IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 16, 2022, accepted April 28, 2022, date of publication May 9, 2022, date of current version May 12, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3173261

A Novel Multimodule Neural Network for EEG Denoising

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 In this paper, a novel multi-module neural network (MMNN) is proposed to remove ocular artifacts (OAs) and myogenic artifacts (MAs) from noisy single-channel electroencephalogram (EEG) signals. This network is a based on deep learning (DL) architecture consisting of multiple denoising modules connected in parallel. Each denoising module is built using one-dimensional convolutions (Conv1Ds) and fully connected (FC) layers, and it estimates not only clean EEG signals but also artifacts. The proposed MMNN has two main advantages. Frist, the multiple denoising modules can purify noisy input EEG signals by continuously removing artifacts in the forward propagation. Second, the parallel architecture allows the parameters of each denoising module to be updated concurrently in the backpropagation, thereby improving the learning capacity of neural networks. We tested the network denoising performance using a recent public database, namely, EEGdenoiseNet. The results revealed that the proposed network reduced the temporal relative root mean square error (T-RRMSE) and spectral relative root mean square error (S-RRMSE) by at least 6% and enhanced the correlation coefficient (CC) by at least 3% over the state-of-the-art approaches. These significant performance improvements were confirmed by observing the deviation distribution between the denoised and clean signals. Furthermore, the proposed network achieved a similar performance efficiency with only 60% of the training data compared to the existing DL models. Our model can be found at: https://github.com/Zhangzhenkut/Multi-Module-Neural-Network-for-EEG-Denoising

INDEX TERMS EEG denoising, multi-module neural network (MMNN), deep learning (DL).

I. INTRODUCTION

Electroencephalography (EEG) is a safe, reliable, and relatively non-invasive measurement tool to study human brain activity. EEG signals are used in various fields, such as the diagnosis and treatment of diseases [1]–[3], investigations of brain's neurobiological mechanisms [4]–[6], and brain-computer interface (BCI) systems [7], [8]. Related studies strongly rely on whether EEG data accurately represent brain activity. However, noise and artifacts are always contained in EEG signals, and they are entangled with brain activity.

Eye movements [9] and facial muscle activity [10] are two common causes of noise and artifacts in EEG epochs. Eye movements distort the electric field around the eyes and over the scalp, thus causing ocular artifacts (OAs) [11]. Facial muscle activity responds to pressure changes in the upper airway, generating electrical amplitude signals, called myogenic artifacts (MAs) [12]. Recently, many approaches, such as regression [13]–[15], adaptive filtering [16], [17], blind source separation (BSS) [18]–[21], and empirical mode decomposition (EMD) [22] have been proposed to remove OAs and MAs from EEG epochs. A high-performance denoising approach should be able to accurately remove artifacts in real-time without distorting the signal of interest and be sufficiently robust to reconstruct EEG data in various formats, especially the signals recorded using only a few electrodes. However, these methods have not fully competent in fulfilling these criteria. The regression and adaptive filtering techniques are not efficient for real-time applications, because they need to estimate the transfer and filtering coefficients before removing the artifacts. BSS and EMD are not functional for single-channel signals, because they remove artifacts by decomposing and reconstructing the EEG signals in the time and frequency domains. However, the signal decomposition relies on the independence between channels that the single-channel signals do not have.

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

Researchers have applied deep learning (DL) technologies to address these issues after witnessing their breakthroughs in the fields of computer vision [23]–[29] and natural language processing [30]–[32]. Recently proposed DL-based EEG denoising approaches [33]–[37] use fully connected (FC) layers, one-dimensional convolutions (Conv1Ds), and long short-term memory (LSTM) to build end-to-end learning models. Such models can automatically output real-time results and perform well even when both the multi-channel EEG information and reference signals are unavailable.

One of the key aspects of the DL models is that artifact removal strategies are not designed by human engineers but are learned from data, so the model performance is greatly influenced by training data [38]. However, in many cases, particularly in real applications, it is highly expensive to collect high-quality training data [39]. Therefore, there is value in studying the DL model performance on limited data. On the other aspect, DL models usually run as black boxes [40]. The lack of transparency may hinder DL applications in the medical field because it is difficult for humans to verify whether a complex DL model has expert medical or signal-processing knowledge. Thus, DL models that provide the explanations for their mechanism deserve to be further explored.

In this study, we propose a novel multi-module neural network (MMNN) for EEG denoising. This network can be implemented in real-time and applied to single-channel EEG data. Our contributions are as follows: 1) Artifact removal is defined as the detachment of pure EEG signals from signals containing additive noise, therefore we create a network flow that can constantly decompose and assemble EEG information using the proposed denoising modules. 2) We designed the denoising modules using Conv1Ds and FC layers, aiming to customize a solution specialized at separating OAs or MAs from noisy EEG signals by network learning. Conv1Ds were used to extract and generalize the informative features of brain activity, and FC layers were used to reconstruct the clean signals and artifacts. Their combination acted as an end-to-end trainable filter. 3) Referring to the work of EEGdenoiseNet [37] that provided a publicly available structured database for EEG denoising studies, we compared the proposed network with the existing DL and conventional techniques under the same condition. 4) The model denoising performance when different amounts of learning data are available were explored, and the visualization of its component was discussed.

The remainder of this paper is structured as follows. The description of the experimental materials is given in Section II; Section III presents the proposed MMNN; the experiments and results are given in Section IV; and the discussion and conclusion are presented in Section V and Section VI.

II. MATERIALS

The database used in this study is summarized in Table 1. This database provides large-scale clean EEG and artifact epochs, involving 4514 clean EEG epochs, 3400 EOG epochs, and 5598 EMG epochs. In the previous DL denoising studies [36], [37], these epochs were used to synthesize the training and testing data. Their extraction process is briefly described as follows:

TABLE 1. The benchmark data used in our EEG denoising study.

	EEG	EOG	EMG
# of Epochs	4514	3400	5598
Sampling rate (Hz)	256 and 512	256	512
Bandpass filter (Hz)	1 to 80	0.3 to 10	1 to 120
# of EEG channels	1	1	1
Selected duration (s)	2	2	2
Notch filter (Hz)	50	50	50
Detrend	Yes	Yes	Yes
Manual check	Yes	Yes	Yes

The database is available at:
https://github.com/ncclabsustech/EEGdenoiseNet/tree/master/data

A. CLEAN EEG DATA

The EEG data are composed of 4514 clean EEG epochs, and each epoch is a single-channel EEG segment of 2s. As described in [36], 64-channel EEG epochs were collected from a public database of motor-imagery BCI [41]. These epochs were then band-pass filtered between 1 and 80 Hz, notch-filtered (50 Hz), detrended, and processed using independent component analysis on ICLabel [42]. Finally, the processed epochs were sampled at 256 and 512 Hz, respectively, cut into single-channel epochs, and manually checked to ensure that each one was clean.

B. ELECTROOCULOGRAM (EOG) DATA

The EOG data contain 3400 single-channel OA epochs with a sample rate of 256 Hz. These epochs were extracted from previous studies [43]–[47]. As described in [36], the data were bandpass filtered between 0.3 and 10 Hz, notch-filtered (50 Hz), and detrended. The extracted OAs were subsequently segmented into 2s per epoch and visually checked by experts.

C. ELECTROMYOGRAPHY (EMG) DATA

The EMG data consist of 5598 MA epochs, and each epoch is a single-channel EEG epoch with a duration of 2s and a sampling rate of 512 Hz. These MA epochs were collected from [48] and band-pass filtered between 1 and 120 Hz. Afterwards, these epochs were notch-filtered (50 Hz), detrended, and visually checked by experts.

III. METHODS

This section describes the proposed MMNN in detail. We first define the EEG denoising problem and then describe the denoising module, which serves as the basic component in our model, followed by the model structure. Finally, we introduce the synthesis process of noisy EEG signals, as well as training and testing data for OA and MA removals.

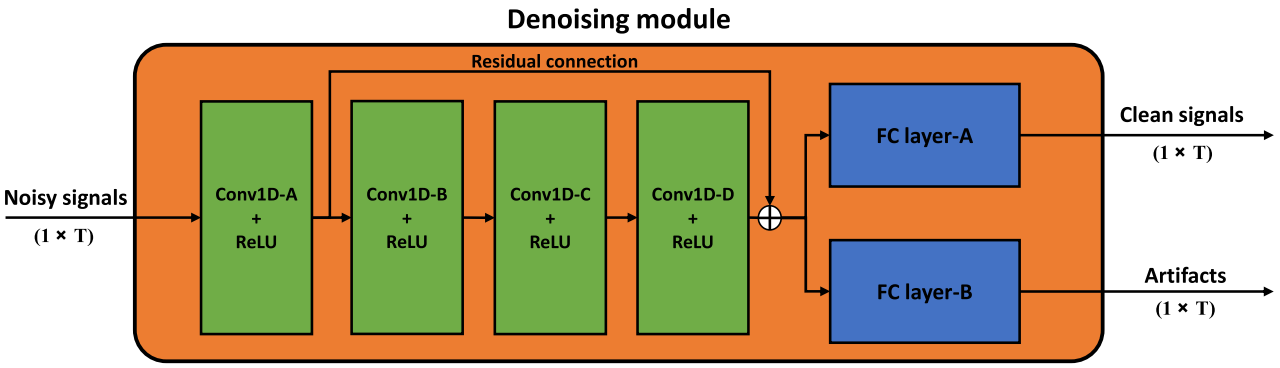


FIGURE 1. The internal structure of the denoising module.

