Research on recognition of O-MI based on CNN combined with SST and LSTM*

LI Penghai** **and LIU Cong**
School of Integrated Circuit Science and Engineering, Tianjin University of Technology, Tianjin 300384, China

(Received 3 April 2022; Revised 22 May 2022)

Recognition algorithms have been widely used in brain computer interface (BCI) for neural paradigms classification. To improve the classification and recognition effect of motor imagery with motor observation (O-MI) in BCI rehabilitation technology, this paper explores the function of convolutional neural network (CNN) combined with synchrosqueezed wavelet transform (SST) and long short-term memory (LSTM) in the recognition and classification of neural activities in the brain motor area. Combining the advantages of SST in signal feature extraction in the pretreatment stage and the ability of LSTM network in time series information modeling, the purpose is to make up for CNN's shortcomings in both aspects. This paper verifies the algorithm on the self-collected O-MI experimental datasets and the public datasets (BCI competition IV datasets 2a). The results show that the composite CNN algorithm incorporating SST and LSTM achieves higher classification accuracy than classic algorithms and the similar new method which is CNN combined with discrete wavelet transform (DWT) and power spectral density (PSD), so it is convenient for practical application in O-MI BCI system.
Document code: A **Article ID:** 1673-1905(2022)09-0566-6
DOI https://doi.org/10.1007/s11801-022-2054-1

Brain computer interface (BCI)-based neural rehabilitation technology mainly includes two ways, motor imagery (MI) and motor observation (MO)[1-3]. MI requires patients to actively repeat the limb movement imagination in their minds. MO requires patients to watch their own limbs or other people performing rehabilitation tasks. None of the two means involves actual limb movements[4]. In the process of MI and MO, a large number of neurons in the sensory motor area of the cerebral cortex will be activated to produce regular electrical activities, so this training method can enhance brain functional potential and rebuild motor nerve pathways[5].

Algorithm recognition effect is particularly important for accurately identifying the electrophysiological activities in the brain. The traditional method to identify electroencephalogram (EEG) signal is event-related desynchronization or synchronization (ERD/ERS), its strength lies in that it can intuitively reflect the characteristic phenomenon of MI and MO neural activities, which is convenient for direct observation[6-9], but it cannot generate ideal efficiency and classification accuracy results when dealing with big data set.

The main classification methods in MI include linear discriminant analysis (LDA), support vector machines (SVM), convolutional neural network (CNN), Bayesian classifier, and so on. The advantage of LDA is simple algorithm implementation and good real-time data processing ability. The disadvantage is that it depends too much on the quality of training set data. SVM is strong in dealing with the data of binary classification, but it involves uncertainty in parameters selection. CNN excels in handling fuzzy information, but each type of CNN network has its defects and cannot be well generalized. Bayes classifier exhibits strength of fast computing speed and high algorithm efficiency, but its classification accuracy is slightly weaker compared with the algorithms mentioned above.

Classical algorithms such as common spatial pattern (CSP) + SVM and wavelet packet decomposition (WPD) + SVM have their limitations because SVM can hardly determine the kernel function for high-dimensional data processing, which poses more difficulty in classification, and SVM only gives good results for two-category experiment with small sample statistics, but it is not ideal for classification with a large data set.

CNN has the following downsides. The pooling layer may lose a lot of valuable information, and consequently CNN neglects the correlation between local parts and the overall entity. The separate treatment process on feature extraction rules out the possibility to change the treatment process by altering parameters, and hence makes it more difficult to improve the network performance. However, this method still surpasses traditional methods in terms of efficiency and classification accuracy when

* This work has been supported by the Natural Science Foundation of Tianjin City (Nos.18JCYBJC95400 and 19JCTPJC56000).
** E-mail: lph1973@tju.edu.cn

dealing with massive EEG data[10]. In order to further improve the accuracy of classification, this paper adopts the synchrosqueezed wavelet transform (SST) and long short-term memory (LSTM) composite CNN method so as to combine the advantages of SST[11] in signal feature extraction and the strength of LSTM network in time series information modeling[12].

