII, and PreIII.

TABLE III
PERFORMANCE COMPARISONS ON FIVE-CATEGORY EPILEPTIC CLASSIFICATION OF THE CHI-MIT DATABASE

Feature	Models	PreI	PreII	PreIII	Interictal	Seizure	accuracy
MAS 19×18×1(2S)	ResNet152	-	-	-	-	-	87.82%
MAS 19×18×1(10S)	ResNet152	94.52%	94.85%	90.98%	91.38%	89.27%	92.20%
MAS 19×18×1	KNN	84.53%	71.15%	73.96%	97.74%	95.51%	84.56±0.38%
MAS 19×18×1	SCNN	85.59%	76.95%	80.52%	98.57%	98.15%	87.95±0.16%
MAS 19×18×1	TFCNN	98.79%	96.24%	97.07%	95.08%	97.65%	96.97±0.17%
MAS 25×23×1	ResNet152	99.21%	87.71%	85.71%	99.77%	93.48%	93.13±0.13%
MAS 25×23×1	KNN	90.06%	77.01%	65.95%	99.81%	90.19%	85.96±0.33%
MAS 25×23×1	ResNet18	**99.48%**	89.50%	89.50%	84.21%	92.88%	93.41±0.27%
MAS 25×23×3	STCARN	99.04%	**99.07%**	**99.67%**	**100%**	**100%**	**99.51±0.15%**

TABLE IV
COMPARISON OF MAS 25 × 23 × 3 FEATURE IN DIFFERENT MODELS

Models	PreI	PreII	PreIII	Interictal	Seizure	accuracy
ResNet18	94.58%	94.63%	98.05%	99.95%	99.82%	97.6±0.09%
EfficientNet_B0	94.77%	91.88%	95.77%	99.95%	99.36%	97.14±0.12%
EfficientNet_B3	93.26%	91.8%	95.44%	100.0%	99.63%	96.29±0.21%
SVM	60.02%	49.05%	75.77%	99.86%	94.5%	77.53±0.01%
KNN	74.19%	60.72%	84.49%	99.95%	89.40%	80.91±0.01%
LSTM	93.57%	90.91%	94.97%	99.77%	99.31%	95.77±0.28%
STCARN	**99.04%**	**99.07%**	**99.67%**	**100%**	**100%**	**99.51±0.15%**

Fig. 7. Confusion matrix comparisons to state-of-the-art algorithms on the 5-category CHB-MIT database. (a) MAS 19 × 18 × 1 + ResNet152(10S). (b) MAS 19 × 18 × 1 + SCNN. (c) MAS 19 × 18 × 1 + TFCNN. (d) MAS 25 × 23 × 1 + ResNet152. (e) MAS 2 5× 23 × 1 + ResNet18. (f) MAS 25 × 23 × 3+ STCARN.

In order to further evaluate the performance of STRCNN, it is compared with six different models, including ResNet18, LSTM, EfficientNetB0, EfficientNetB3, SVM, and KNN on MAS 25 × 23 × 3. The experimental results are listed in Table IV. An interesting observation is that our STRCNN model achieves the best detection accuracy. This indicates that STRCNN can better extract spatiotemporal features for improving the detection accuracy.

C. Evaluation of Three Innovations

To evaluate the effectiveness of the proposed STCARN, an ablation experiment was carried out. Two variants of the