TABLE 2. Hyperparameters of the denoising module.

Input	Noisy signals (1 × T)
Conv1d-A + ReLU	Input channel = 1; Output channel = c; Kernel size = k; Zero-padding = (k-1)/2
Conv1d-B + ReLU	Input channel = c; Output channel = c; Kernel size = k; Zero-padding = (k-1)/2
Conv1d-C + ReLU	Input channel = c; Output channel = c; Kernel size = k; Zero-padding = (k-1)/2
Conv1d-D +ReLU	Input channel = c; Output channel = c; Kernel size = k; Zero-padding = (k-1)/2
FC-layer-A	Input features = c × T; Output features = T
FC-layer-B	Input features = c × T; Output features = T
Outputs	Clean signals (1 × T) Artifacts (1 × T)

*T indicates the number of discrete time points of the input EEG epoch.

A. PROBLEM DEFINITION

OAs and MAs belong to ambient noises, also called background noises [49], [50]. They are generated independent of the clean signals, therefore the relationship among clean signals, OAs or MAs, and noisy signals in EEG recordings can be expressed as [51], [52]:

$$Y = X + Z \tag{1}$$

where X, Z and Y denote clean signals, ocular or myogenic artifacts, and noisy signals, respectively.

The essence of EEG denoising is to estimate the clean signals using the noisy signals Y. For DL denoising models, it is challenging to use the prior knowledge learned from Z distribution to filter Y [53].

B. DENOISING MODULE

In our design, the denoising module is constructed by four Conv1Ds with rectified linear units (ReLUs), a residual connection, and two FC layers, as shown in Figure 1. Table 2 provides the details of the hyperparameters, and the parameter tuning process of c and k is given in Section III. Notably, the proposed denoising module outputs both clean signals and artifacts.

1) Conv1Ds WITH ReLUs

The objective of Conv1Ds is to decompose noisy EEG signals, and their parameters need to be learned from training data. Within each Conv1D, the output channel and kernel size determine the computational complexity and filter length of feature extraction, respectively. The zero-padding operation can maintain the structural consistency between the inputs and outputs. ReLUs can improve the model's nonlinearity and avoid the vanishing gradient problem in the learning stage. Conv1D-A first dismantles noisy single-channel EEG signals into features in multiple dimensions. Conv1D-B, C, and D are subsequently used to continuously generalize the extracted features. The combination of multiple Conv1Ds with ReLUs can build complex mappings, thus dismantling EEG signals more finely. (The discussion regarding the number of Conv1Ds is provided in Appendix. A)

2) RESIDUAL CONNECTIONS

Residual connections can accelerate network convergence and improve the model's learning ability.

3) FC LAYERS

The function of the FC layers is to reconstruct clean signals and artifacts by connecting the generalized features. The output clean signals and artifacts have the same data size as the input (data size: 1 × T).

C. NETWORK STRUCTURE

The proposed MMNN is built using multiple denoising modules, their number is flexible and can be adjusted according

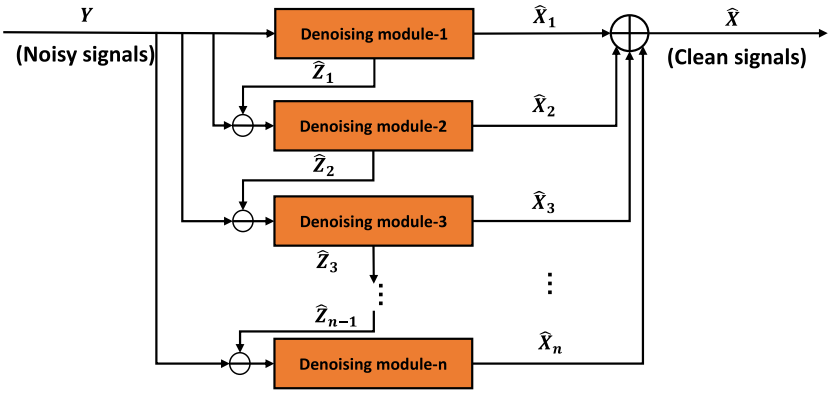


FIGURE 2. The structure of MMNN-n (Multi-Module Neural Network-n).

to different denoising tasks. An MMNN assembled using n denoising modules (MMNN-n) is shown in Figure 2, where the inputs and outputs of n denoising modules are Y, $Y - \hat{Z}_1$ to $Y - \hat{Z}_{n-1}$ and $\hat{X}_1, \hat{X}_2$ to $\hat{X}_n$. The final estimation of clean signals is the sum of $\hat{X}_1$, $\hat{X}_2$ to $\hat{X}_n$. In our designed structure, the proposed MMNN constantly purifies the inputs for each denoising module by removing the artifact estimation. Specifically, $Y - \hat{Z}_{i-1}$ replaces Y itself as the input for the i_{th} denoising module. According to (1), the former is a purer EEG signal than the latter. Therefore, there is a high probability that the outputs of the i_{th} denoising module, $\hat{X}_i$ and $\hat{Z}_i$, are closer to the ground truth of EEG signals and artifacts in theory. The related discussion is presented in Section V.

Based on the above, a workflow of multiple denoising modules can constantly improve the network performance in theory. However, multi-stacking structures may lead to vanishing gradient problem during backpropagation [54]. To hedge this risk, the proposed model is designed as a parallel architecture, thus allowing the parameters from each denoising module to be updated synchronously. The network architecture is expressed as:

$$\hat{X}_i, \hat{Z}_i = \mathcal{F}_i\left(Y - \hat{Z}_{i-1}\right) \tag{2}$$

$$\hat{X} = \mathcal{G}(Y) = \sum_{i=1}^{n} \hat{X}_i \tag{3}$$

where $\hat{X}$ is the final estimation of the clean signal; $\mathcal{G}$ is the proposed MMNN, Y is the input noisy signal, and n is the number of the denoising modules; F_i indicates the i_{th} denoising module, $\hat{X}_i$ and $\hat{Z}_i$ are the reconstructed clean signals and artifacts, respectively, and $\hat{Z}_0 = 0$.

D. NOISY SIGNAL SYNTHESIS

Using the clean EEG, EOG, and EMG epochs from the mentioned database, we synthesized noisy EEG epochs for model training and testing. The synthesized noisy EEG epochs and clean EEG epochs were the data and labels, respectively. In the training stage, the Adam optimizer [55] was adopted to minimize the mean squared error (MSE) [56] between the model outputs and labels. The details of the noisy signal synthesis are as follows.

To synthesize noisy signals with different noise levels, the signal-to-noise ratio (SNR) as a reference is first given, as shown in (4). It describes the ratio of the true signal to the background noise and is widely used to evaluate noise levels.

$$SNR = 10 \log \frac{RMS(X)}{RMS(Z)} \tag{4}$$

where X and Z are the discrete-time clean EEG signal and artifact, respectively; and RMS is the root mean squared value, as defined:

$$RMS(P) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} p_i^2} \tag{5}$$

where p_i indicates the i_{th} discrete time point in an epoch of P, and n is the number of time points in the epoch.

For signal synthesis with the given database, X denotes any clean single-channel EEG epoch, and Z is expressed as $\lambda \times N$, where N is any single-channel EOG or EMG epoch and λ is the parameter used to control the SNR of the noisy signal. By (4), λ can be derived as:

$$\lambda = \frac{RMS(X)}{10^{0.1 \times SNR} \times RMS(N)} \tag{6}$$

According to (1), a clean EEG epoch X and an EOG or EMG epoch N can be simulated into a noisy EEG epoch Y with any SNR level, as shown in (7):

$$Y = X + \lambda \times N \tag{7}$$

E. TRAINING DATA AND TESTING DATA

Previous studies [57]–[59] have demonstrated that the SNR values of OAs and MAs are commonly between -7 and 2dB, thus the noisy signals were synthesized within this SNR range.

The OA removal task was implemented using 30000 pairs and 4000 pairs of training and testing samples, respectively. The OA removal task was implemented using 30000 pairs

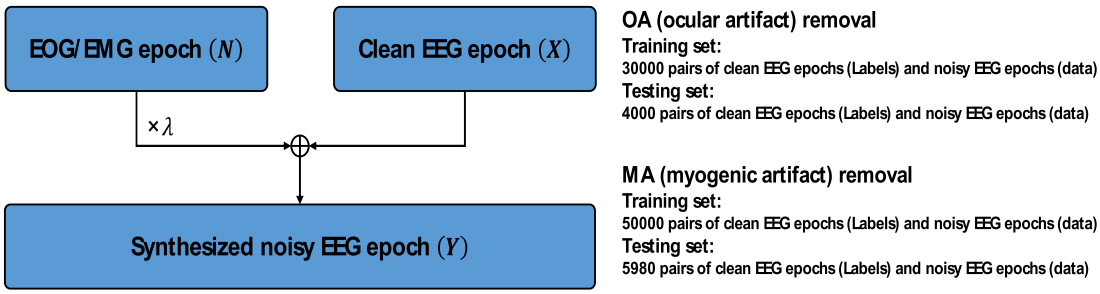


FIGURE 3. Noisy signal synthesis for model training and testing.

