IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 4, 2021, accepted November 24, 2021, date of publication December 1, 2021,
date of current version December 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3132024

Spatial-Temporal Neural Network for P300 Detection

ZHEN ZHANG[1], XIAOYAN YU[2], (Member, IEEE), XIANWEI RONG[2], AND MAKOTO IWATA[1], (Member, IEEE)

[1]School of Information, Kochi University of Technology, Kochi 7828502, Japan
[2]Department of Physics and Electronic Engineering, Harbin Normal University, Harbin 150025, China

Corresponding author: Zhen Zhang (248002m@gs.kochi-tech.ac.jp)

ABSTRACT P300 spellers are common brain-computer interface (BCI) systems designed to transfer information between human brains and computers. In most P300 detections, the P300 signals are collected by averaging multiple electroencephalographic (EEG) changes to the same target stimuli, so the participants are obliged to endure multiple repeated stimuli. In this study, a spatial-temporal neural network (STNN) based on deep learning (DL) is proposed for P300 detection. It detects P300 signals by combining the outputs from a temporal unit and a spatial unit. The temporal unit is a flexible framework consisting of several temporal modules designed for analyzing brain potential changes in the time domain. The spatial unit combines one-dimensional convolutions (Conv1Ds) and linear layers to generalize P300 features from the space domain, and it can decode EEG signals recorded using different numbers of electrodes. Both amyotrophic lateral sclerosis (ALS) patients and healthy subjects can benefit from this study. In the within-subject P300 detection and the cross-subject P300 detection, our approach gained higher performance with fewer repeated stimuli than other comparative approaches. Furthermore, we applied the proposed STNN in the P300 detection challenge of BCI Competition III. The accuracy score was 89% in the fifth round of repeated stimuli, outperforming the best result in the literature (accuracy = 80%) to the best of our knowledge. The results demonstrate that the proposed STNN performs well with limited stimuli and is robust enough for various P300 detections. Our model can be found at: https://github.com/Zhangzhenkut/STNN.

INDEX TERMS P300 detection, spatial-temporal neural network (STNN), deep learning (DL).

I. INTRODUCTION

Brain-computer interface (BCI) systems enable neural signals to control external devices directly. In recent years, BCIs have been applied in many fields, such as environmental control [1], communication [2], and neurofeedback rehabilitation [3]. Electroencephalography (EEG) monitoring is one of the most popular measurement tools in BCI applications because of its non-invasiveness, mobility, and relatively low cost [4].

The P300 speller, as an EEG-based BCI paradigm, was first proposed by Farwell and Donchin [5], as shown in Figure. 1. During the spelling, the participants are required to focus their gaze on the lighted characters when the rows or columns of 36 alphanumeric characters are randomly intensified. In this process, the participants' brain activity changes evoked by the target characters are called event-related potentials (ERPs). Within the ERPs, the P300 signal is one of the most robust components that corresponds to a positive deflection, occurring 250-500ms after a target presentation [6].

An efficient P300 detection technique is a valuable contribution for the BCI community. Humans, particularly amyotrophic lateral sclerosis (ALS) patients, who suffer from progressive physical disabilities caused by the degeneration of the motor neuron system [7], will benefit from this research. The challenges, however, are that EEG signals inherently have a low signal-to-noise ratio (SNR) and differ significantly between individuals. Even for the same individual, EEG changes can differ in responding to the same target stimuli when affected by internal states and external surroundings. Thus, we usually average multiple EEG responses to a target stimulus to weaken noise and highlight

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno Garcia.

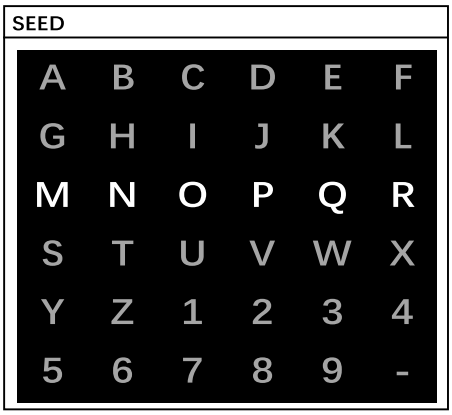


FIGURE 1. Farwell and donchin's paradigm.

features. However, it adds inconvenience for the participants, who are obliged to spend more time and endure multiple repeated stimuli for the same target. To cope with this challenge, researchers should make a reasonable tradeoff among time, cost, accuracy, and complexity when designing a P300 detection approach.

In this paper, we proposed a spatial-temporal neural network (STNN) for P300 detection. It performs better and is more robust in various P300 detections with limited data and repeated stimuli. Our main contributions in the proposed network are as follows: 1) We proposed a parallel network consisting of a temporal unit and a spatial unit to simultaneously learn spatial and temporal features from raw EEG signals; 2) In our design, the spatial unit is mainly constructed using Conv1Ds and linear layers. It can generalize the spatial features of EEG signals recorded using different numbers of electrodes (for example, 8, 16, or 64); 3) We designed a temporal unit inspired by [8]. This unit is used to analyze brain potential changes in the time domain by stacking multiple temporal modules. The number of the temporal modules can be adjusted according to the corresponding P300 detection task.

We demonstrate the effectiveness of our model using three public databases: P300 speller with ALS patients [9], covert and overt ERP-based BCI [10], and BCI Competition III-dataset II [11].

The remainder of this paper is structured as follows. Section II introduces related work; The description of databases and data preprocessing procedures are presented in Section III; Section IV details the proposed STNN; the results and discussion are in Sections V and VI; and Section VII concludes the paper.

II. RELATED WORK

The current mainstream P3000 detection approaches can be categorized into two types: deep learning (DL) and traditional technologies using statistical features and classifiers. In the traditional ones, the feature extraction mainly includes measures such as independent component analysis (ICA) [12], canonical correlation analysis (CCA) [13], common spatial patterns (CSP) [14], and XDAWN spatial filter [15]. Commonly used classifiers include linear discriminant analysis (LDA) [16], support vector machine (SVM) [17], and Riemannian geometry classifier (RGC) [18], among others. Of these, the combination of XDAWN and RGC is perhaps the most potent approach for P300 detection [19], which exhibits a strong generalization capability for variable EEG signals. Nevertheless, it is still not as competitive as DL approaches [20].

Convolutional neural network (CNN) as a representative DL framework [21]–[26] has attracted widespread attention from the BCI community. In 2010, Cecotti *et al.* [23] first proposed a CNN-based P300 detection approach that won the third BCI competition. This method adopts a four-layer CNN architecture to extract channel features and temporal features in sequence, demonstrating that CNN can capture both spatial peculiarities and latent serial dependencies from EEG signals. However, although CNN improved the detection accuracy to an unprecedented level, there are still two major obstacles that lie ahead for such methods. Firstly, the network accuracy depends on the quality and quantity of training data, while the amount of high-quality data commonly remains limited in P300 tasks because of the high cost of time and labor. Secondly, the P300 response is a relatively small potential change presented at a high resolution in the time domain [24], yet the CNN-based frameworks are not skilled at decoding sequential information with limited EEG data.

To resolve the above problems, some of the recent DL approaches tend to strengthen the learning capability of neural networks when limited data are available, such as [27]–[29], or adopt more advanced architectures to optimize the feature extraction procedure, such as [30]–[33]. EEGNet [33] as a generic DL network implemented by depth-wise and separable convolutions is proposed, which yields the satisfactory results in various EEG detections. This network extracts temporal features from the EEG signals firstly and then performs spatial filtering on each temporal feature map. With this design, the network can directly perform sequential learning using raw EEG signals and then generalize the captured dependencies in the space domain. It is more competitive for P300 detection than other DL-based pure sequence models, such as recurrent neural networks and long-short-term memory networks. However, this network relies on multiple repeated stimuli to collect EEG signals.

III. MATERIALS

A. DATASET 1: P300 SPELLER WITH ALS PATIENTS

In Dataset 1 [9], EEG signals from eight ALS patients (five males and three females, mean age $= 59.7 \pm 12.3$ years) were recorded using BCI2000 [34] and Farwell and Donchin's paradigm (Figure. 1). The EEG signals were digitized at 256 Hz from eight channels (Fz, Cz, Pz, Oz, P3, P4, PO7, and PO8) according to 10-10 standard [35] and bandpass filtered between 0.1 and 30 Hz.

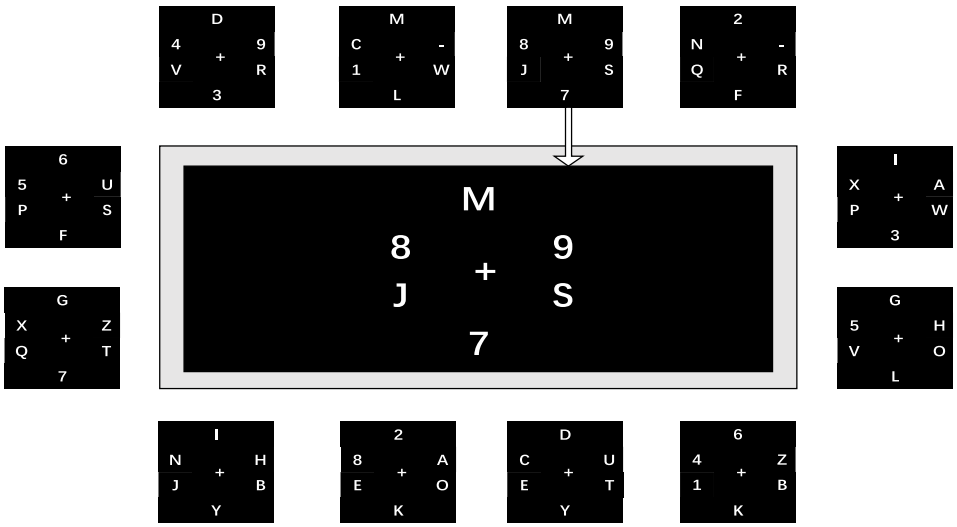


FIGURE 2. GeoSpell (geometric speller) paradigm.