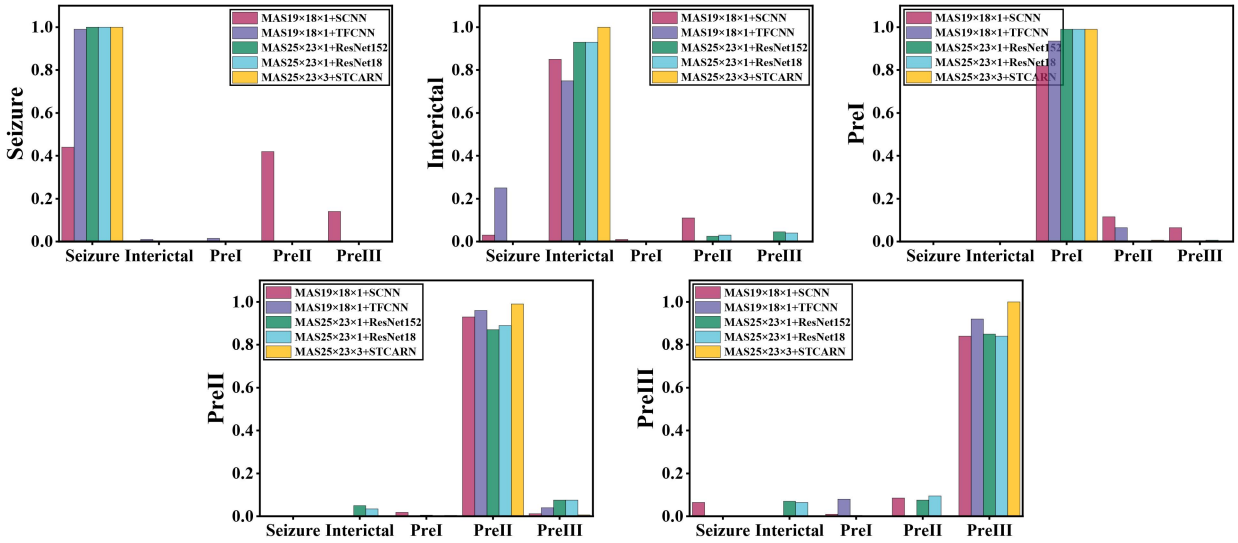


Fig. 8. Probabilities obtained in the Softmax layer.

TABLE V
BASELINE MODEL AND IMPROVED MODEL FOR EPILEPSY RECOGNITION

Model	residual convolution	self-attention	LSTM
STCARN01	✓	×	×
STCARN02	✓	×	✓
STCARN	✓	✓	✓

TABLE VI
COMPARISON BETWEEN STCARN AND ITS TWO VARIANTS

Feature	Models	accuracyc
MAS25×23×3	STCARN01	97.6±0.09%
MAS25×23×3	STCARN02	99.39±0.12%
MAS25×23×3	STCARN	**99.51±0.1%**

STCARN structure were designed experimentally. The first variant of STCARN is called STCARN01, where the attention mechanism and LSTM module were not used to learn the features of MAS 25 × 23 × 1 maps and detect epilepsy state. The second variant of STCARN is called STCARN02, where the attention mechanism was not used. The detailed information on these compared models is listed in Table V.

Table VI lists the average recognition accuracy of STCARN01, STCARN02, and STCARN on MAS 25 × 23 × 1 feature for the epilepsy five classification task. STCARN01 provides an average accuracy of 97.6%, while STCARN02 does that at 99.39%. STCARN02 has a 1.97% higher detection accuracy than STCARN01. This is because STCARN02 rationally combines residual convolution and LSTM to effectively obtain the spatiotemporal features of EEG signals. However, STCARN01 uses only residual convolution to obtain the spatial features. Furthermore, the proposed STCARN has a 0.12% higher accuracy than STCARN02. This is because STCARN uses the channel attention mechanism to enhance the effective channel weights and decline the unimportant channel weights compared with STCARN02. This also demonstrates the effectiveness of the recurrent network (LSTM) and channel-attention mechanisms for improving the performance of STCARN. In general, because the proposed STCARN rationally combines residual convolution, RNN, and channel attention mechanism, it can provide extremely competitive detection performance in epilepsy detection.

IV. CONCLUSION

In this article, we propose a data-driven STCARN model with extended series MAS for EEG-based epilepsy detection. Unlike traditional epilepsy detection studies, which focus on seizure/nonseizure detection, STCARN can obtain extremely promising performance on both seizure and preseizure lasting 60 min. The experimental research and comparison with existing state-of-the-art methods demonstrate the effectiveness of STCARN. The main reasons behind this fact are given as follows: 1) the extended series MAS representation involves multiple MAS features in continuous time, which can effectively represent the related spatiotemporal information of the EEG signals of the activities of the brain and 2) STCARN covers the spatial and temporal feature extraction modules, which can map the extended series MAS features for extracting the spatiotemporal feature information of EEG signals, respectively. The spatial extraction module applies three residual networks with the channel attention mechanism to effectively extract a series of MAS features in related time, instead of only one MAS feature. The temporal extraction module utilizes LSTM to extract features of the extended series MAS, which can obtain the effective temporal relationship features between different EEG channels. Therefore, through rationally combining the spatial and temporal feature extraction modules with the extended series MAS features, STCARN receives 99.98% and 99.51% accuracy in the three-classification and five-classification tasks, respectively, for the detection of preseizure. In the future, we will explore how to efficiently detect epilepsy using a very small number of EEG channels. Also, we will further explore how to significantly reduce the time and space complexity of the STCARN model while maintaining the high epilepsy detection accuracy of the model.