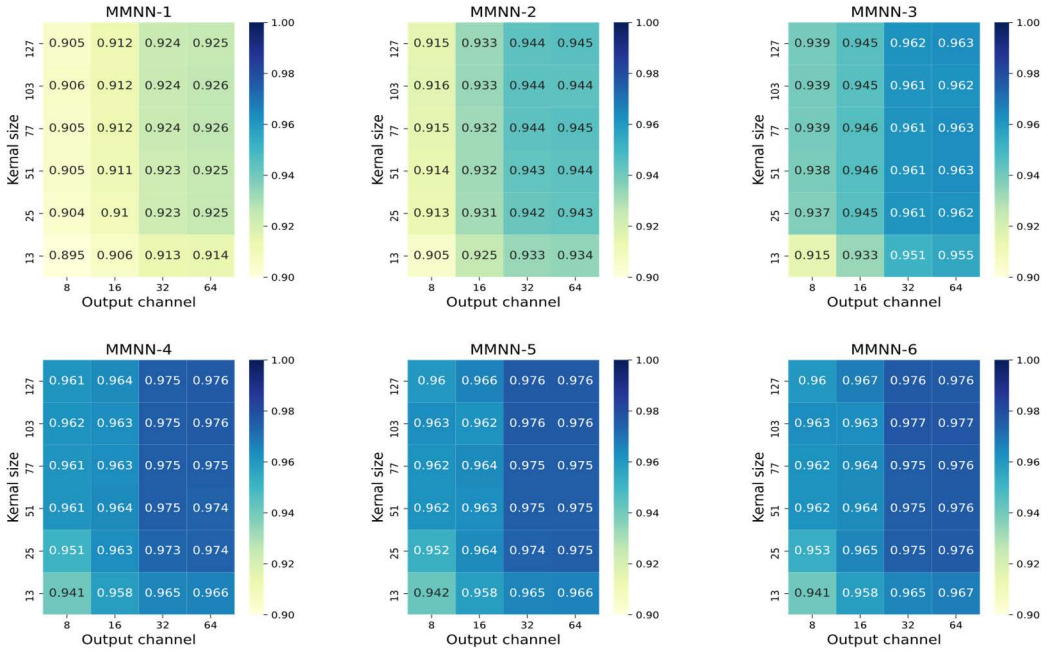


FIGURE 4. The CC results of 10-fold cross-validation in the OA removal task.

and 4000 pairs of training and testing samples, respectively. They were synthesized using 3400 clean EEG epochs (randomly selected from 4514 clean EEG epochs) and all 3400 EOG epochs, according to (6) and (7). For the noisy EEG synthesis of training set, the *SNR* values were followed a uniform distribution from −7 to 2dB, X and N were 3000 of 3400 clean epochs and EOG epochs, respectively. The testing set were synthesized using the remaining 400 pairs of EEG epochs and EOG epochs, and the *SNR* values ranged from −7dB to 2dB at an interval of one. (−7dB, −6dB, −5dB, −4dB, −3dB, −2dB, −1dB, 0dB, 1dB, 2dB).

In the MA removal task, all 4514 clean EEG epochs and 5598 EMG epochs were utilized, where we randomly copied 1084 clean EEG epochs into original EEG epochs, thus producing 5598 clean EEG epochs. Finally, we used 5000 of 5598 EEG epochs and EMG epochs to construct 50000 pairs of training samples, and the remaining 598 pairs were used to construct 5980 pairs of testing samples. The synthesis process was followed the OA removal task.

Figure 3 briefly summarizes the above process.

IV. EXPERIMENTS AND RESULTS

In this section, we first present the experimental hardware and evaluation metrics. Then, the hyperparameter tuning process of the denoising module is described. Finally, we compare the proposed model with other DL and conventional approaches through scoring and visualization.

A. HARDWARE AND EVALUATION METRICS

All the experiments were implemented using Pytorch [60] and two GeForce GTX 1080 GPUs in a Linux system. The evaluation metrics included the temporal relative root mean square error (T-RRMSE), spectral relative root mean square error (S-RRMSE), and correlation coefficient (CC), as shown in (8)(9) and (10).

$$T-RRMSE = \frac{RMS\left(\mathcal{G}\left(y\right)-x\right)}{RMS\left(x\right)} \tag{8}$$

$$S-RRMSE = \frac{RMS\left(PSD\left(\mathcal{G}\left(y\right)\right)-PSD\left(x\right)\right)}{RMS\left(PSD\left(x\right)\right)} \tag{9}$$

$$CC = \frac{Cov\left(f\left(y\right),x\right)}{\sqrt{Var\left(f\left(y\right)\right)\times Var\left(x\right)}} \tag{10}$$

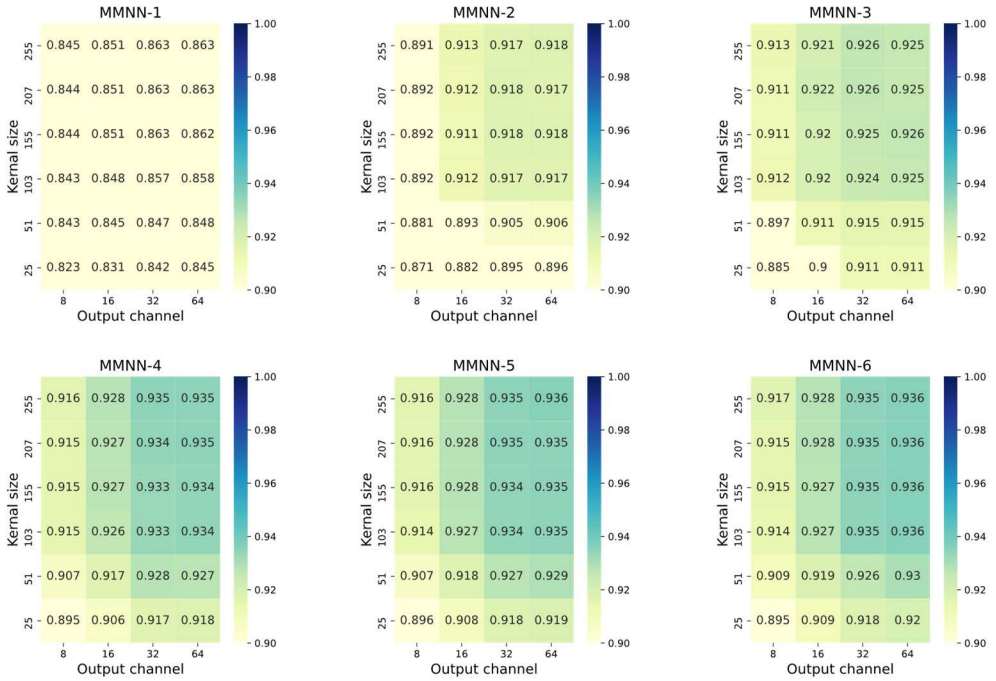


FIGURE 5. The CC results of 10-fold cross-validation in the MA removal task.

TABLE 3. Average denoising performance in the OA removal task.

Approach	T-RRMSE	S-RRMSE	CC
FCNN	0.367	0.387	0.906
Simple CNN	0.359	0.361	0.916
Complex CNN	0.336	0.343	0.923
RNN	0.411	0.389	0.900
MMNN-1	0.321	0.323	0.925
MMNN-2	0.301	0.309	0.936
MMNN-3	0.289	0.295	0.941
MMNN-4	0.273	0.279	0.958
MMNN-5	0.273	0.277	0.959
MMNN-6	0.273	0.276	0.959

Blue: Best scores of the reference models; Red: Best scores of the proposed models

TABLE 4. Average denoising performance in the MA removal task.

Approach	T-RRMSE	S-RRMSE	CC
FCNN	0.569	0.552	0.804
Simple CNN	0.639	0.643	0.787
Complex CNN	0.632	0.604	0.786
RNN	0.561	0.521	0.816
Novel CNN	0.448	0.442	0.863
MMNN-1	0.523	0.501	0.815
MMNN-2	0.434	0.415	0.878
MMNN-3	0.403	0.406	0.885
MMNN-4	0.386	0.378	0.896
MMNN-5	0.385	0.375	0.896
MMNN-6	0.385	0.376	0.896

Blue: Best scores of the reference models; Red: Best scores of the proposed models

where x and y are the clean EEG epoch and input noisy EEG epoch, respectively; $\mathcal{G}$ indicates the proposed model; *PSD* is the power spectral density function; *Cov* and *Var* are short for the covariance function and variance function. In general, the smaller the T-RRMSE and S-RRMSE values are, the closer to 1 the CC value is, the better the performance is.