Every participant in the study went through 35 trials, with 10 rounds of repeated stimuli in each trial. Every round of stimuli contained two target stimuli and 10 nontarget stimuli, where a stimulus was a random intensification of a row or a column. Two target stimuli indicated the intensifications of the row and the column of the target character, respectively. The non-target stimuli were the intensifications of the rows and columns of the nontarget characters. The time between the onset of two adjacent stimuli, called stimulus onset asynchrony (SOA), was 250 ms, where the intensification time and the inter-stimulus interval (ISI) were both 125 ms.

B. DATASET 2: COVERT AND OVERT ERP-BASED BCI

In Dataset 2 [10], 10 healthy subjects (six males and four females, mean age $= 26.8 \pm 5.6$ years) took part in the experiment. The EEG signals were collected with BCI2000, digitized at 256Hz from 16 channels (Fz, FCz, Cz, CPz, Pz, Oz, F3, F4, C3, C4, CP3, CP4, P3, P4, PO7, and PO8) and bandpass filtered between 0.1 and 20 Hz. This study was performed on two speller paradigms: Farwell and Donchin's paradigm and the Geometric Speller (GeoSpell, Figure. 2).

The recordings using the two interfaces both included three sessions, with six trials in each session. Each trial contained eight rounds of repeated stimuli with 12 stimuli (two target stimuli and 10 nontarget stimuli) within every round of stimuli. The SOA and the ISI were 250 ms and 125 ms, respectively. For the stimulating patterns, the rows or columns were illuminated on Farwell and Donchin's interface as described in Dataset 1, whereas the GeoSpell interface displayed six characters per time interval until all 36 had appeared twice.

C. DATASET 3: BCI COMPETITION III-DATASET II

In Dataset 3 [11], the EEG signals recorded using Farwell and Donchin's interface, were bandpass filtered between 0.1 and 60 Hz and digitized at 240 Hz from 64 channels. Both EEG signals from two subjects (A and B) were divided into a training set (85 trials) and a testing set (100 trials). Every trial contained 15 rounds of repeated stimuli, and the intensification time and the ISI were 100 ms and 75 ms in each round.

D. DATA PREPROCESSING

The EEG signals of Dataset 1-3 were downsampled to 128, 128, and 120 Hz, respectively. Then, they were bandpass filtered between 0.1 and 20 Hz with the fifth-order Butterworth filter [36] to remove the short-term fluctuations and leave the longer-term trends [37]. At last, they were extracted from 0 to 0.5 s after each stimulus onset, as shown in Table 1.

IV. METHODS

This section describes the proposed STNN, where the temporal unit and spatial unit are connected concurrently, as shown in Figure 3. The details are as follows.

A. PARALLEL MECHANISM

The proposed model adopts a parallel mechanism to perform simultaneous analysis of EEG information in the time and space domains, which is expressed as:

$$\hat{y} = Sigmoid\left(t\left(X;\theta_t\right) + s\left(X;\theta_s\right)\right), \tag{1}$$

where X and $\hat{y}$ denote the input of EEG signals and the output of predicted results, the ideal output $\hat{y}$ is either 1 (target) or 0 (nontarget), $t\left(X;\theta_t\right)$ and $s\left(X;\theta_s\right)$ represent the functions of the temporal unit and the spatial unit, θ_t and θ_s are the network parameters, and *Sigmoid* is an S-shaped activation function.

B. SPATIAL UNIT

The spatial unit utilizes the global features of the EEG signals in the space domain for P300 detection. It is composed of Conv1Ds, linear layers, weight norms (WNs) [38],

TABLE 1. Data description and preprocessing procedure.

	Dataset 1	Dataset 2	Dataset 3
Original sampling rate (Hz)	256	256	240
Bandpass filter (Hz)	0.1-30	0.1-20	0.1-60
# of Subjects	8 (ALS)	10 (Healthy)	2 (Healthy)
# of Trials each subject	35	18	185
# of Repeated stimuli	10	8	15
# of EEG channels	8	16	64
Target vs Non-target	1 vs 5	1 vs 5	1 vs 5
Speller paradigm	F	F & G	F
Data preprocessing procedure			
Subsampling rate (Hz)	128	128	120
Butterworth filter (Hz)	0.1-20	0.1-20	0.1-20
Selected duration (s)	0.5	0.5	0.5
EEG format (C × T)	8 × 64	16 × 64	64 × 60

* F/G: Farwell and Donchin's paradigm/GeoSpell paradigm; C/T: The number of channels / time points.

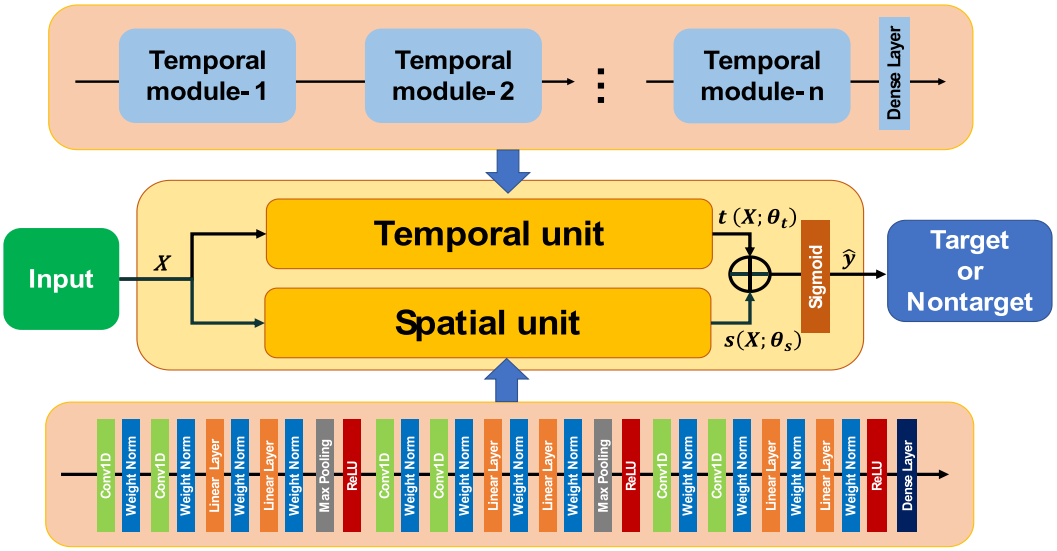


FIGURE 3. Overall architecture of the proposed spatial-temporal neural network (STNN).

max-pooling operations, rectified linear units (ReLUs), and a dense layer, as shown in Table 2. The hyperparameter tunning process is given in Figure. 11 and Figure. 12. This unit generalizes spatial features from the horizontal and vertical dimensions by the combination of multiple Conv1Ds and linear layers. To improve the model's robustness in different P300 detections, a multi-stage feature generalization and compression (Conv1D to Linear layer to Max pooling) is built in the unit.

1) Conv1Ds

Conv1Ds as single-dimensional filters, can generalize EEG channel features in the vertical dimension. By setting the kernel size to 1, the correlation between EEG electrodes at each time point is extracted.

2) LINEAR LAYERS AND MAX POOLING OPERATIONS

Linear layers can balance the extratced feature sizes in the horizontal and vertical dimensions, thus minimizing the information loss when compressing feature with max-pooling operations.

3) ReLUs AND WNs

ReLUs can improve the model's nonlinearity and avoid vanishing gradients, and WNs can accelerate network convergence.

4) DENSE LAYER

The dense layer is connected to the extracted features, producing the predicted results of the spatial unit.

C. TEMPORAL UNIT

Temporal unit detects P300 signals by learning the features of temporal changes in EEG signals. It comprises n temporal modules and a dense layer, as shown in Figure.3. The number of the temporal modules can be customized according to the input EEG signals.

1) TEMPORAL MODULE

As shown in Figure.4, each temporal module is assembled of a temporal analyzer and a global generalizer with a residual connection. The temporal analyzer performs sequence analysis, which is the core of learning the features of temporal

TABLE 2. Hyperparameters of the spatial unit.

Layer	# Params	Output
Input	/	C × T
Conv1D +WN	(C, 128, 1)	128 × T
Conv1D + WN	(128, 128, 1)	128 × T
Linear Layer + WN	(T, 128)	128 × 128
Linear Layer + WN	(128, 128)	128 × 128
Max Pooling + ReLU	(2, 2)	64 × 64
Conv1D +WN	(64, 64, 1)	64 × 64
Conv1D + WN	(64, 32, 1)	32 × 64
Linear Layer + WN	(64, 64)	32 × 64
Linear Layer + WN	(64, 32)	32 × 32
Max Pooling + ReLU	(2, 2)	16 × 16
Conv1D +WN	(16, 16, 1)	16 × 16
Conv1D + WN	(16, 4, 1)	4 × 16
Linear Layer + WN	(16, 16)	4 × 16
Linear Layer + WN + ReLU	(16, 4)	4 × 4
Dense Layer	(16, 1)	1 × 1
Output	/	1 × 1

* Conv1D: (Input channel, Output channel, Kernel size); * Linear layers: (Input channel, Output channel); * C/T: Number of channels/time points

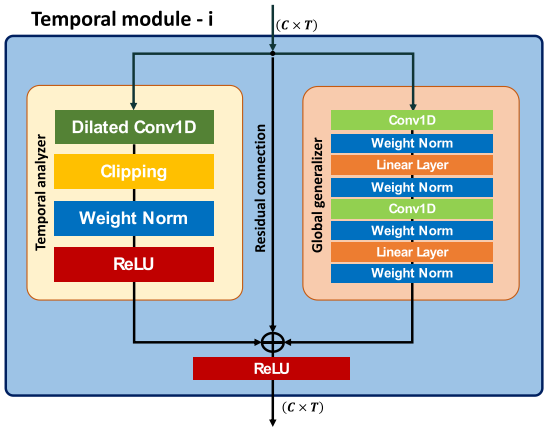


FIGURE 4. Internal structure of the temporal module.

changes. The global generalizer generalizes features from the raw EEG signals or the outputs from the former layer of the temporal module. It provides the global information for the next sequence analysis, which can be expressed as:

$$y^{(i)}_{:,T-1}, \ldots, y^{(i)}_{:,0} = f_i\left(x^{(i-1)}_{:,T-1}, \ldots, x^{(i-1)}_{:,0}\right), \tag{2}$$

where $\left[x^{(i-1)}_{:,T-1}, \ldots, x^{(i-1)}_{:,0}\right]$ and $\left[y^{(i)}_{:,T-1}, \ldots, y^{(i)}_{:,0}\right]$ are the inputs and outputs of each temporal module, and both are the same size of $C \times T$; C and T indicate the number of channels and time points, respectively; and f_i represents the i_{th} temporal module.