The strength of MI and MO is merged in this research to improve the rehabilitation effect, and motor imagery with motor observation (O-MI) paradigm is adopted as the paradigm in this paper. At the same time, in order to improve the recognition and classification effect of neural activities in brain motor area, SST is used at first to extract the characteristic frequency band of the original data, mainly α frequency band (8—13 Hz) and β frequency band (13—30 Hz). Then the processed EEG data are sent to CNN for motor imagery feature extraction. Finally, considering the excel modeling ability of LSTM network on date series, the motor imagery feature sequence composed of three channels (CZ, C3, C4) data is successively input into LSTM network, so that multi-channel combined features are extracted, which is used to improve the effect of recognition and classification under O-MI paradigm. The processing flow is shown in Fig.1. To verify the effectiveness of this method, the research results are compared with those generated by other methods.

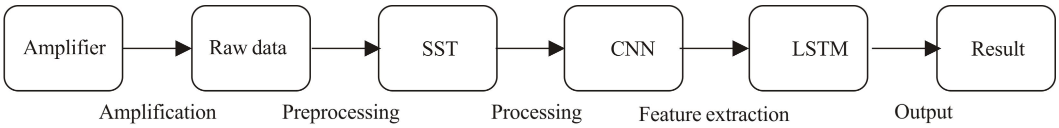

Fig.1 Data processing flow chart

The experiment was performed in the Lab of EEG Acquisition and Application, Tianjin University of Technology, China. The experimental equipment Grael is an electrophysiological amplifier developed by Compumedics Company in Australia. It has 32 channels, and the electrode position follows the setting of the international 10/20 system. In this paper, EEG data of 10 college students with good physical and mental health conditions were collected, including 7 males and 3 females, aged 20—24, who had not participated in BCI test before. Before the experiment, the subjects were told the points for attention and procedures in the experiment process, and signed the experimental informed consent. In the whole experiment, the subjects were asked to keep their bodies relaxed and sit on the chair quietly.

The upper limb rehabilitation movement is to turn over the hands, and the lower limb rehabilitation movement is to point the ground alternately with both feet.

The experiment under each mode was repeated in two groups, with 30 trials in each group, and a total of 60 trials, the upper limb movements and the lower limb movements are equal in number.

The sequence of the experimental mode is shown in Fig.2. In the −2—0 s stage, the white word "Ready" appears in the center of the black screen on the monitor and the loudspeaker sends out the warning voice, announcing that the experiment is about to start, and the subjects should keep quiet and concentrated. From 0—6 s, the scene of hands motion or feet motion appears in the center of the screen. During this period, the subjects are required to imagine the same motions shown in the video. Then the white warning word "Rest" appears in the center of the black screen during the 6—8 s, reminding the subjects to relax and rest to relieve fatigue, announcing the end of the experiment. After the rest, the next trial resumes and the subjects repeat the above process.

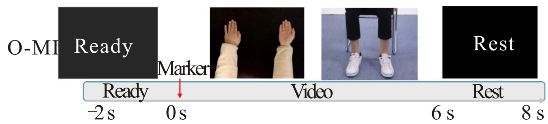

Fig.2 Sequence of the experimental mode

SST is a method combining wavelet analysis and reallocation, which is derived from empirical mode decomposition (EMD)[6]. Compared with the traditional short time Fourier transform (STFT)[7], SST's signal decomposition produces more significant frequency domain characteristics.

In order to demonstrate the advantages of SST data processing, this paper uses simulated EEG signal to generate SST and STFT time-frequency spectrum, so as to compare the accuracy of the two methods in the analysis of signal frequency-domain characteristics. The signal frequency band is 9—10 Hz, the duration is 8 s, and the sampling frequency is 256 Hz. The signal is constructed by MATLAB. As shown in Fig.3(a) and Fig.3(b), the accuracy of SST in frequency domain is much greater than that of STFT for the simulated EEG signal.

Fig.3(c) and Fig.3(d) compare the SST and STFT's processing results of actual C3 lead single trial EEG signal of one subject. It should be noted that the parameters of SST and STFT adopt default values except frequency. The two-dimensional array of SST processing results is 288×1 536 (frequency×time) and that of STFT is 33×47 (frequency×time). In the frequency and time domain, SST has higher resolution for data processing and can obtain more precise decomposition.