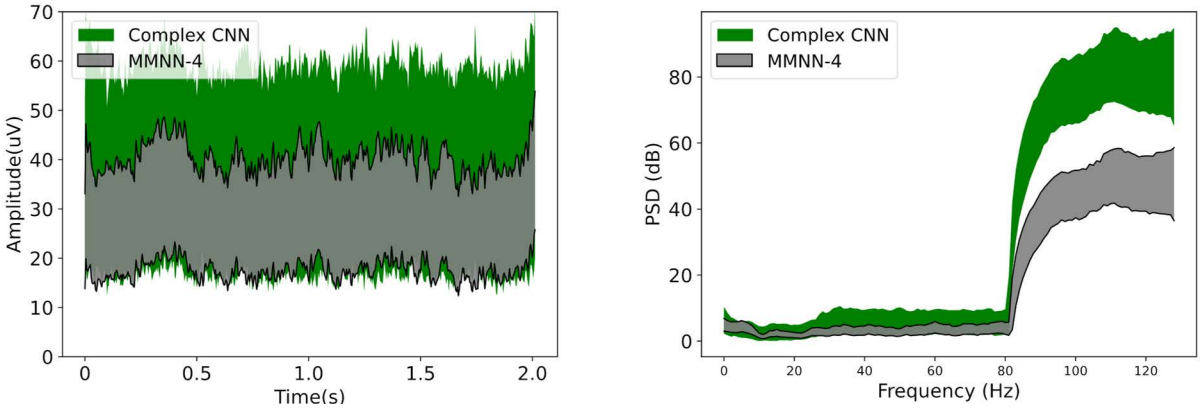


FIGURE 6. Distribution of signal deviation in the time and frequency domains for MA removal (Confidence interval = 0.95).

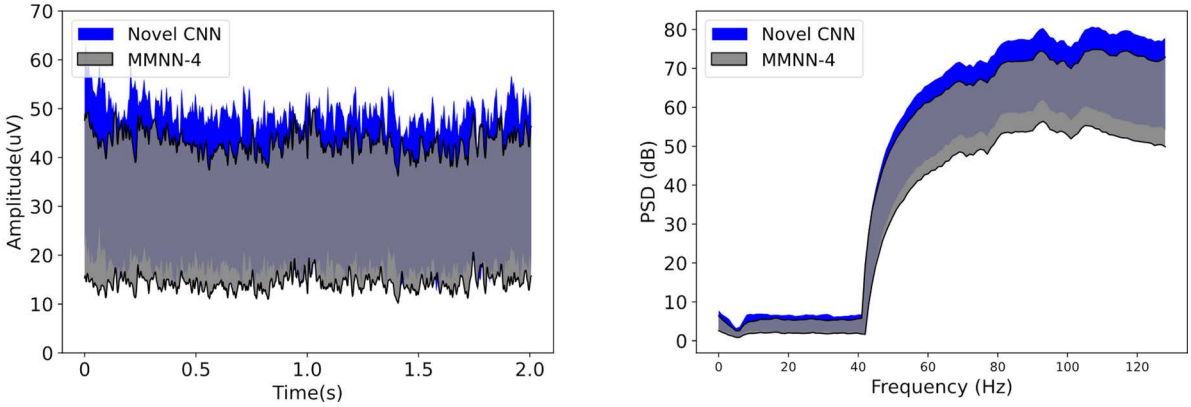


FIGURE 7. Distribution of signal deviation in the time and frequency domains for OA removal (Confidence interval = 0.95).

B. HYPERPARAMETER TUNING

To select the hyperparameters (c, k) of the denoising module, we performed a 10-fold cross-validation on the given 30000 and 50000 pairs of training samples in the OA and MA removals, respectively. In comparison with the specified validation dataset, a cross-validation strategy can avoid the problems caused by the unreasonable division of the dataset.

In the hyperparameter tuning process, the number of output channels was set as 8, 16, 32, and 64. The filter length for feature extraction was set as 0.05s, 0.1s, 0.2s, 0.3s, 0.4s, and 0.5s, which corresponds to the kernel size of 13, 25, 51, 77, 103, and 127 in the OA (256Hz) task and the kernel size of 25, 51, 103, 155, 207, and 255 in the MA (512Hz) task, respectively.

The CC results of the 10-fold cross-validation from MMNN-1 to MMNN-6 are shown in Figures 4 and 5. We can see that the model performance did not significantly improve, but the computational complexity increased when the number of output channels exceeded 32 in both the OA and MA removal tasks. Moreover, the filter lengths of 0.1s (kernel size = 25) and 0.2s (kernel size = 103) were capable of obviously improving the model performance with fewer training parameters in the OA and MA removals. Therefore, we separately chose (32, 25) and (32, 103) as the module's hyperparameters for removing OAs and MAs.

C. RESULTS OF OA AND MA REMOVALS

We performed EEG denoising using MMNN-1 to MMNN-6. The reference DL models (Appendix. B) included fully connected neural network (FCNN), simple convolution neural network (Simple CNN), complex convolution neural network (Complex CNN), and recurrent neural network (RNN) from [36], and novel Convolutional Neural Network (Novel CNN) [37]. To fairly compare the model performance under the same condition, we trained and tested the models using the same amount of dataset as the references [36], [37], as shown in Figure 3. The learning rate and batch size were 0.0001 and 128, respectively. The trained parameters of our models at the 10th iteration, were used to test the denoising performance.

Tables 3 and 4 show the denoising performance on the 4000 pairs of testing samples in the OA removal and 5980 pairs of testing samples in the MA removal. The results show that the scores of the proposed model can be constantly improved when using one to four denoising modules, whereas more than four denoising modules cannot significantly enhance its performance. Moreover, compared to the best results of the reference models, the proposed model

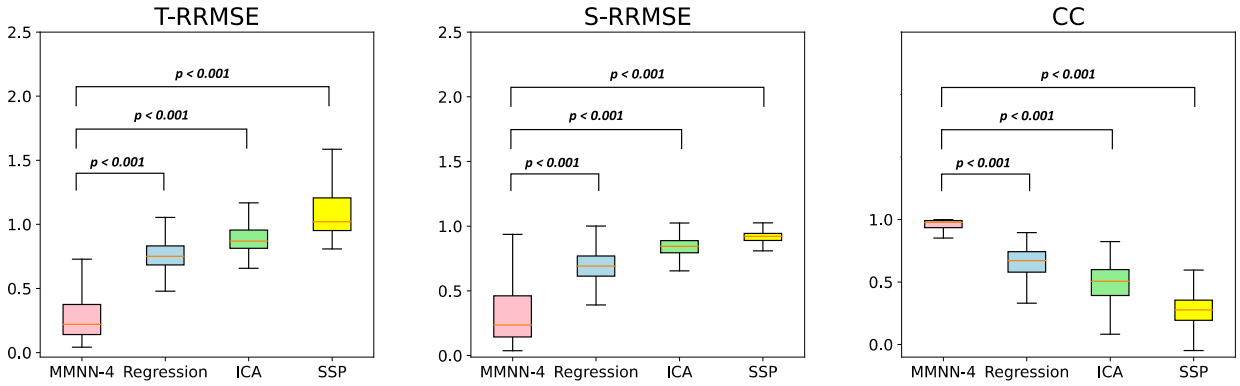


FIGURE 8. Performance comparison between the proposed model and traditional models in the OA removal.

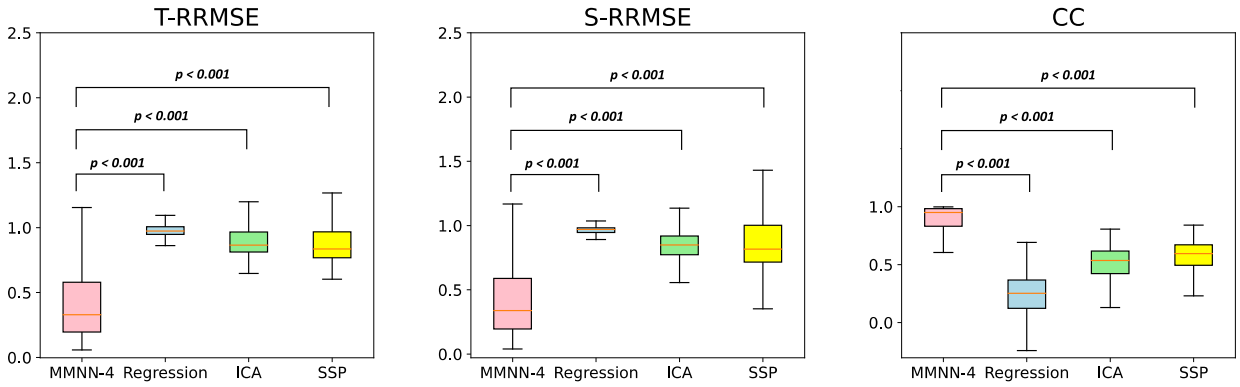


FIGURE 9. Performance comparison between the proposed model and traditional models in the MA removal.

(MMNN-4) reduced the T-RRMSE and S-RRMSE by at least 6.3% and 6.4%, respectively, and improved the CC by at least 3.5% when removing OAs. In the MA removal, it reduced the T-RRMSE and S-RRMSE by at least 6.2% and 6.4%%, respectively, and improved the CC by at least 3.3%. These results illustrate that the proposed model performs well on the given database.