REFERENCES

[1] "Reports by world health organization (WHO)." 2019. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/epilepsy

[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] S. Noachtar and J. Rémi, "The role of EEG in epilepsy: A critical review," *Epilepsy Behav.*, vol. 15, no. 1, pp. 22–33, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1525505009000924

[4] 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, May 1972. [Online]. Available: https://doi.org/10.1001/archneur.1972.00490110043004

[5] J. Gotman, "Automatic recognition of epileptic seizures in the EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 54, no. 5, pp. 530–540, 1982. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0013469482900384

[6] L. Guo, D. Rivero, and A. Pazos, "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks," *J. Neurosci. Methods*, vol. 193, pp. 156–163, Oct. 2010.

[7] 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, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482513000905

[8] S. Ouelha and B. Boashash, "Automatic signal abnormality detection using time-frequency features and machine learning: A newborn EEG seizure case study," *Knowl. Based Syst.*, vol. 106, pp. 38–50, Aug. 2016.

[9] 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.

[10] Y. Qiu, W. Zhou, N. Yu, and P. Du, "Denoising sparse autoencoder-based ictal EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 9, pp. 1717–1726, Sep. 2018.

[11] J.-L. Song and R. Zhang, "Application of extreme learning machine to epileptic seizure detection based on lagged poincaŕe plots," *Multidimensional Syst. Signal Process.*, vol. 28, pp. 945–959, Jul. 2017.

[12] E. Q. Wu, Z.-R. Tang, P. Xiong, C.-F. Wei, A. Song, and L.-M. Zhu, "ROpenPose: A rapider openpose model for astronaut operation attitude detection," *IEEE Trans. Ind. Electron.*, vol. 69, no. 1, pp. 1043–1052, Jan. 2022.

[13] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482517303153

[14] X. Zhang, L. Yao, M. Dong, Z. Liu, Y. Zhang, and Y. Li, "Adversarial representation learning for robust patient-independent epileptic seizure detection," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 10, pp. 2852–2859, Oct. 2020.

[15] M. T. Avcu, Z. Zhang, and D. W. S. Chan, "Seizure detection using least EEG channels by deep convolutional neural network," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 1120–1124.

[16] 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.

[17] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[18] P. Bizopoulos, G. I. Lambrou, and D. Koutsouris, "Signal2Image modules in deep neural networks for EEG classification," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2019, pp. 702–705.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017. [Online]. Available: https://doi.org/10.1145/3065386

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, *arXiv:1409.1556*.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2261–2269.

[23] X. Chen, J. Ji, T. Ji, and P. Li, "Cost-sensitive deep active learning for epileptic seizure detection," in *Proc. ACM Int. Conf. Bioinform. Comput. Biol. Health Inform.*, 2018, pp. 226–235.

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[25] S. Roy, I. Kiral-Kornek, and S. Harrer, "ChronoNet: A deep recurrent neural network for abnormal EEG identification," in *Proc. Conf. Artif. Intell. Med. Europe*, 2019, pp. 47–56.

[26] W. Hu, J. Cao, X. Lai, and J. Liu, "Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks," *J. Ambient Intell. Humanized Comput.*, pp. 1–11, Feb. 2019. [Online]. Available: https://doi.org/10.1007/s12652-019-01220-6

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

[28] Y. Wang, J. Cao, J. Wang, D. Hu, and M. Deng, "Epileptic signal classification with deep transfer learning feature on mean amplitude spectrum," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2019, pp. 2392–2395.

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

[30] R. Peng, J. Jiang, G. Kuang, H. Du, D. Wu, and J. Shao, "EEG-based automatic epilepsy detection: Overview and outlook," *Acta Automatica Sinica*, vol. 48, no. 2, pp. 335–350, 2022.

[31] J. Martinerie et al., "Epileptic seizures can be anticipated by non-linear analysis," *Nat. Med.*, vol. 4, pp. 1173–1176, Nov. 1998.

[32] L. D. Iasemidis and J. C. Sackellares, "Chaos theory and epilepsy," *Neuroscientist*, vol. 2, no. 2, pp. 118–126, 1996.

[33] P. Celka and P. Colditz, "A computer-aided detection of EEG seizures in infants: A singular-spectrum approach and performance comparison," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 5, pp. 455–462, May 2002.

[34] Y. Park, T. Netoff, and K. Parhi, "Seizure prediction with spectral power of EEG using cost-sensitive support vector machines," *J. Med. Devices*, vol. 4, no. 2, Aug. 2010, Art. no. 27542. [Online]. Available: https://doi.org/10.1115/1.3455144

[35] A. Berdakh and S. H. Don, "Epileptic seizures detection using continuous time wavelet based artificial neural networks," in *Proc. 6th Int. Conf. Inf. Technol. New Gener.*, Jan. 2009, pp. 1456–1461.

[36] A. S. Zandi, M. Javidan, G. A. Dumont, and R. Tafreshi, "Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1639–1651, Jul. 2010.