2) TEMPORAL ANALYZER

The temporal analyzer is composed of four components: a dilated Conv1D [39], a clipping operation, a weight norm, and a ReLU. Within the temporal analyzer of the i_{th} temporal module, the hyperparameters of the dilated Conv1D include input channel, output channel, kernel size, dilation, and zero-padding, where the input channel and output channel are the number of the electrodes of input EEG signals, and the kernel size, dilation, and zero-padding parameter are $k, 2^{i-1}$, and $(k-1)\times 2^{i-1}$, respectively. By them, the range of learning the temporal changes can be constantly extended. The function of clipping operation is to cut off $\left[y^{(i)}_{:,-1}, \ldots, y^{(i)}_{:,-(k-1)\times 2^{i-1}}\right]$ for structural consistency between the inputs and outputs. The functions of the ReLU and weight norm are similar to those in the spatial unit.

Figure.5 shows an example of two stacked temporal modules where the kernal size of the dilated Conv1D was set up to 2. The two temporal analyzers in Temporal module-1 and Temporal module-2 construct a sequential mapping from $\left[x^{(0)}_{:,3}, x^{(0)}_{:,2}, x^{(0)}_{:,1}, x^{(0)}_{:,0}\right]$ to $\left[y^{(1)}_{:,2}, y^{(1)}_{:,0}\right]$ to $\left[x^{(1)}_{:,2}, x^{(1)}_{:,0}\right]$ to $\left[y^{(2)}_{:,0}\right]$. The output $\left[y^{(2)}_{:,0}\right]$ represents the two-level temporal features of $\left[x^{(0)}_{:,3}, x^{(0)}_{:,2}, x^{(0)}_{:,1}, x^{(0)}_{:,0}\right]$.

As for a temporal unit stacked by n temporal modules, the output $\left[y^{(n)}_{:,0}\right]$ are the n-level mapping of the raw EEG signals $\left[x^{(0)}_{:,l-1}, \ldots, x^{(0)}_{:,0}\right]$, which yields the final result of the temporal unit by connecting with a dense layer; l indicates the length of receiving field, as calculated in (3).

$$l = k \times 2^{n-1}, \tag{3}$$

where k is the kernel size of the dilated Conv1D within the i_{th} temporal module.

For EEG epochs in different P300 detection tasks, we can customize the temporal detection range by adjusting n (the number of stacked temporal modules) and k (the kernel size of the dilated Conv1D in the dilated Conv1D).

3) TEMPORAL GENERALIZER

The structure of the temporal generalizer is similar to the spatial unit, consisting of Conv1Ds, linear layers, and WNs, as shown in Table 3. The hyperparameter tuning process is given in Figure. 13 and Figure. 14.

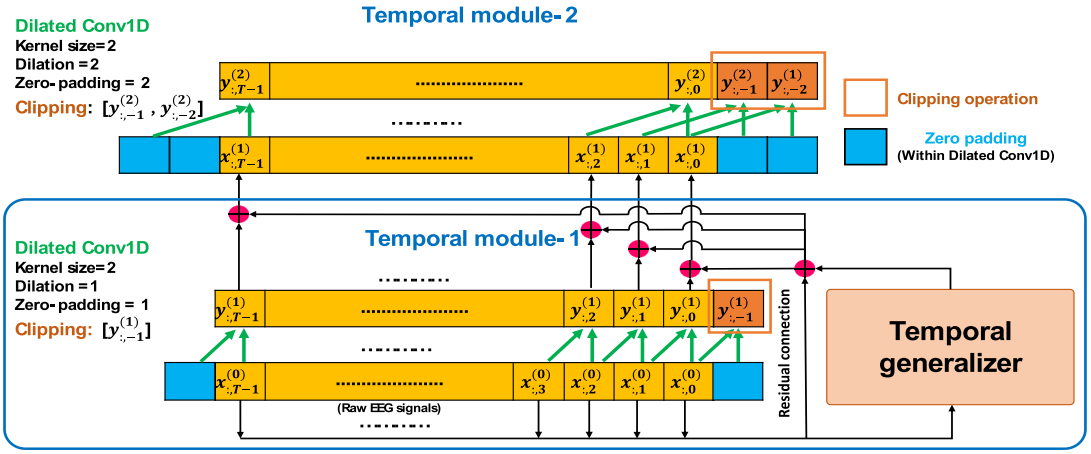


FIGURE 5. Mapping relations when stacking two temporal modules.

TABLE 3. Hyperparameters of the temporal generalizer.

Layer	# Params	Output
Input	/	C × T
Conv1D +WN	(C, 128, 1)	128 × T
Linear Layer + WN	(T, 128)	128 × 128
Conv1D +WN	(128, C, 1)	C × 128
Linear Layer + WN	(128, T)	C × T
Output	/	C × T

* Conv1D: (Input channel, Output channel, Kernel size)
* Linear layers: (Input channel × Output channel)
* C/T: Number of channels/time points

4) RESIDUAL CONNECTION

The residual connection simplifies the network learning process, especially when multiple temporal modules are stacked.

V. EXPERIMENTS

We performed three experiments to evaluate the proposed STNN: 1) P300 detection under multiple repeated stimuli with applications to ALS patients; 2) P300 detection on two speller paradigms to healthy subjects; 3) an ablation and combination study using BCI Competition III-dataset II.

All the experiments involved 30 iterations of network training within 5 mins, where the Adam optimizer [40] (learning rate = 0.001) was used to minimize the binary cross entropy (BCE) [41] between the outputs and the labels in Pytorch [42] environment. The evaluation metrics included accuracy, area under the receiver operating characteristic curve (AUC) [43], F1-score, and Kappa coefficient [44].

The reference approaches were 3D Input CNN [45], the winner in BCI Competition III [23], and EEGNet-*t*&*p* [27], where t and p were the number of temporal filters and pointwise filters, respectively.

A. EXPERIMENT 1

The first experiment explored our model performance under multiple rounds of repeated stimuli to ALS patients using Dataset 1 [9]. As described in Section III, Dataset 1 was composed of 8-channel EEG signals from eight ALS subjects, and there were 35 trials of each subject. Each trial was included of EEG signals under 10 rounds of repeated stimuli, there were 12 stimuli (two target and 10 nontarget stimuli) in each round. In a trial, by averaging the EEG epochs under the same stimuli from 1 to i rounds, there were 12 EEG epochs for training or testing the model performance under the i_{th} round of the repeated stimuli.

The proposed model was represented as STNN-*n*&*k*, where n was the number of temporal modules and k was the kernel size of the dilated Conv1D in each module. STNN-3&15, 4&7, 4&8, 5&3, and 5&4 were used in the experiment, and they produced the receptive fields of length 60, 56, 64, 48 and 64, covering the main portion of the EEG signals (The data length is 64 in Databset 1) in the temporal domain. The reference models were EEGNet-4&2, 8&2, 16&2, 4&4, 8&4, 16&4, and the most used in [33] were EEGNet-8&2 and EEGNet-4&2.

We implemented a within-subject P300 detection and a cross-subject P300 detection, respectively. In the within-subject task, we randomly selected 20 trials (240 EEG epochs) for model training from each subject and the remaining 15 trials (180 EEG epochs) for testing the model. The average results of eight subjects with 1-10 rounds of repeated stimuli are shown in Tables 4 and 5. From the performance comparison of these models, we can see that the average AUC and F1 scores of the proposed models are higher than its competitors under 1-10 rounds of repeated stimuli. All our models reach above 0.95 AUC scores using the EEG signals from the first five rounds of repeated stimuli, while EEGNet cannot reach it until at least the ninth round of stimuli. Moreover, the average F1 score of our models under the fifth round of repeated stimuli improved 25.3% than EEGNet in the same condition. This result is close to that of the reference models using 10 rounds of stimuli. It is demonstrated that the proposed models can attain the similar high detection accuracy using fewer repeated stimuli, thereby reducing the number of stimuli for ALS patients in applications.

TABLE 4. The AUC results of 1-10 rounds of repeated stimuli in the within-subject P300 detection.

Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
STNN-3&15	0.691	0.828	0.879	0.941	0.950	0.959	0.968	0.969	0.974	0.975
STNN-4&7	0.695	0.833	0.883	0.936	0.953	0.962	0.970	0.973	0.979	0.979
STNN-4&8	0.705	0.841	0.891	0.938	0.953	0.961	0.968	0.971	0.977	0.978
STNN-5&3	0.742	0.875	0.932	0.973	0.975	0.980	0.982	0.991	0.993	0.995
STNN-5&4	0.739	0.877	0.938	0.975	0.969	0.985	0.988	0.993	0.993	0.993
STNN-Average	**0.714**	**0.851**	**0.904**	**0.952**	**0.960**	**0.969**	**0.975**	**0.979**	**0.983**	**0.984**
EEGNet-4&2	0.677	0.788	0.836	0.877	0.905	0.918	0.909	0.946	0.963	0.967
EEGNet-8&2	0.709	0.813	0.874	0.940	0.934	0.944	0.951	0.977	0.970	0.970
EEGNet-16&2	0.709	0.810	0.877	0.935	0.935	0.941	0.949	0.979	0.975	0.975
EEGNet-4&4	0.677	0.788	0.841	0.881	0.917	0.921	0.915	0.950	0.965	0.970
EEGNet-8&4	0.679	0.793	0.845	0.891	0.921	0.925	0.936	0.955	0.968	0.969
EEGNet-16&4	0.705	0.811	0.867	0.940	0.938	0.944	0.953	0.976	0.971	0.973
EEGNet-Average	**0.692**	**0.801**	**0.857**	**0.911**	**0.925**	**0.932**	**0.935**	**0.963**	**0.968**	**0.971**

TABLE 5. The F1scores of 1-10 rounds of repeated stimuli in the within-subject P300 detection.

Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
STNN-3&15	0.253	0.531	0.651	0.709	0.774	0.815	0.832	0.875	0.888	0.885
STNN-4&7	0.286	0.549	0.678	0.731	0.793	0.833	0.850	0.883	0.894	0.886
STNN-4&8	0.295	0.551	0.677	0.723	0.795	0.833	0.851	0.886	0.897	0.899
STNN-5&3	0.455	0.571	0.728	0.787	0.823	0.845	0.863	0.902	0.918	0.921
STNN-5&4	0.440	0.563	0.731	0.793	0.813	0.843	0.841	0.895	0.925	0.923
STNN-Average	**0.346**	**0.553**	**0.693**	**0.749**	**0.800**	**0.834**	**0.847**	**0.888**	**0.904**	**0.903**
EEGNet-4&2	0.053	0.128	0.241	0.410	0.433	0.541	0.641	0.773	0.748	0.772
EEGNet-8&2	0.095	0.233	0.347	0.454	0.587	0.643	0.732	0.790	0.787	0.751
EEGNet-16&2	0.107	0.258	0.372	0.471	0.591	0.648	0.739	0.791	0.795	0.785
EEGNet-4&4	0.062	0.134	0.271	0.399	0.456	0.571	0.643	0.783	0.745	0.781
EEGNet-8&4	0.116	0.245	0.357	0.463	0.597	0.638	0.741	0.783	0.792	0.789
EEGNet-16&4	0.131	0.241	0.371	0.453	0.620	0.645	0.735	0.790	0.789	0.789
EEGNet-Average	**0.094**	**0.207**	**0.326**	**0.442**	**0.547**	**0.614**	**0.705**	**0.785**	**0.776**	**0.778**

TABLE 6. The AUC results of 1-10 rounds of repeated stimuli in the cross-subject P300 detection.

Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
STNN-3&15	0.734	0.788	0.815	0.885	0.888	0.910	0.913	0.913	0.917	0.918
STNN-4&7	0.779	0.813	0.843	0.889	0.908	0.915	0.921	0.920	0.923	0.925
STNN-4&8	0.778	0.805	0.839	0.893	0.907	0.921	0.923	0.925	0.924	0.925
STNN-5&3	0.792	0.815	0.845	0.901	0.911	0.919	0.925	0.938	0.935	0.937
STNN-5&4	0.789	0.809	0.851	0.899	0.909	0.918	0.923	0.935	0.936	0.936
STNN-Average	**0.774**	**0.806**	**0.836**	**0.893**	**0.904**	**0.916**	**0.921**	**0.926**	**0.927**	**0.928**
EEGNet-4&2	0.675	0.775	0.803	0.833	0.855	0.870	0.895	0.901	0.903	0.905
EEGNet-8&2	0.701	0.799	0.825	0.865	0.891	0.896	0.900	0.903	0.905	0.907
EEGNet-16&2	0.703	0.808	0.827	0.887	0.894	0.897	0.908	0.911	0.905	0.910
EEGNet-4&4	0.685	0.788	0.813	0.845	0.874	0.888	0.901	0.905	0.904	0.909
EEGNet-8&4	0.705	0.803	0.825	0.883	0.901	0.909	0.910	0.905	0.917	0.910
EEGNet-16&4	0.705	0.811	0.831	0.899	0.893	0.905	0.907	0.903	0.907	0.907
EEGNet-Average	**0.695**	**0.797**	**0.821**	**0.869**	**0.884**	**0.894**	**0.903**	**0.904**	**0.906**	**0.908**

In the cross-subject P300 detection, we utilized all trials of five random subjects for network training and the trials from the remaining three subjects to evaluate the network performance. The average AUC and F1-score results of five experiments following the above steps are listed in Tables 6 and 7, where we can see that our models obtain improvements of 2% in the average AUC score and 12.4% in the average F1 score under 10 rounds of stimuli. Notably, our models using EEG signals of six rounds of stimuli can reach the similar performance compared to the reference ones of 10 rounds of stimuli.

The average results of the kappa coefficient are given in Figure.6. According to [44], a study has substantial reliability when the kappa coefficient is greater than 0.6. STNN-average reached this standard under the 3rd and the 6th rounds of repeated stimuli in the with-subject detection and cross-subject detection, respectively. While EEGNet-average fulfilled the criterion under the 7th and the 10th rounds of stimuli, respectively. Overall, the proposed models achieved advantages both in the within-subject and cross-subject P300 detections, and they can reduce four rounds of repeated stimuli than the reference ones when gaining the similar results.

TABLE 7. The F1scores of 1-10 rounds of repeated stimuli in the cross-subject P300 detection.

Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
STNN-3&15	0.328	0.401	0.486	0.543	0.592	0.632	0.639	0.688	0.721	0.755
STNN-4&7	0.356	0.414	0.505	0.560	0.611	0.640	0.645	0.700	0.735	0.774
STNN-4&8	0.356	0.407	0.499	0.574	0.605	0.639	0.641	0.705	0.740	0.776
STNN-5&3	0.381	0.445	0.503	0.599	0.635	0.669	0.671	0.711	0.753	0.785
STNN-5&4	0.385	0.443	0.510	0.607	0.643	0.671	0.681	0.709	0.761	0.779
STNN-Average	**0.361**	**0.422**	**0.501**	**0.577**	**0.617**	**0.650**	**0.656**	**0.703**	**0.742**	**0.774**
EEGNet-4&2	0.157	0.287	0.352	0.411	0.555	0.578	0.600	0.628	0.639	0.638
EEGNet-8&2	0.187	0.305	0.405	0.421	0.576	0.590	0.615	0.631	0.645	0.671
EEGNet-16&2	0.195	0.311	0.370	0.450	0.563	0.584	0.607	0.630	0.639	0.658
EEGNet-4&4	0.143	0.277	0.348	0.399	0.541	0.573	0.599	0.611	0.642	0.645
EEGNet-8&4	0.185	0.315	0.401	0.433	0.588	0.601	0.606	0.605	0.635	0.648
EEGNet-16&4	0.199	0.322	0.399	0.441	0.571	0.591	0.617	0.622	0.639	0.641
EEGNet-Average	**0.178**	**0.303**	**0.379**	**0.426**	**0.566**	**0.586**	**0.607**	**0.621**	**0.640**	**0.650**

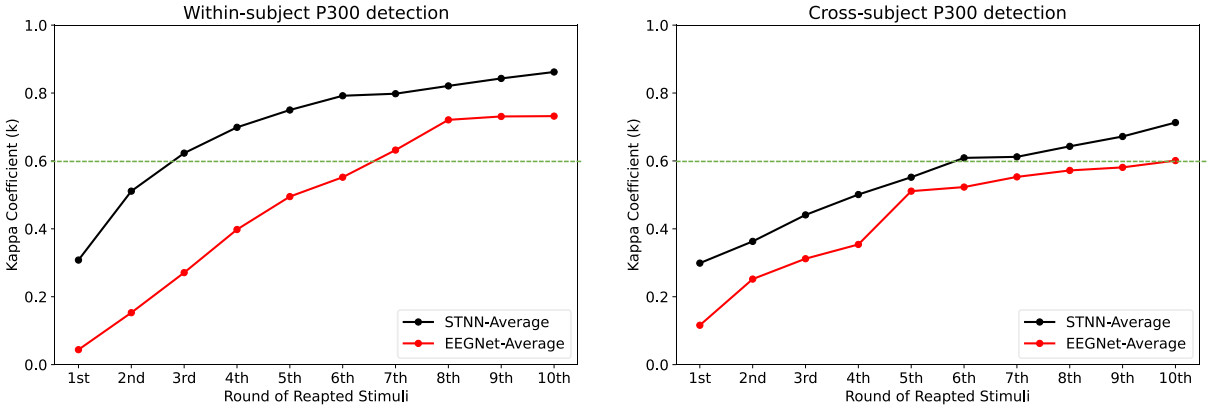


FIGURE 6. Kappa coefficient results in the within-subject P300 detection and cross-subject P300 detection.

B. EXPERIMENT 2

The second experiment studied our model performance on the two P300 speller paradigms (Farwell and Donchin's paradigm and the GeoSpell paradigm) to healthy subjects using Databset 2 [10]. Databset 2 were 16-channel EEG signals from 10 healthy subjects. There were three sessions (six trials in each session) in each subject's recordings. In each trial, the subject experienced eight rounds of repeated stimuli. Because the data from Databets 1 and 2 have the same length in the time domain, we still used STNN-3&15, 4&7, 4&8, 5&3, and 5&4 to implement the P300 detection tasks.

In the within-subject experiment, two sessions were randomly chosen as the training set from each healthy subject, and the remaining one was used for testing the models. Figure.7 and Figure.8 give the AUC and F1 scores and Kappa coefficient results of the proposed models and reference models using Farwell and Donchin's paradigm and the GeoSpell paradigm. We can see that the proposed STNN-3&15, 4&7, 4&8, 5&3, and 5&4 all achieved perfect detection (AUC, F1 scores and Kappa coefficient results were equal to 1) under the second rounds of stimuli on Farwell and Donchin's paradigm and under the fourth rounds of stimuli on the GeoSpell paradigm, while the reference ones need at least four and six rounds of repeated stimuli to reach this goal on Farwell and Donchin's and the GeoSpell paradigm, respectively. It shows that our models realize better performance using fewer stimuli to healthy subjects.