Subsequently, through the visualization of the denoised results, we compared the robustness between the proposed model and the top-scoring reference models for OA and MA removals (Complex CNN and Novel CNN). As shown in Figure 6 and Figure 7, we presented the signal deviation in the time and frequency domains between the denoised results and the sample labels by calculating the absolute values of the noise-free epoch minus the denoised epoch. From the deviation results of the OA and MA removals within the 95% confidence interval, we can observe that the signal deviation of the proposed model (MMNN-4) is closer to the horizontal axis (noise-free situation) and exhibits a smaller range of deviation than the other competitors in both the time and frequency domains, which confirms the relative robustness of the proposed model.

D. PROPOSED MODELS VS CONVENTIONAL MODELS

We further compared the proposed model with three conventional models: Regression [14], ICA [61], and SSP [62]. These models are classic EEG denoising approaches applied to MNE toolbox [63]. Figure 8 and Figure 9 show the score distributions of the OA removal (4000 testing epochs) and MA removal (5980 testing epochs), respectively, where the proposed model achieves higher CC and smaller T-RRMSE and S-RRMSE scores than the conventional ones. According to the ANOVA results with Holm-Bonferroni correction, the performance differences between the proposed model and the classical approaches are significant (all p-values < 0.001) in both the OA and MA removals.

E. OA AND MA REMOVALS ON LIMITED TRAINING DATA

In the applications of DL-based EEG denoising, sufficient high-quality training data are usually unavailable. Therefore, we investigated the robustness of DL models when using limited training data, as shown in Table 5 and Table 6. The proposed MMNN-4 was compared with the top-scoring reference models for the OA and MA removals, Complex CNN, and Novel CNN, respectively, where 10 to 100% of the training data were separately selected from the given database for network learning, and the training iterations and parameters were consistent with the former settings. The results show that the proposed model always has a superior performance over its competitors when using the same amount of training data both for the OA and MA removals. Notably, our model can reach scores similar to the reference ones using only 60% of the training data when removing OAs and MAs.

TABLE 5. Denoising performance with limited training data in the OA removal.

Training data	Testing data	Complex CNN (T-RRMSE/S-RRMSE/CC)	MMNN-4 (T-RRMSE/S-RRMSE/CC)
3000	4000	0.641/0.557/0.737	0.562/0.501/0.790
6000	4000	0.543/0.507/0.810	0.495/0.459/0.840
9000	4000	0.481/0.462/0.849	0.437/0.417/0.871
12000	4000	0.445/0.454/0.872	0.382/0.375/0.902
15000	4000	0.427/0.442/0.879	0.346/0.351/0.916
18000	4000	0.410/0.438/0.892	0.322/0.340/0.929
21000	4000	0.381/0.393/0.902	0.307/0.328/0.935
24000	4000	0.366/0.377/0.908	0.298/0.303/0.945
27000	4000	0.342/0.361/0.917	0.287/0.297/0.950
30000	4000	0.336/0.343/0.923	0.273/0.279/0.958

Blue: Best scores of the reference models; Red: Better scores than the best of the reference models

TABLE 6. Denoising performance with limited training data in the MA removal.

Training data	Testing data	Novel CNN (T-RRMSE/S-RRMSE/CC)	MMNN-4 (T-RRMSE/S-RRMSE/CC)
5000	5980	0.770/0.680/0.597	0.615/0.579/0.745
10000	5980	0.597/0.554/0.751	0.537/0.518/0.804
15000	5980	0.555/0.537/0.776	0.503/0.494/0.826
20000	5980	0.516/0.515/0.814	0.474/0.453/0.843
25000	5980	0.499/0.508/0.829	0.446/0.437/0.855
30000	5980	0.485/0.489/0.840	0.430/0.430/0.865
35000	5980	0.491/0.473/0.844	0.420/0.410/0.872
40000	5980	0.462/0.464/0.851	0.415/0.404/0.874
45000	5980	0.457/0.452/0.857	0.395/0.395/0.883
50000	5980	0.448/0.442/0.863	0.386/0.378/0.896

Blue: Best scores of the reference models; Red: Better scores than the best of the reference models

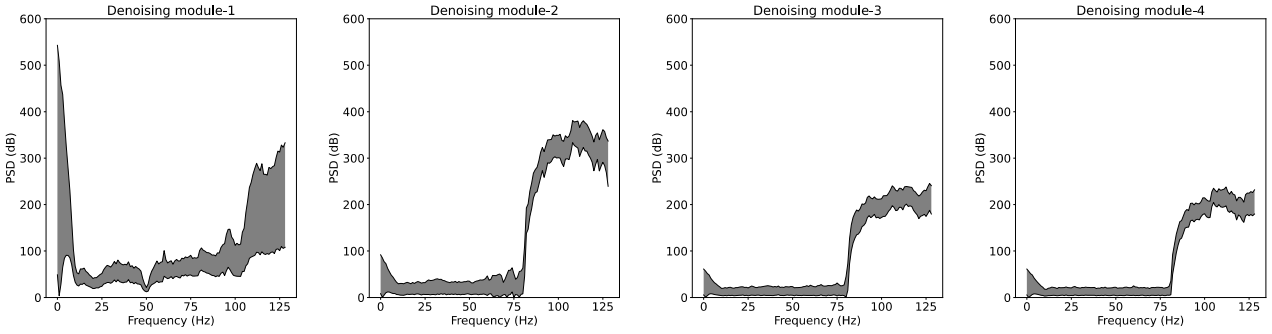


FIGURE 10. Deviation distribution between the input signals of denoising modules and clean signals for OA removal (Confidence interval = 0.95).

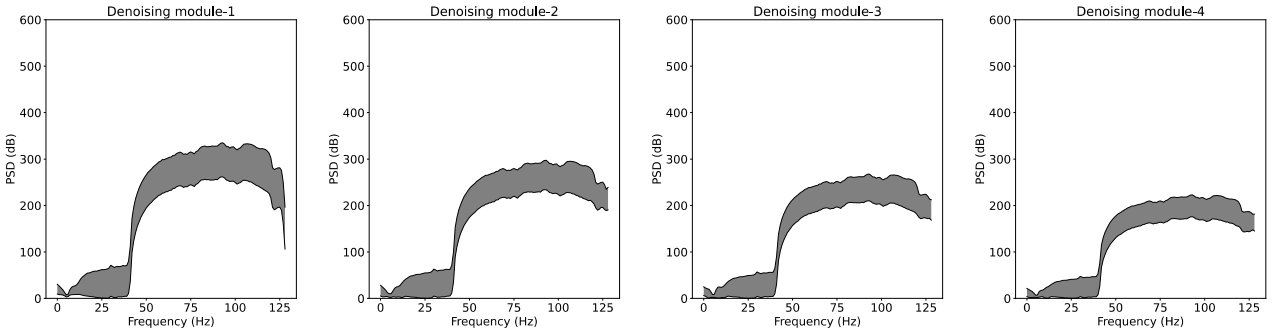


FIGURE 11. Deviation distribution between the input signals of denoising modules and clean signals for MA removal (Confidence interval = 0.95).

V. DISCUSSION

In this paper, we proposed a novel DL-based EEG denoising model called MMNN. This model achieved smaller T-RRMSE and S-RRMSE scores and higher CC scores than the other models when removing OAs and MAs. It can reach a performance similar to that of the reference DL

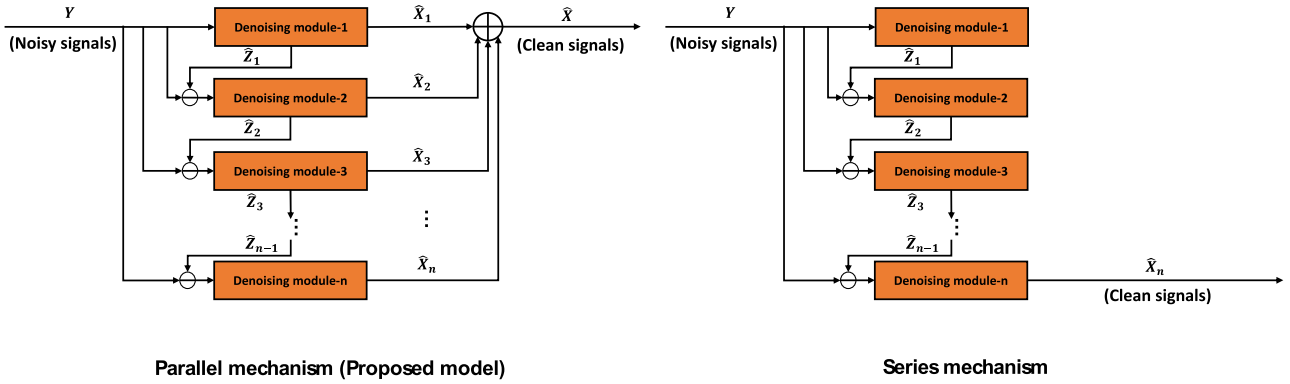


FIGURE 12. Parallel and series mechanisms.