[37] K.-C. Hsu and S.-N. Yu, "Detection of seizures in EEG using subband nonlinear parameters and genetic algorithm," *Comput. Biol. Med.*, vol. 40, no. 10, pp. 823–830, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482510001204

[38] 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, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809410000054

[39] V. Joshi, R. B. Pachori, and A. Vijesh, "Classification of ictal and seizure-free EEG signals using fractional linear prediction," *Biomed. Signal Process. Control*, vol. 9, pp. 1–5, Jan. 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809413001195

[40] Y. Li, X.-D. Wang, M.-L. Luo, K. Li, X.-F. Yang, and Q. Guo, "Epileptic seizure classification of EEGs using time–frequency analysis based multiscale radial basis functions," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 2, pp. 386–397, Mar. 2018.

[41] Y. R. Aldana, B. Hunyadi, E. J. M. Reyes, V. R. Rodríguez, and S. Van Huffel, "Nonconvulsive epileptic seizure detection in scalp EEG using multiway data analysis," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 660–671, Mar. 2019.

[42] S. Khanmohammadi and C.-A. Chou, "Adaptive seizure onset detection framework using a hybrid PCA–CSP approach," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 154–160, Jan. 2018.

[43] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, and R. Boots, "EEG-based intention recognition from spatio-temporal representations via cascade and parallel convolutional recurrent neural networks," Aug. 2017, *arXiv:1708.06578*.

[44] Y. Yang et al., "Performance comparison of gesture recognition system based on different classifiers," *IEEE Trans. Cogn. Devel. Syst.*, vol. 13, no. 1, pp. 141–150, Mar. 2021.

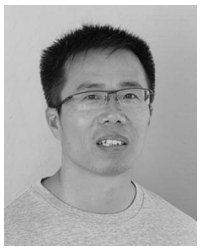
[45] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," M.S. thesis, Dept. MIT Division Health Sci. Technol., Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.
[46] J. Malmivuo and R. Plonsey, "Bioelectromagnetism. 13. Electroencephalography," *Bioelectromagnetism—Principles and Applications of Bioelectric and Biomagnetis Fields*. Oxford, U.K.: Oxford Univ. Press, Jan. 1995, pp. 247–264.
[47] W. Tao et al., "EEG-based emotion recognition via channel-wise attention and self attention," *IEEE Trans. Affect. Comput.*, early access, Sep. 22, 2022, doi: 10.1109/TAFFC.2020.3025777.
[48] E. Q. Wu, P. Xiong, Z.-R. Tang, G.-J. Li, A. Song, and L.-M. Zhu, "Detecting dynamic behavior of brain fatigue through 3-D-CNN-LSTM," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 1, pp. 90–100, Jan. 2022.
[49] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2019.

Qi Wang received the B.S. degree from Jiangxi University of Science and Technology, Nanchang, China, in 2020. He is currently pursuing the master's degree in computer technology with the Department of Computer Science, College of Mathematics and Computer Sciences, Nanchang University, Nanchang.

His current research directions are EEG signal processing and deep learning.

Chenxi Huang received the B.S. degree in computer science and technology from Minnan Normal University, Zhangzhou, China, in 2019. He is currently pursuing the master's degree in computer technology with the Department of Computer Science, College of Mathematics and Computer Sciences, Nanchang University, Nanchang, China.

His research interests include EEG-based emotion recognition, image recognition, and machine learning.



Qingpeng Zeng received the M.Sc. degree from Guizhou University, Guiyang, China, in 2004.

He has been with the College of Mathematics and Computer Sciences, Nanchang University, Nanchang, China, since 2004, where he is currently a Vice Professor. He has published over 20 research articles. His current interests include computing intelligence, and network and information security.

Chunquan Li (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees from Nanchang University, Nanchang, China, in 2002, 2007, and 2015, respectively.

He has been with the School of Information Engineering, Nanchang University since 2002, where he is currently a Professor and a Ganjiang Distinguished Professor. He has published over 50 research articles. He was also a Visiting Professor with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada. His current interests include computing intelligence, haptics, virtual surgery simulation, robotics, and their applications to biomedical engineering.

Ting Shu received the B.Sc., M.Sc., and Ph.D. degrees from Nanchang University, Nanchang, China, in 2006, 2010, and 2021, respectively.

She has been working with the Second Affiliated Hospital, Nanchang University since 2002, where she is currently an Attending Doctor with the Radiology Department. Her current interests include the brain imaging changes, brain function and structure changes in epilepsy machine learning techniques, and Bayesian classifier in detecting seizure focus.