In the cross-subject experiment, we utilized all the trials from five random subjects to train network parameters. The trials from the remaining five subjects were used for the network testing. Figure.9 and Figure.10 show that our models always score higher than or equal to the reference ones under 1-8 rounds of repeated stimuli and reach perfect detection under the second round of stimuli over the two paradigms, while the reference ones fulfill this condition in the third or the fourth stimuli. Therefore, it can be seen that the proposed models can reduce repeated stimuli to healthy subjects and is robust to the two different P300 speller paradigms.

C. EXPERIMENT 3

To measure the contribution of individual components and component combinations on the model performance, the third experiment was an ablation and combination study on Dataset 3 (BCI Competition III-dataset II, 64 channels) [11], where the training and testing sets of two subjects (A and B) were described in Section II. In order to compare to other models, we implemented the P300 detection using the same evaluation metrics and rounds of stimuli as in the literatures [23], [45].

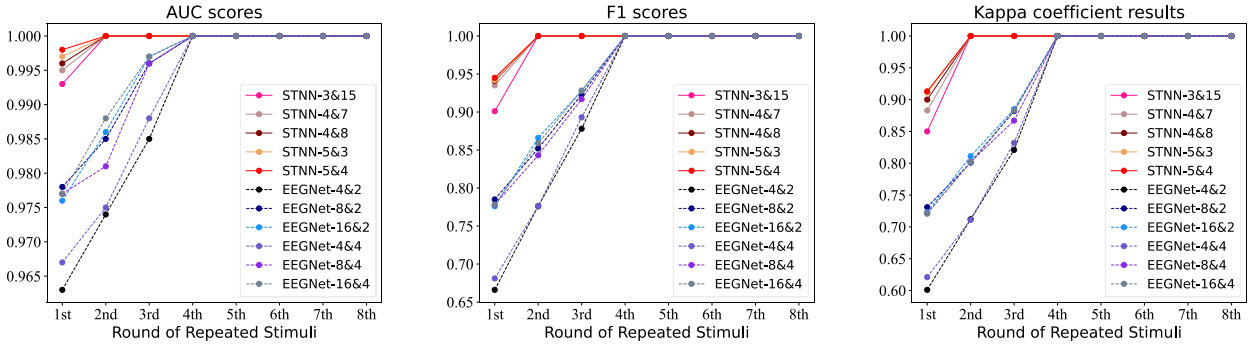


FIGURE 7. The AUC, F1 scores and kappa coefficient results using farwell and donchin's paradigm in the within-subject P300 detection.

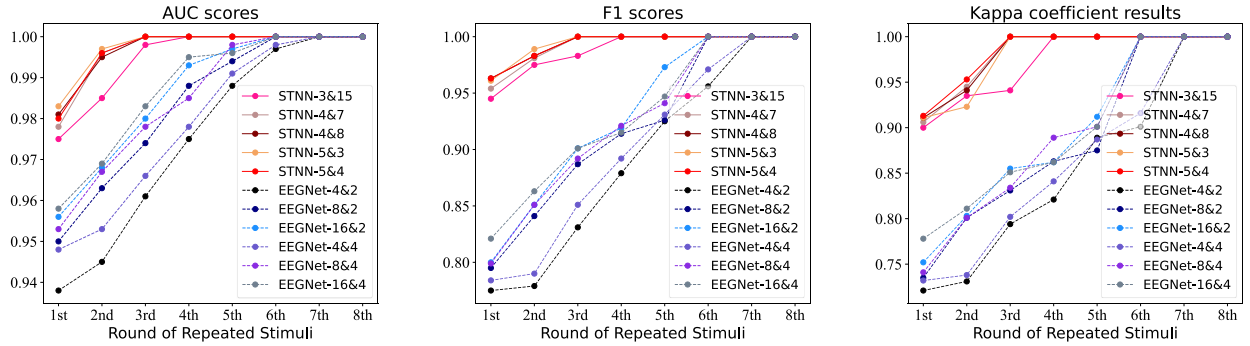


FIGURE 8. The AUC, F1 scores and kappa coefficient results using the GeoSpell paradigm in the within-subject P300 detection.

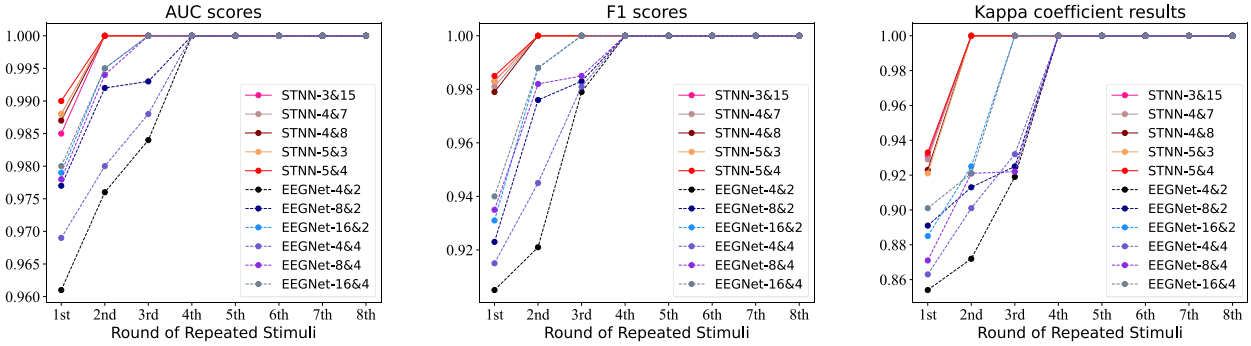


FIGURE 9. The AUC, F1 scores and kappa coefficient results using farwell and donchin's paradigm in the cross-subject P300 detection.

In the combination study, we tested the accuracy of multiple combinations of STNN, including STNN-1&60, 2&30, 3&15, 4&7, 4&8, 5&3, 5&4, and 6&2. In the ablation study, the above combinations were compared with their temporal units and STNN assembled with only the spatial unit. The results are shown in Table 8, where we can see that 1) the network performance is continuously improved by stacking one to four temporal modules in the temporal unit, while the performance does not continue to be enhanced when stacking more than four modules; 2) the parallel mechanism of the temporal unit and the spatial unit improves the accuracy by 1-3% over the temporal unit working alone; 3) the temporal unit is superior to the spatial unit in terms of the average accuracy; 4) to the best of our knowledge, the proposed STNN stacked with four or more temporal units outperforms the best state-of-the-art model in the literatures by at least 9% in accuracy.

VI. DISCUSSION

In this paper, we propose a novel DL model called STNN for P300 detection. The results prove that STNN performs better than other DL model and reduces the number of repeated stimuli in different P300 detections. Both healthy subjects and ALS patients can benefit from this research, even with limited data.

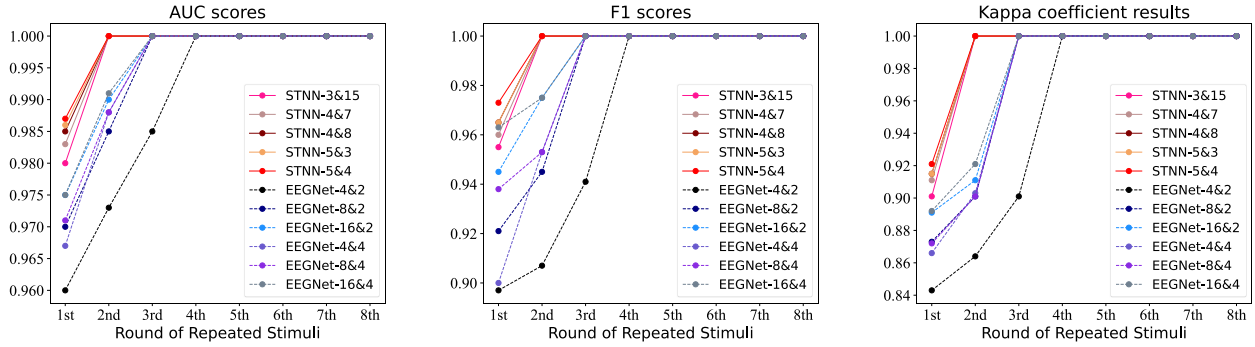


FIGURE 10. The AUC and F1 scores and kappa coefficient results using the GeoSpell paradigm in the cross-subject P300 detection.