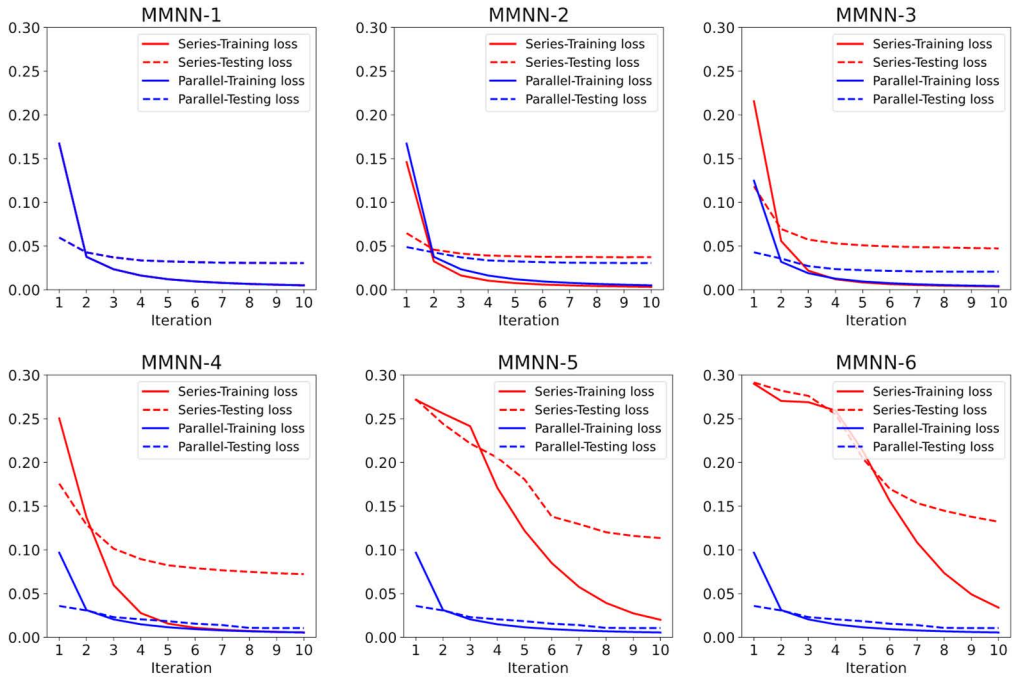


FIGURE 13. Training and testing losses using the given database in the OA removal.

models with only 60% of the training data. Through the visualization of signal deviation distribution, the performance differences between the reference and the proposed models are clearly observed in the time and frequency domains.

Overall, the proposed model has a superior performance compared to the reference DL models. There are two reasons for this. First, the proposed model enables constantly providing more purified input signals for denoising modules. As shown in Figure 10 and Figure 11, we present the signal deviation distribution between the inputs of four denoising modules and clean signals. In Figure 10, the OAs in the range of 0-80 Hz were rapidly suppressed using two denoising modules, and the OAs above 80 Hz were gradually reduced when more modules were used. In Figure 11, the MAs of the input signals were suppressed by degrees from the first to the fourth module. Second, the parallel architecture of the proposed model allows the gradients to flow through each denoising module directly in the backpropagation, thereby avoiding the vanishing gradient problem and enhancing the network learning ability. For further clarification, the parallel and series mechanisms of the denoising modules are presented in Figure 12. The training and testing losses of the two mechanisms (batch size $=$ 128 and learning rate $=$ 0.0001) using the given database are shown in Figure 13 and Figure 14. The followings were observed, respectively: 1) both the two models converge in 10 iterations; 2) the training and testing losses of the parallel mechanism are smaller than those of the series mechanism when more than two denoising modules are assembled within our model, which illustrates that the parallel mechanism of the proposed model possesses a stronger learning capacity; 3) for the series

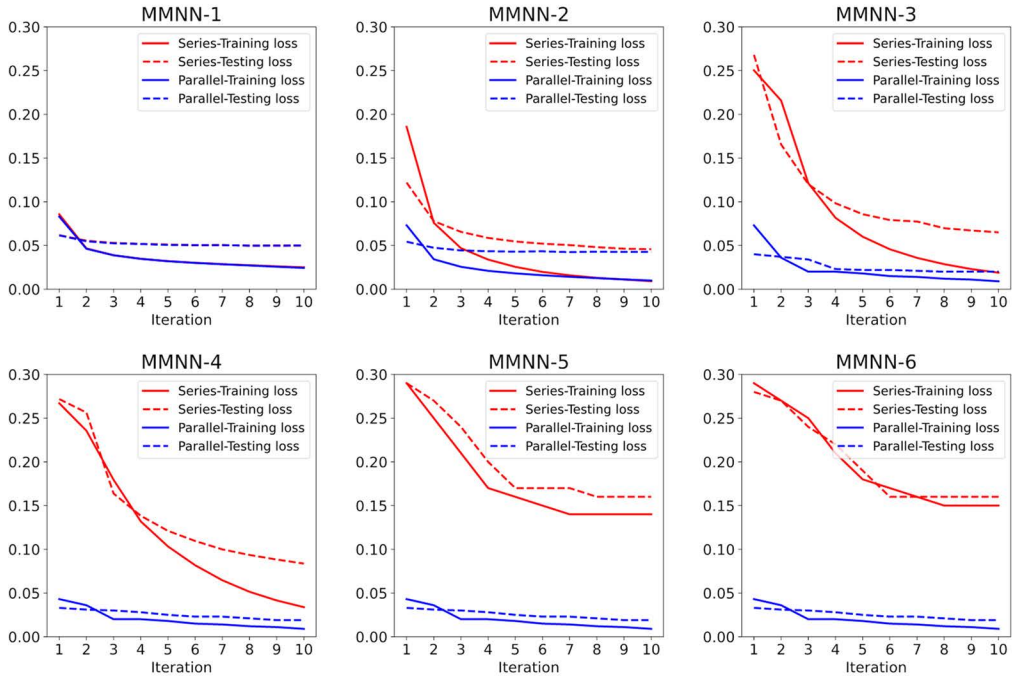


FIGURE 14. Training and testing losses using the given database in the OA removal.

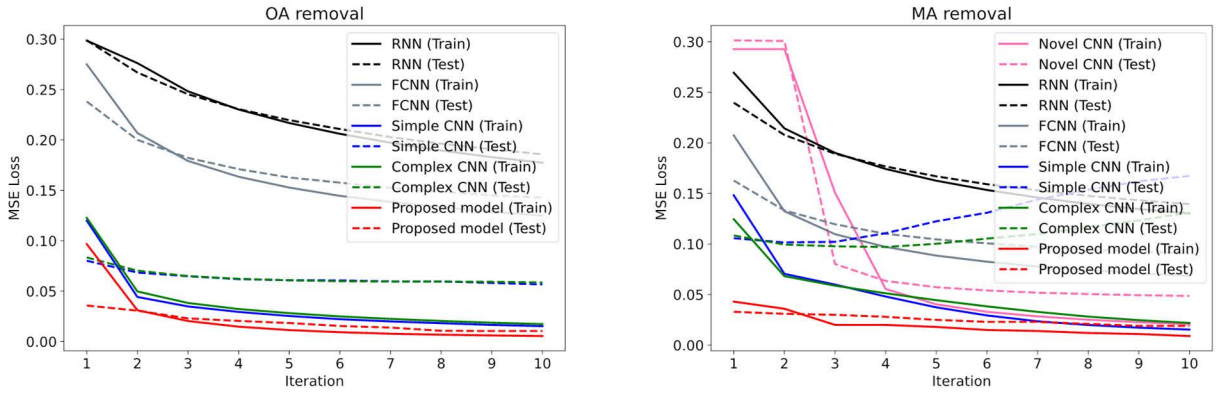
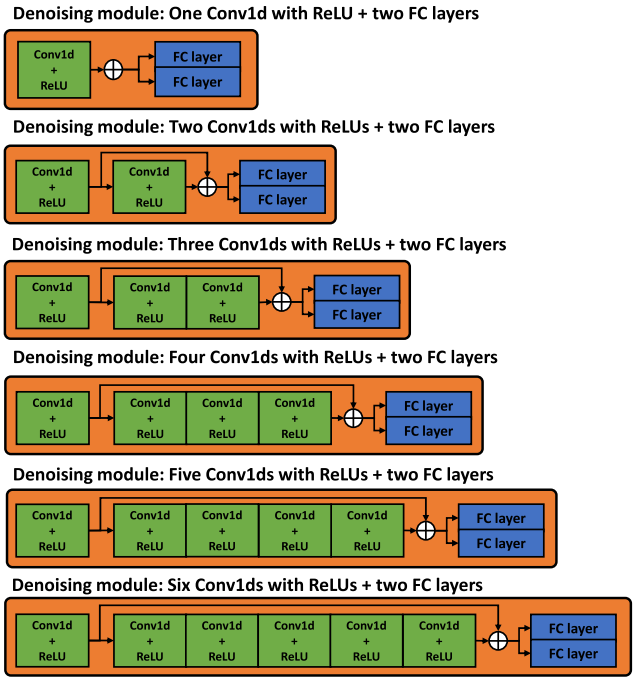


FIGURE 15. Loss comparison between the proposed model and the other DL models.

mechanism, the network learning capacity is weakened when more denoising modules were stacked in the model, which is possibly caused by the vanishing gradient problem in the learning process; 4) in contrast with the series mechanism, the parallel mechanism can improve the learning capacity when more denoising modules were used. However, there was a limitation to the improvement of network learning. We can see that the training and testing losses of MMNN-4, MMNN-5, and MMNN-6 are almost the same for the parallel model, which explains their similar scores in the experiment. Furthermore, the loss comparison between the proposed MMNN-4 and the other DL models is given in Figure 15, where our model has smaller training and testing losses and can converge faster than the others, which is possibly the reason why it performs well with fewer training data in both the OA and MA removals.