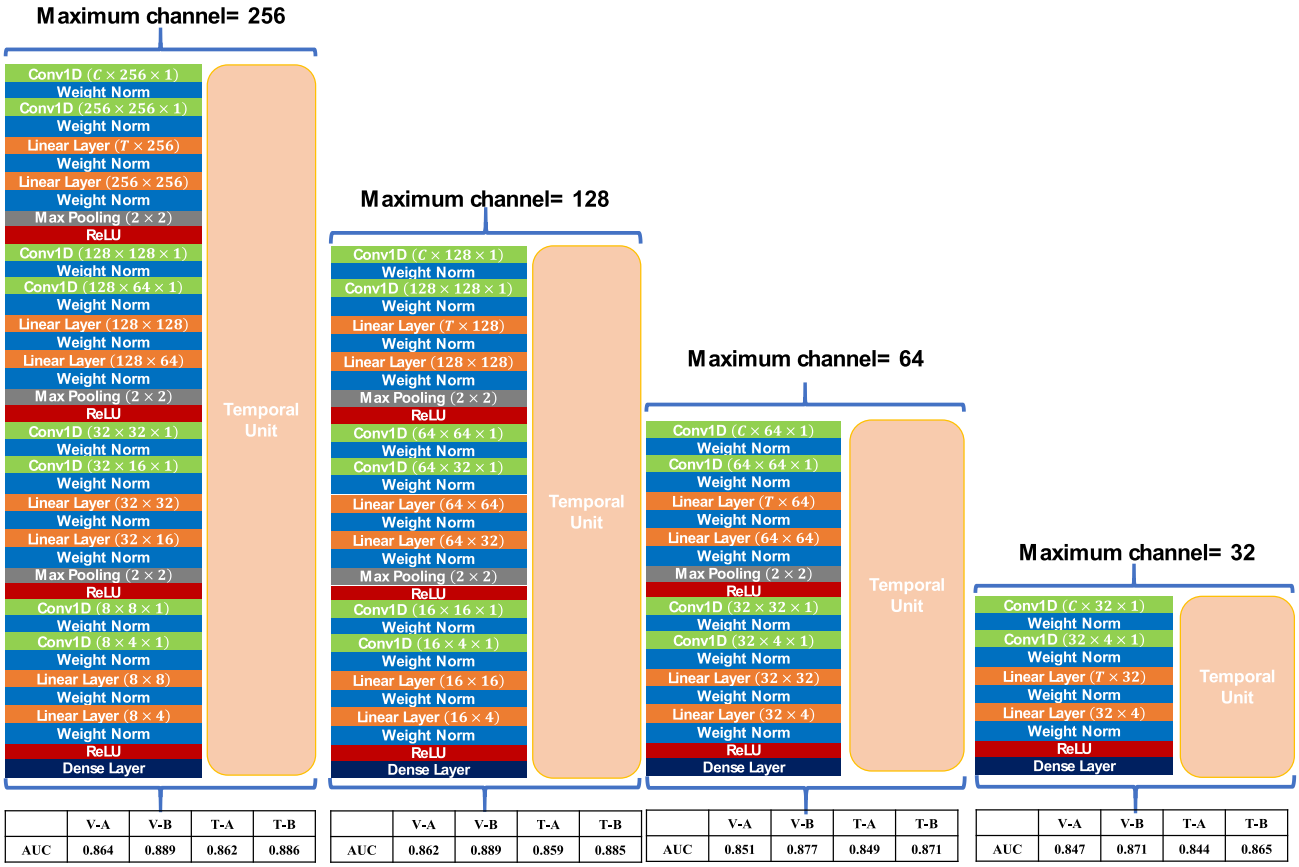
TABLE 8. Ablation and combination study: accuracy under 5 rounds of repeated stimuli on database 3.

Method	Subject-A	Subject-B	Average
STNN-1&60 ($l = 60$)	82.5%	89.1%	85.8%
STNN-1&60 - Only with the temporal unit	81.0%	86.2%	83.6%
STNN-2&30 ($l = 60$)	83.5%	88.5%	86.0%
STNN-2&30 - Only with the temporal unit	82.2%	87.4%	84.8%
STNN-3&15 ($l = 60$)	88.3%	89.5%	88.9%
STNN-3&15 - Only with the temporal unit	85.8%	89.2%	87.5%
STNN-4&7 ($l = 56$)	88.0%	90.0%	89.0%
STNN-4&7 - Only with the temporal unit	87.6%	87.7%	87.7%
STNN-4&8 ($l = 64$)	87.5%	90.5%	89.0%
STNN-4&8 - Only with the temporal unit	86.9%	87.5%	87.2%
STNN-5&3 ($l = 48$)	87.7%	90.5%	89.1%
STNN-5&3 - Only with the temporal unit	85.4%	88.8%	87.1%
STNN-5&4 ($l = 64$)	88.3%	89.7%	89.0%
STNN-5&4 - Only with the temporal unit	84.7%	87.9%	86.3%
STNN-6&2 ($l = 64$)	88.4%	89.8%	89.2%
STNN-6&2 - Only with the temporal unit	86.1%	85.9%	86.0%
STNN - Only with the spatial unit	82.5%	83.5%	83.0%
Competitors			
3D Input CNN (Best result in literatures)	74%	86%	80%
Winner in BCI Competition III	60%	87%	73.5%

*l represents the receiving field size.

The main reasons are as follows: 1) the temporal unit, as a flexible DL-based network dedicated to time-domain modeling, can capture the temporal dependencies from brain potential changes by constructing an end-to-end multi-level sequential mapping, so it is more sensitive than the previously mentioned approaches when detecting P300 signals; 2) the spatial unit can constantly generalize and compress P300 features in the space domain, which hedges complex noise interference to a certain extent; 3) a joint decision-making mechanism is built into the network by connecting the temporal unit and the spatial unit concurrently, which can utilize the above advantages of the two units, thus achieving both better performance and stronger robustness, as shown in Experiments 1 and 2.

Furthermore, it should be emphasized that stacking multiple temporal modules within the temporal unit is critical for sequential modeling, as shown in Experiment 3. The network accuracy is constantly improved when one to four temporal modules are stacked, which demonstrates a more complicated multi-level sequence model is more suitable for characterizing temporal changes in human brain regions. Nevertheless, the over-stacking of temporal modules cannnot endlessly improve its performance but rather increases the model complexity due to the larger number of training parameters, as we can see that the accuracy scores of STNN-4&7, 4&8, 5&3, 5&4, and 6&2 are almost equivalent in Table 5. Even so, our results still significantly outperform the best methods in the literature, to the best of our knowledge, in BCI Competition III. This is possible because, driven by the great success of 2D or 3D CNNs in image processing and video analysis, some current state-of-the-art DL frameworks are commonly obsessed with high-dimensional feature extraction from EEG data. However, the P300 signals present significant 1D features (the deflections in the time domain) rather than high-dimensional ones. 2D or 3D frameworks are not skilled at decoding features from EEG signals recorded with a small number of channels, because the EEG data inherently lack the spatial resolution [46]. In contrast, our network focuses more



	V-A	V-B	T-A	T-B
AUC	0.864	0.889	0.862	0.886

	V-A	V-B	T-A	T-B
AUC	0.862	0.889	0.859	0.885

	V-A	V-B	T-A	T-B
AUC	0.851	0.877	0.849	0.871

	V-A	V-B	T-A	T-B
AUC	0.847	0.871	0.844	0.865

FIGURE 11. Hyperparameter tuning of the spatial unit.

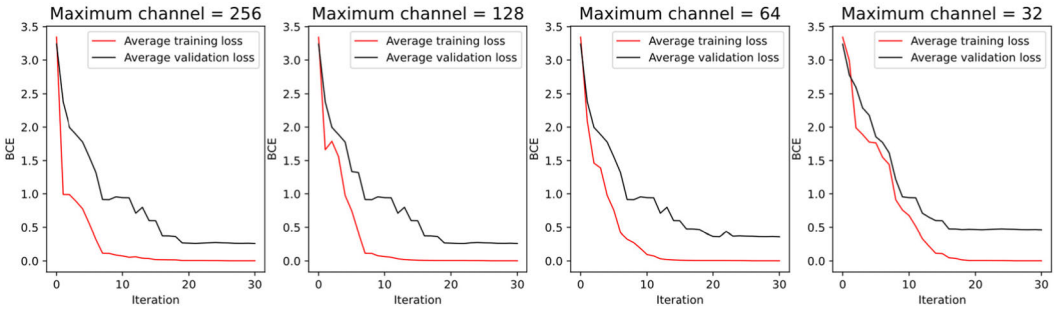


FIGURE 12. Average training loss and validation loss of the spatial unit (batch size = 32).

on the temporal activities within the P300 signals and EEG channel generalization in the space domain, which thus can capture more hidden information from EEG signals at a low SNR.

In the future, the proposed STNN is predicted to reach a high information transfer rate (ITR) when implementing online P300 detection because the information transferred per unit of time is likely to increase because of the decrease in the rounds of repeated stimuli. Moreover, we consider that this network has potential for applications in EEG-BCI systems and some other areas of signal processing, such as Electrocardiogram (ECG) classification [47], seeing that it is designed with a flexible structure and can be fast training and testing with limitied data, as shown in B.1-4 (Appendix-B).

VII. CONCLUSION

Spatial-temporal neural network (STNN), a DL-based P300 detection network, is proposed in this paper. The network is a parallel architecture consisting of a temporal unit and a spatial unit. It can perform EEG channel generalization and analyze the brain's potential changes simultaneously.

The results using three public databases reveal that our network performs better with fewer rounds of stimuli than

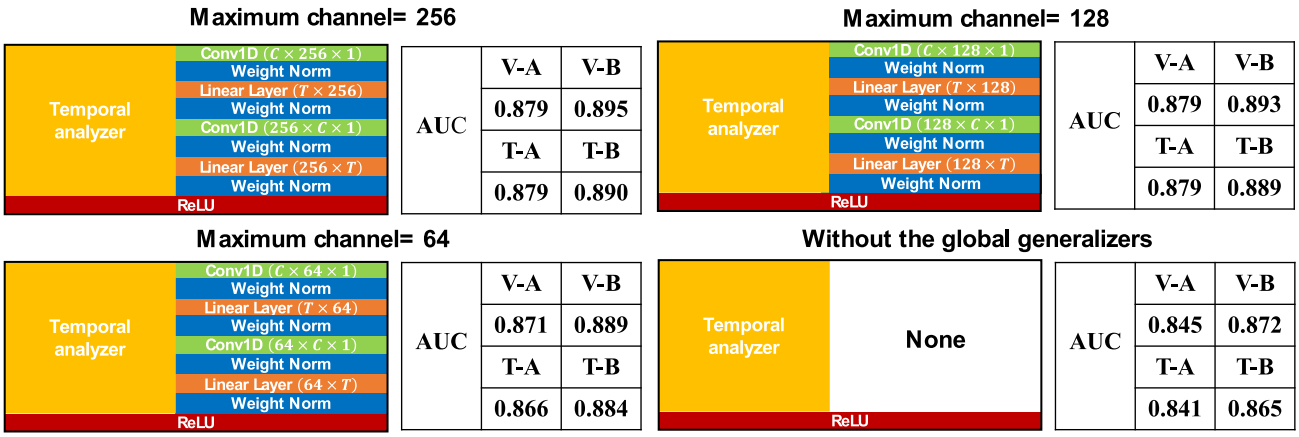


FIGURE 13. Hyperparameter tuning of the global generalizers in the temporal module.