In the future, there are some challenges worth exploring using the proposed model. Specifically, we used the denoising modules of the different filter sizes in the OA and MA removals. Whether the filter size of feature extraction is caused by noise feature differences should be further studied. Moreover, OAs and MAs are entangled with motion artifacts in a real EEG epoch, however, for the mixed signals, there is no available public database to evaluate the model performance [64]. Given that the proposed model offers significant advantages over the conventional and DL models in this study, the related research is within the scope of the further work.



	Number of parameters	Running time (s)	CC
OA removal	16779328	189.3	0.886
MA removal	67112000	767.8	0.711

	Number of parameters	Running time (s)	CC
OA removal	16813152	191.1	0.892
MA removal	67145824	787.3	0.743

	Number of parameters	Running time (s)	CC
OA removal	16846976	202.5	0.911
MA removal	67179648	834.1	0.778

	Number of parameters	Running time (s)	CC
OA removal	16880800	205.2	0.925
MA removal	67213472	913.5	0.815

	Number of parameters	Running time (s)	CC
OA removal	16914624	212.3	0.925
MA removal	67247296	977.2	0.815

	Number of parameters	Running time (s)	CC
OA removal	16948448	221.9	0.925
MA removal	67281120	1025.6	0.815

FIGURE 16. Configuration of the denoising modules.

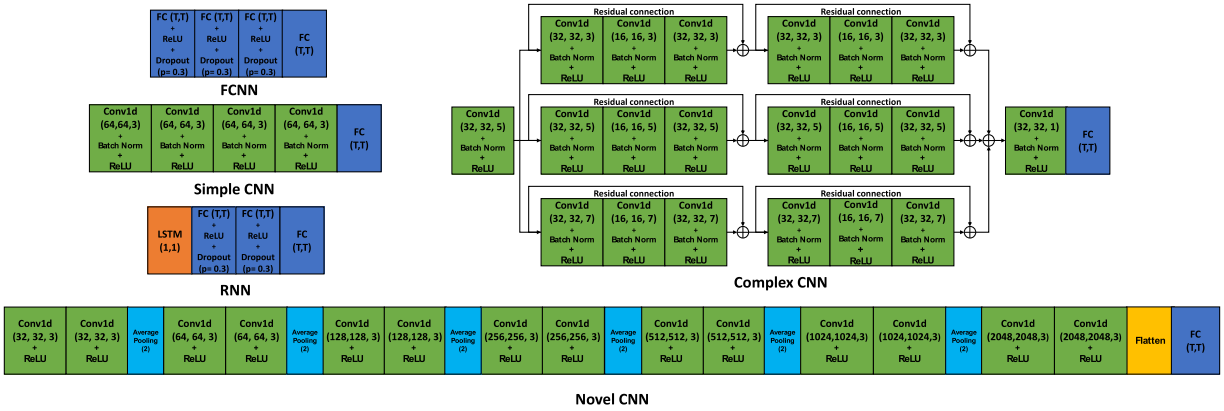


FIGURE 17. Architecture of the reference models. Hyperparameter formats of 'Conv1D", "FC", and "LSTM" are (Input_channel, Output_channel, Kernel_size), (Input_feature, Output_feature), (Input_channel, Output_channel), respectively. "T" the number of discrete time points of the input EEG epoch.

VI. CONCLUSION

A novel MMNN (multi-module neural network) is proposed in this study, which is a parallel architecture assembled with multiple denoising modules. The results revealed that the proposed model can automatically remove OAs and MAs from single-channel noisy EEG signals. Compared to the existing models, it achieved higher signal reconstruction accuracy and reached this goal with less training data. In the future, we expect that our model will play a critical role in EEG denoising applications.

APPENDIX A
CONFIGURATION OF DENOISING MODULES

Figure 16 shows the different configurations of the denoising modules, in which the number of parameters, and running time increased with the number of Conv1Ds. However, the module performance reaches its limit when using four Conv1Ds. Therefore, we configured four Conv1Ds for the denoising module in our model.

APPENDIX B
ARCHITECTURE OF THE REFERENCE MODELS

Figure 17 presents the architecture of the reference models, including FCNN, Simple CNN, Complex CNN, and RNN from [36], and Novel CNN from [37]. These models decomposed noisy EEG signals using the different combinations of Conv1D, FC, and LSTM blocks, and then reconstructed clean signals using an FC block.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude for the invaluable help from Dr. Ruimin Wang, Dr. Shinichi Yoshida, Dr. Masaki Takeda, and Dr. Kiminori Matsuzaki.

REFERENCES

[1] M. Tacke, K. Janson, K. Vill, F. Heinen, L. Gerstl, K. Reiter, J. Remi, and I. Borggraefe, "The influence of information about the circumstances of EEG recordings on the ability to identify seizure patterns," *Seizure*, vol. 88, pp. 125–129, May 2021.

[2] P. M. Rodrigues, B. C. Bispo, C. Garrett, D. Alves, J. P. Teixeira, and D. Freitas, "Lacsogram: A new EEG tool to diagnose Alzheimer's disease," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 9, pp. 3384–3395, Sep. 2021.

[3] C. Biancardi, G. Sesso, G. Masi, U. Faraguna, and F. Sicca, "Sleep EEG microstructure in children and adolescents with attention deficit hyperactivity disorder: A systematic review and meta-analysis," *Sleep*, vol. 44, no. 7, Jul. 2021, Art. no. zsab006.

[4] Y. Huang, A. Mohan, D. D. Ridder, S. Sunaert, and S. Vanneste, "The neural correlates of the unified percept of alcohol-related craving: A fMRI and EEG study," *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, Dec. 2018.

[5] R. Rajkumar, C. R. Brambilla, T. Veselinović, J. Bierbrier, C. Wyss, S. Ramkiran, L. Orth, M. Lang, E. R. Kops, and J. Mauler, "Excitatory—Inhibitory balance within EEG microstates and resting-state fMRI networks: Assessed via simultaneous trimodal PET-MR-EEG imaging," *Transl. Psychiatry.*, vol. 11, no. 1, pp. 1–15, 2021.

[6] U. Halsband and T. Gerhard Wolf, "Functional changes in brain activity after hypnosis: Neurobiological mechanisms and application to patients with a specific phobia—Limitations and future directions," *Int. J. Clin. Experim. Hypnosis*, vol. 67, no. 4, pp. 449–474, Oct. 2019.

[7] A. Myrden and T. Chau, "A passive EEG-BCI for single-trial detection of changes in mental state," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 345–356, Apr. 2017.

[8] B. Kerous, F. Skola, and F. Liarokapis, "EEG-based BCI and video games: A progress report," *Virtual Reality*, vol. 22, no. 2, pp. 119–135, Jun. 2018.

[9] J. L. Whitton, F. Lue, and H. Moldofsky, "A spectral method for removing eye movement artifacts from the EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 44, no. 6, pp. 735–741, Jun. 1978.

[10] R. D. O'Donnell, J. Berkhout, and W. R. Adey, "Contamination of scalp EEG spectrum during contraction of cranio-facial muscles," *Electroencephalogr. Clin. Neurophysiol.*, vol. 37, no. 2, pp. 145–151, 1974.

[11] M. A. Shaffer, "Problem record of the month, No. 3: Asymmetrical eye-blink artifact," *Amer. J. EEG Technol.*, vol. 10, no. 4, pp. 153–156, Dec. 1970.

[12] O. P. Mathew, Y. K. Abu-Osba, and B. T. Thach, "Influence of upper airway pressure changes on genioglossus muscle respiratory activity," *J. Appl. Physiol.*, vol. 52, no. 2, pp. 438–444, Feb. 1982.

[13] R. J. Croft and R. J. Barry, "Removal of ocular artifact from the EEG: A review," *Neurophysiol. Clin./Clin. Neurophysiol.*, vol. 30, no. 1, pp. 5–19, Feb. 2000.

[14] G. Gratton, M. G. H. Coles, and E. Donchin, "A new method for off-line removal of ocular artifact," *Electroencephalogr. Clin. Neurophysiol.*, vol. 55, no. 4, pp. 468–484, Apr. 1983.

[15] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[16] P. He, G. Wilson, C. Russell, and M. Gerschutz, "Removal of ocular artifacts from the EEG: A comparison between time-domain regression method and adaptive filtering method using simulated data," *Med. Biol. Eng. Comput.*, vol. 45, no. 5, pp. 495–503, Apr. 2007.

[17] S. Boudet, L. Peyrodie, G. Forzy, A. Pinti, H. Toumi, and P. Gallois, "Improvements of adaptive filtering by optimal projection to filter different artifact types on long duration EEG recordings," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 234–249, Oct. 2012.

[18] H. Hallez, M. D. Vos, B. Vanrumste, P. V. Hese, S. Assecondi, K. V. Laere, P. Dupont, W. V. Paesschen, S. V. Huffel, and I. Lemahieu, "Removing muscle and eye artifacts using blind source separation techniques in ictal EEG source imaging," *Clin. Neurophysiol.*, vol. 120, no. 7, pp. 1262–1272, 2009.

[19] M. A. Klados, C. Papadelis, C. Braun, and P. D. Bamidis, "REG-ICA: A hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts," *Biomed. Signal Process. Control*, vol. 6, no. 3, pp. 291–300, Jul. 2011.

[20] L. Shoker, S. Sanei, and M. A. Latif, "Removal of eye blinking artifacts from EEG incorporating a new constrained BSS algorithm," in *Proc. Process. Workshop Proc., Sensor Array Multichannel Signal*, Jul. 2004, pp. 177–181.

[21] K. H. Ting, P. C. W. Fung, C. Q. Chang, and F. H. Y. Chan, "Automatic correction of artifact from single-trial event-related potentials by blind source separation using second order statistics only," *Med. Eng. Phys.*, vol. 28, no. 8, pp. 780–794, Oct. 2006.

[22] K. T. Sweeney, S. F. McLoone, and T. E. Ward, "The use of ensemble empirical mode decomposition with canonical correlation analysis as a novel artifact removal technique," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 97–105, Jan. 2013.

[23] J.-E. Liu and F.-P. An, "Image classification algorithm based on deep learning-kernel function," *Sci. Program.*, vol. 2020, pp. 1–14, Jan. 2020.

[24] M. Kim, G.-S. Jeng, I. Pelivanov, and M. O'Donnell, "Deep-learning image reconstruction for real-time photoacoustic system," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3379–3390, Nov. 2020.

[25] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super resolution: A survey," 2019, *arXiv:1902.06068*.

[26] D. A. Wood, S. Kafiabadi, A. Al Busaidi, E. L. Guilhem, J. Lynch, M. K. Townend, A. Montvila, M. Kiik, J. Siddiqui, N. Gadapa, M. D. Benger, A. Mazumder, G. Barker, S. Ourselin, J. H. Cole, and T. C. Booth, "Deep learning to automate the labelling of head MRI datasets for computer vision applications," *Eur. Radiol.*, vol. 32, no. 1, pp. 725–736, Jan. 2022.

[27] M. H. Alkinani, W. Z. Khan, and Q. Arshad, "Detecting human driver inattentive and aggressive driving behavior using deep learning: Recent advances, requirements and open challenges," *IEEE Access*, vol. 8, pp. 105008–105030, 2020.

[28] S. Oprea, P. Martinez-Gonzalez, A. Garcia-Garcia, J. A. Castro-Vargas, S. Orts-Escolano, J. Garcia-Rodriguez, and A. Argyros, "A review on deep learning techniques for video prediction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 2806–2826, Jun. 2022.

[29] J. Guo, H. He, T. He, L. Lausen, M. Li, H. Lin, X. Shi, C. Wang, J. Xie, and S. Zha, "GluonCV and GluonNLP: Deep learning in computer vision and natural language processing," *J. Mach. Learn. Res.*, vol. 21, no. 23, pp. 1–7, 2020.

[30] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.

[31] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 604–624, Feb. 2021.

[32] M. Gardner, J. Grus, M. Neumann, O. Tafjord, P. Dasigi, N. Liu, M. Peters, M. Schmitz, and L. Zettlemoyer, "AllenNLP: A deep semantic natural language processing platform," 2018, *arXiv:1803.07640*.

[33] B. Yang, K. Duan, C. Fan, C. Hu, and J. Wang, "Automatic ocular artifacts removal in EEG using deep learning," *Biomed. Signal Process. Control*, vol. 43, pp. 148–158, May 2018.

[34] W. Sun, Y. Su, X. Wu, and X. Wu, "A novel end-to-end 1D-ResCNN model to remove artifact from EEG signals," *Neurocomputing*, vol. 404, pp. 108–121, Sep. 2020.

[35] C. Hanrahan. (2019). *Noise Reduction in EEG Signals Using Convolutional Autoencoding Techniques*. Masters University of Dublin [Online]. Available: https://arrow.tudublin.ie/scschcomdis/188/

[36] H. Zhang, M. Zhao, C. Wei, D. Mantini, Z. Li, and Q. Liu, "EEGdenoiseNet: A benchmark dataset for deep learning solutions of EEG denoising," *J. Neural Eng.*, vol. 18, no. 5, Oct. 2021, Art. no. 056057.

[37] H. Zhang, C. Wei, M. Zhao, Q. Liu, and H. Wu, "A novel convolutional neural network model to remove muscle artifacts from EEG," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 1265–1269.

[38] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, Sep. 2015.

[39] X. Li, H. Jiang, K. Zhao, and R. Wang, "A deep transfer nonnegativity-constraint sparse autoencoder for rolling bearing fault diagnosis with few labeled data," *IEEE Access*, vol. 7, pp. 91216–91224, 2019.

[40] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," 2017, *arXiv:1708.08296*.

[41] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. C. Jun, "EEG datasets for motor imagery brain computer interface," *Giga Sci.*, vol. 6, no. 7, pp. 1–8, 2017.

[42] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "ICLabel: An automated electroencephalographic independent component classifier, dataset, and website," *NeuroImage*, vol. 198, pp. 181–197, Sep. 2019.
[43] S. Kanoga, M. Nakanishi, and Y. Mitsukura, "Assessing the effects of voluntary and involuntary eyeblinks in independent components of electroencephalogram," *Neurocomputing*, vol. 193, pp. 20–32, Jun. 2016.
[44] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey," *Clin. Neurophysiol.*, vol. 118, no. 3, pp. 480–494, 2007.
[45] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, "Separability of four-class motor imagery data using independent components analysis," *J. Neural Eng.*, vol. 3, no. 3, p. 208, 2006.
[46] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clin. Neurophysiol.*, vol. 118, no. 1, pp. 98–104, Jan. 2007.
[47] C. Brunner, M. Naeem, R. Leeb, B. Graimann, and G. Pfurtscheller, "Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis," *Pattern Recognit. Lett.*, vol. 28, no. 8, pp. 957–964, 2007.
[48] V. Rantanen, M. Ilves, A. Vehkaoja, A. Kontunen, J. Lylykangas, E. Makela, M. Rautiainen, V. Surakka, and J. Lekkala, "A survey on the feasibility of surface EMG in facial pacing," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 1688–1691.
[49] Y. Yasui, "A brainwave signal measurement and data processing technique for daily life applications," *J. Physiolog. Anthropol.*, vol. 28, no. 3, pp. 145–150, 2009.
[50] J. T. Jacobson and C. A. Jacobson, "The effects of noise in transient EOAE newborn hearing screening," *Int. J. Pediatric Otorhinolaryngol.*, vol. 29, no. 3, pp. 235–248, Jun. 1994.
[51] D. Safieddine, A. Kachenoura, L. Albera, G. Birot, A. Karfoul, A. Pasnicu, A. Biraben, F. Wendling, L. Senhadji, and I. Merlet, "Removal of muscle artifact from EEG data: Comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–15, Dec. 2012.
[52] G. Gomez-Herrero, W. D. Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen, "Automatic removal of ocular artifacts in the EEG without an EOG reference channel," in *Proc. 7th Nordic Signal Process. Symp. (NORSIG)*, Jun. 2006, pp. 130–133.
[53] P. J. Allen, G. Polizzi, K. Krakow, D. R. Fish, and L. Lemieux, "Identification of EEG events in the MR scanner: The problem of pulse artifact and a method for its subtraction," *NeuroImage*, vol. 8, no. 3, pp. 229–239, Oct. 1998.
[54] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 6, no. 2, pp. 107–116, Jun. 1998.
[55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[56] M. Tuchler, A. C. Singer, and R. Koetter, "Minimum mean squared error equalization using *a priori* information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.
[57] G. Wang, C. Teng, K. Li, Z. Zhang, and X. Yan, "The removal of EOG artifacts from EEG signals using independent component analysis and multivariate empirical mode decomposition," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 5, pp. 1301–1308, Sep. 2016.
[58] X. Chen, H. Peng, F. Yu, and K. Wang, "Independent vector analysis applied to remove muscle artifacts in EEG data," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1770–1779, Jul. 2017.
[59] W. D. Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2583–2587, Nov. 2006.
[60] 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.
[61] J. Dammers, M. Schiek, F. Boers, C. Silex, M. Zvyagintsev, U. Pietrzyk, and K. Mathiak, "Integration of amplitude and phase statistics for complete artifact removal in independent components of neuromagnetic recordings," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 10, pp. 2353–2362, Oct. 2008.
[62] M. A. Uusitalo and R. J. Ilmoniemi, "Signal-space projection method for separating MEG or EEG into components," *Med. Biol. Eng. Comput.*, vol. 35, no. 2, pp. 135–140, 1997.
[63] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, pp. 446–460, Feb. 2014.
[64] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.

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 at 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, in 1996, the M.E. degree from the School of Information and Communication, Harbin Engineering University, China, in 2010. He is currently a Professor at 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 (KUT), Kochi, Japan, in 1997, as an Associate Professor and became a Professor, in 2002. 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, CA, USA. He also worked as the Director of the Research Institute, KUT, where he is currently a Professor with the School of Information. 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.