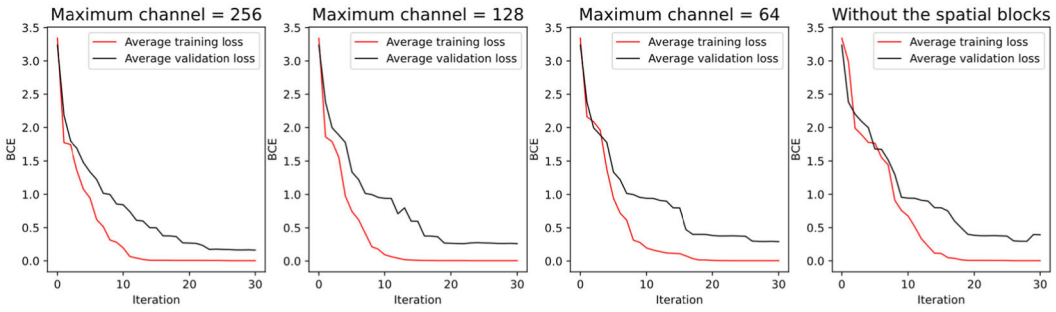


FIGURE 14. Average training loss and validation loss of the global generalizers in the temporal module (batch size = 32).

TABLE 9. Average running time in the first experiment.

Model (# of parameters)	Training time(s)	Testing time(s)
Within-subject P300 detection, Batch size = 32		
STNN-3&15(48827)	16.15	0.026
STNN-4&7(56299)	21.05	0.027
STNN-4&8(56811)	21.27	0.030
STNN-5&3(64283)	23.69	0.038
STNN-5&4(64923)	23.93	0.037
Cross-subject P300 detection, Batch size = 32		
STNN-3&15(48827)	95.5	0.037
STNN-4&7(56299)	112.5	0.039
STNN-4&8(56811)	111.6	0.040
STNN-5&3(64283)	126.1	0.042
STNN-5&4(64923)	125.9	0.042

other competitors. It is robust with limited data and is suitable for decoding EEG data recorded with various electrodes. In the future, we expect the proposed network to play a critical role in online P300 detection and other areas of EEG-BCI systems.

APPENDIX A
HYPERPARAMETER TUNING

Figure 11. lists the hyperparameter tuning process of the spatial unit, where the average AUC scores of STNN-3&15, 4&7, 4&8, 5&3, and 5&4 are given. We utilized Dataset 3 for training, validating, and testing models, where we performed 5-fold cross-validation on the training dataset (85 trials), and the testing dataset (100 trials) was given in Section II. V-A, V-B, T-A, and T-B are short for the validation result of subject-A, validation result of subject-B, testing result of subject-A, and testing result of subject-B. Figure 12. gives the average training loss and validation loss of subjects A and B.

We can see that the model performance can be improved by increasing the number of hyperparameters in the spatial unit, while the excessive increase in the hyperparameters does not

TABLE 10. Average running time in the second experiment (F).

Model (# of parameters)	Training time(s)	Testing time(s)
Within-subject P300 detection, Batch size = 32		
STNN-3&15(70355)	9.07	0.023
STNN-4&7(72371)	10.53	0.029
STNN-4&8(74419)	10.53	0.029
STNN-5&3(76435)	11.57	0.032
STNN-5&4(78994)	11.58	0.032
Cross-subject P300 detection, Batch size = 32		
STNN-3&15(70355)	51.05	0.036
STNN-4&7(72371)	59.30	0.041
STNN-4&8(74419)	59.78	0.042
STNN-5&3(76435)	66.55	0.046
STNN-5&4(78994)	66.74	0.047

TABLE 11. Average running time in the second experiment (G).

Model (# of parameters)	Training time(s)	Testing time(s)
Within-subject P300 detection, Batch size = 32		
STNN-3&15(70355)	9.22	0.025
STNN-4&7(72371)	10.31	0.027
STNN-4&8(74419)	10.52	0.028
STNN-5&3(76435)	11.44	0.030
STNN-5&4(78994)	11.46	0.030
Cross-subject P300 detection, Batch size = 32		
STNN-3&15(70355)	51.73	0.036
STNN-4&7(72371)	60.50	0.041
STNN-4&8(74419)	60.53	0.043
STNN-5&3(76435)	67.90	0.048
STNN-5&4(78994)	67.91	0.048

TABLE 12. Average running time in the third experiment.

Model (# of parameters)	Training time(s)	Testing time(s)
STNN-1&60(130019)	29.43	0.029
STNN-1&60-Only with the temporal unit (108737)	27.53	0.027
STNN-2&30(147171)	35.70	0.032
STNN-2&30-Only with the temporal unit (125889)	33.79	0.031
STNN-3&15(201443)	40.62	0.035
STNN-3&15-Only with the temporal unit (180161)	39.40	0.034
STNN-4&7(259331)	45.07	0.036
STNN-4&7-Only with the temporal unit (238049)	44.15	0.035
STNN-4&8(262099)	45.36	0.036
STNN-4&8-Only with the temporal unit (240817)	43.70	0.036
STNN-5&3(272987)	50.40	0.037
STNN-5&3-Only with the temporal unit (251705)	48.90	0.036
STNN-5&4(281947)	50.50	0.037
STNN-5&4-Only with the temporal unit (260665)	48.95	0.040
STNN-6&2(322563)	54.60	0.041
STNN-6&2-Only with the temporal unit (301281)	53.91	0.041
STNN - Only with the spatial unit (21282)	41.05	0.028

significantly improve its performance. Therefore, the spatial unit with maximum channel = 128 was adopted in our P300 detection study.

Figure 13. gives the hyperparameter tuning process of the global generalizers in the temporal module using Dataset 3. Figure 14. shows the average training loss and validation loss of subjects A and B. We separately assembled the global generalizers with different maximum channels into STNN-3&15, 4&7, 4&8, 5&3, and 5&4. According to the average 5-fold cross-validation and testing AUC scores of these five models, we can see that the model performance can be improved using the global generalizers in the temporal modules. However, a huge amount of training parameters led to computational redundancy but did not obviously improve

the model performance. Therefore, the output channel of the temporal feature generalizer was set up to 128 in our P300 detection study.

APPENDIX B
RUNNING TIME

Tables. 9–12 present the average running time of the three experiments. All the experiments were implemented using a Linux PC with two GeForce GTX 1080 GPUs.

Table 9 shows the processing time of the within-subject P300 detection and cross-subject P300 detection in the first experiment, where we list the average training and testing times of 1-10 rounds of stimuli.

In the second experiment, we calculated the average training and testing times of 1-8 rounds of stimuli in the within-subject P300 detection and cross-subject P300 detection. The results in Farwell and Donchin's paradigm and the GeoSpell paradigm are given in Table 10 and Table 11, respectively. Table 12 gives the average training and testing times of two subjects using our model components and combinations in the third experiment.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude for the insightful suggestions given by Dr. S. Yoshida, Dr. M. Takeda, and Dr. K. Matsuzaki. They are also indebted to the experts who reviewed this article and generously lent them invaluable help.

REFERENCES

[1] Y. Liu, Y. Liu, J. Tang, E. Yin, D. Hu, and Z. Zhou, "A self-paced BCI prototype system based on the incorporation of an intelligent environment-understanding approach for rehabilitation hospital environmental control," *Comput. Biol. Med.*, vol. 118, Mar. 2020, Art. no. 103618.

[2] M. Balconi and G. Fronda, "The use of hyperscanning to investigate the role of social, affective, and informative gestures in non-verbal Communication. Electrophysiological (EEG) and inter-brain connectivity evidence," *Brain Sci.*, vol. 10, no. 1, p. 29, Jan. 2020.

[3] A. Mishra, S. Sharma, S. Kumar, P. Ranjan, and A. Ujlayan, "Effect of hand grip actions on object recognition process: A machine learning-based approach for improved motor rehabilitation," *Neural Comput. Appl.*, vol. 33, no. 7, pp. 2339–2350, Apr. 2021.

[4] X. Wang, G. Gong, N. Li, and Y. Ma, "A survey of the BCI and its application prospect," in *Proc. Asian Simulation Conf.*, Singapore: Springer, 2016, pp. 102–111.

[5] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalograph. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, Dec. 1988.

[6] C. Başar-Eroglu, T. Demiralp, M. Schürmann, and E. Başar, "Topological distribution of oddball 'P300' responses," *Int. J. Psychophysiol.*, vol. 39, nos. 2–3, pp. 213–220, 2001.

[7] M. Abidi, G. Marco, A. Couillandre, M. Feron, E. Mseddi, N. Termoz, G. Querin, P. Pradat, and P. Bede, "Adaptive functional reorganization in amyotrophic lateral sclerosis: Coexisting degenerative and compensatory changes," *Eur. J. Neurol.*, vol. 27, no. 1, pp. 121–128, Jan. 2020.

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

[9] A. Riccio, L. Simione, F. Schettini, A. Pizzimenti, M. Inghilleri, M. O. Belardinelli, D. Mattia, and F. Cincotti, "Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis," *Frontiers Hum. Neurosci.*, vol. 7, p. 732, Nov. 2013.

[10] F. Aloise, P. Aricò, F. Schettini, A. Riccio, S. Salinari, D. Mattia, F. Babiloni, and F. Cincotti, "A covert attention P300-based brain–computer interface: Geospell," *Ergonomics*, vol. 55, no. 5, pp. 538–551, May 2012.

[11] B. Blankertz, K. Muller, D. Krusienski, G. Schalk, J. Wolpaw, A. Schlogl, G. Pfurtscheller, J. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.

[12] J. Cheng, L. Li, C. Li, Y. Liu, A. Liu, R. Qian, and X. Chen, "Remove diverse artifacts simultaneously from a single-channel EEG based on SSA and ICA: A semi-simulated study," *IEEE Access*, vol. 7, pp. 60276–60289, 2019.

[13] T. Wang, P. Liu, X. An, Y. Ke, J. Xu, M. Xu, L. Kong, W. Liu, and D. Ming, "Modeling strategies and spatial filters for improving the performance of P300-speller within and across individuals," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl.*, Jun. 2019, pp. 1–5.

[14] M. J. Monesi and S. Hajipour Sardouie, "Extended common spatial and temporal pattern (ECSTP): A semi-blind approach to extract features in ERP detection," *Pattern Recognit.*, vol. 95, pp. 128–135, Nov. 2019.

[15] P. Schembri, R. Anthony, and M. Pelc, "The feasibility and effectiveness of P300 responses using low fidelity equipment in three distinctive environments," in *Proc. 5th Int. Conf. Physiolog. Comput. Syst.*, 2018, pp. 77–86.

[16] J. Jin, S. Li, I. Daly, Y. Miao, C. Liu, X. Wang, and A. Cichocki, "The study of generic model set for reducing calibration time in P300-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 3–12, Jan. 2020.

[17] S. Kundu and S. Ari, "P300 based character recognition using convolutional neural network and support vector machine," *Biomed. Signal Process. Control*, vol. 55, Jan. 2020, Art. no. 101645.

[18] D. Krzemiński, S. Michelmann, M. Treder, and L. Santamaria, "Classification of P300 component using a Riemannian ensemble approach," in *Proc. 15th Medit. Conf. Med. Biol. Eng. Comput.*, Sep. 2019, pp. 1885–1889.

[19] F. Li, Y. Xia, F. Wang, D. Zhang, X. Li, and F. He, "Transfer learning algorithm of P300-EEG signal based on XDAWN spatial filter and Riemannian geometry classifier," *Appl. Sci.*, vol. 10, no. 5, p. 1804, Mar. 2020.

[20] M. Simões, D. Borra, E. Santamaría-Vázquez, M. Bittencourt-Villalpando, D. Krzemiński, A. Miladinović, T. Schmid, H. Zhao, C. Amaral, B. Direito, J. Henriques, P. Carvalho, and M. Castelo-Branco, "BCIAUT-P300: A multi-session and multi-subject benchmark dataset on autism for P300-based brain-computer-interfaces," *Frontiers Neurosci.*, vol. 14, p. 978, Sep. 2020.

[21] X. Zhang, L. Yao, X. Wang, J. Monaghan, D. McAlpine, and Y. Zhang, "A survey on deep learning-based non-invasive brain signals: Recent advances and new frontiers," *J. Neural Eng.*, vol. 18, no. 3, Jun. 2021, Art. no. 031002.

[22] Z. Cao, "A review of artificial intelligence for EEG-based brain–computer interfaces and applications," *Brain Sci. Adv.*, vol. 6, no. 3, pp. 162–170, Sep. 2020.

[23] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011.

[24] S. Ghiselli, F. Gheller, P. Trevisi, E. Favaro, A. Martini, and M. Ermani, "Restoration of auditory network after cochlear implant in prelingual deafness: A P300 study using LORETA," *Acta Otorhinolaryngolog. Italica*, vol. 40, no. 1, pp. 64–71, Feb. 2020.

[25] M. Awais, X. Long, B. Yin, S. F. Abbasi, S. Akbarzadeh, C. Lu, X. Wang, L. Wang, J. Zhang, J. Dudink, and W. Chen, "A hybrid DCNN-SVM model for classifying neonatal sleep and wake states based on facial expressions in video," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 5, pp. 1441–1449, May 2021.

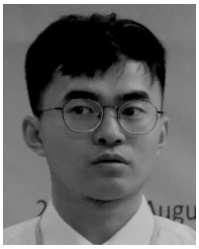
[26] S. F. Abbasi, J. Ahmad, A. Tahir, M. Awais, C. Chen, M. Irfan, H. A. Siddiqa, A. B. Waqas, X. Long, B. Yin, S. Akbarzadeh, C. Lu, L. Wang, and W. Chen, "EEG-based neonatal sleep-wake classification using multilayer perceptron neural network," *IEEE Access*, vol. 8, pp. 183025–183034, 2020.

[27] M. Liu, W. Wu, Z. Gu, Z. Yu, F. F. Qi, and Y. Li, "Deep learning based on batch normalization for P300 signal detection," *Neurocomputing*, vol. 275, pp. 288–297, Jan. 2018.

[28] D. Borra, S. Fantozzi, and E. Magosso, "Convolutional neural network for a P300 brain-computer interface to improve social attention in autistic spectrum disorder," in *Proc. 15th Medit. Conf. Med. Biol. Eng. Comput.*, vol. 76, Coimbra, Portugal, 2019, pp. 206–212.

[29] Z. Lu, Q. Li, N. Gao, T. Wang, J. Yang, and O. Bai, "A convolutional neural network based on batch normalization and residual block for P300 signal detection of P300-speller system," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Aug. 2019, pp. 2303–2308.

[30] O. Tal and D. Friedman, "Recurrent neural networks for P300-based BCI," 2019, *arXiv:1901.10798*.
[31] A. Ditthapron, N. Banluesombatkul, S. Ketrat, E. Chuangsuwanich, and T. Wilaiprasitporn, "Universal joint feature extraction for P300 EEG classification using multi-task autoencoder," *IEEE Access*, vol. 7, pp. 68415–68428, 2019.
[32] R. Maddula, J. Stivers, M. Mousavi, S. Ravindranand, and V. de Sa, "Deep recurrent convolutional neural networks for classifying P300 BCI signals," in *Proc. 7th Graz Brain-Comput. Interface, Conf.*, Sep. 2017, pp. 18–22.
[33] 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.
[34] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
[35] G. E. Chatrian, E. Lettich, and P. L. Nelson, "Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities," *Amer. J. EEG Technol.*, vol. 25, no. 2, pp. 83–92, Jun. 1985.
[36] I. W. Selesnick and C. S. Burrus, "Generalized digital Butterworth filter design," *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1688–1694, Jun. 1998.
[37] S. F. Abbasi, M. Awais, X. Zhao, and W. Chen, "Automatic denoising and artifact removal from neonatal EEG," in *Proc. 3rd Int. Conf. Biol. Inf. Biomed. Eng.*, Jun. 2019, pp. 1–5.
[38] T. Salimans and D. P. Kingma, "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," 2016, *arXiv:1602.07868*.
[39] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.
[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[41] A. Creswell, K. Arulkumaran, and A. A. Bharath, "On denoising autoencoders trained to minimise binary cross-entropy," 2017, *arXiv:1708.08487*.
[42] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Pytorch: An imperative style high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8026–8037.
[43] J. M. Lobo, A. Jiménez-Valverde, and R. Real, "AUC: A misleading measure of the performance of predictive distribution models," *Global Ecol. Biogeogr.*, vol. 17, no. 2, pp. 145–151, 2008.
[44] M. L. McHugh, "Interrater reliability: The Kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012.
[45] Z. Oralhan, "3D input convolutional neural networks for P300 signal detection," *IEEE Access*, vol. 8, pp. 19521–19529, 2020.
[46] F. Miwakeichi, E. Martínez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space–time–frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, Jul. 2004.
[47] J. Zhang, A. Liu, M. Gao, X. Chen, X. Zhang, and X. Chen, "ECG-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network," *Artif. Intell. Med.*, vol. 106, Jun. 2020, Art. no. 101856.



ZHEN ZHANG was born in China. He received the B.E. degree from the Department of Automation, Harbin University of Science and Technology, in 2016, and the M.E. degree from the Department of Physics and Electronic Engineering, Harbin Normal University, in 2020. He is currently pursuing the Ph.D. degree with the School of Information, Kochi University of Technology, Kochi, Japan. His current research interests include deep learning and biological signal processing.

XIAOYAN YU (Member, IEEE) received the B.E. and M.E. degrees from the Department of Physics, Harbin Normal University, Harbin, China, and the Ph.D. degree from the School of Information, Kochi University of Technology, Kochi, Japan. She is currently a Professor with Harbin Normal University. Her current research interests include image processing and computer vision.

XIANWEI RONG received the B.E. degree from the Department of Physics, Harbin Normal University, Harbin, China, 1996, and the M.E. degree from the School of Information and Communication, Harbin Engineering University, China, in 2010. He is currently a Professor with Harbin Normal University. His current research interests include image processing and embedded systems.

MAKOTO IWATA (Member, IEEE) received the B.E. and M.E. degrees in electronic engineering and the Ph.D. degree in information systems engineering from Osaka University, in 1986, 1988, and 1997, respectively. He joined the Department of Information Systems Engineering, Graduate School of Engineering, Osaka University, in 1991, as an Assistant Professor. After that, he joined the Department of Information Systems Engineering, Kochi University of Technology, Kochi (KUT), Japan, in 1997, as an Associate Professor and became a Professor, in 2002. He also worked as the Director of the Research Institute, KUT, where he is currently a Professor with the School of Information. He was a Visiting Associate Professor with the Research Center for 21st-Century Information Technology (IT-21 Center), Research Institute of Electrical Communication, Tohoku University, Sendai, Japan, from 2002 to 2005, and then he worked as a Visiting Professor with the IT-21 Center, from 2006 to 2009. In 2008, he spent with the Department of Electrical Engineering and Computer Science, University of California at Irvine, Irvine, USA. His current research interests include low-power and dependable data-driven architecture and its self-timed circuit implementation. He is also interested in brain computing and its VLSI implementation.