In general, due to the higher resolution and more refined frequency domain resolution characteristics of SST, the accuracy of data decomposition in frequency domain is improved, which is conducive to CNN's extraction of EEG frequency domain features. Therefore, SST is chosen as the means of signal frequency domain decomposition in this paper.

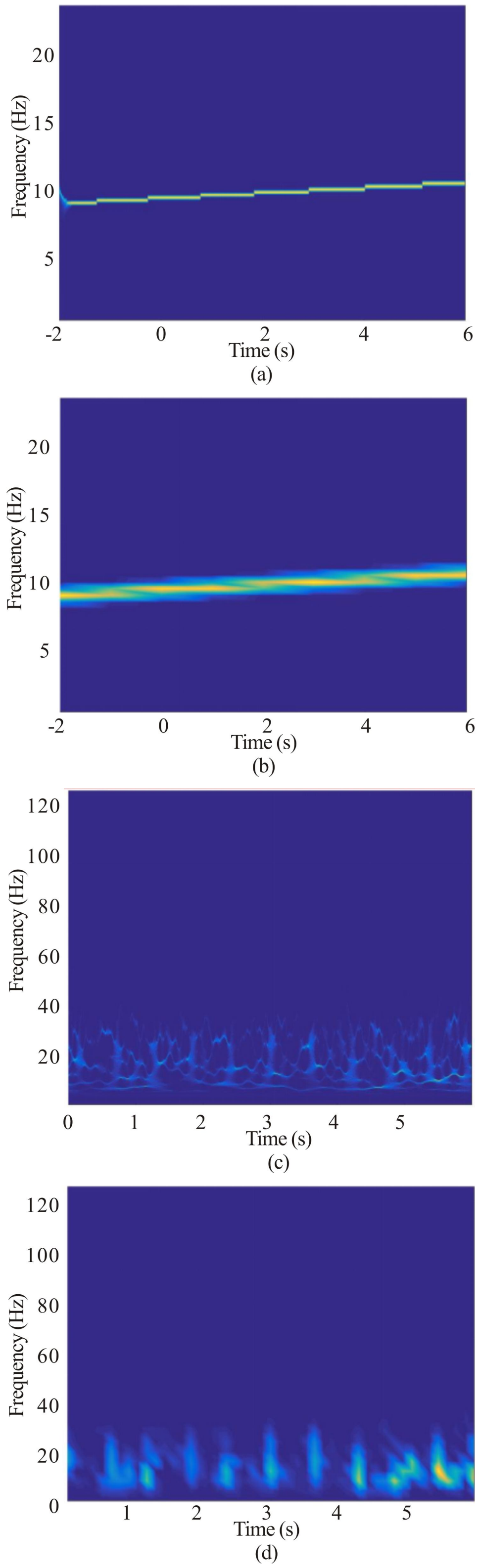

Fig.3 Results comparison: (a) SST's result diagram of the constructed signal; (b) STFT's result diagram of the constructed signal; (c) SST's result diagram of the EEG signal; (d) STFT's result diagram of the EEG signal

The LSTM unit regulates the information flow by introducing three gating units: forget gate, input gate and output gate as the internal mechanism[13,14]. The forget gate can determine what information should be discarded or retained in the information output produced by the preceding LSTM unit. The input gate is used to update the unit state, and the output gate determines the information input into the next LSTM unit[15-17].

The work flow of LSTM is described below. Firstly, the hidden state information $\boldsymbol{h}_{t-1}$ generated by the preceding LSTM unit and the currently input information $\boldsymbol{x}_t$ are input into the sigmoid function to give $\boldsymbol{O}_t$. Secondly, the current unit state $\boldsymbol{C}_t$ is input into the tanh function to give tanh ($\boldsymbol{C}_t$), then tanh ($\boldsymbol{C}_t$) and $\boldsymbol{O}_t$ multiply to produce the information $\boldsymbol{h}_t$ that the hidden state should carry, and finally the new unit state $\boldsymbol{C}_t$ and the new hidden state $\boldsymbol{h}_t$ are transmitted to the LSTM unit at the next moment. $\boldsymbol{O}_t$ is calculated as follows

$$\boldsymbol{O}_t = \mathrm{sigmoid}(\boldsymbol{W}_\mathrm{o} * [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_\mathrm{o}), \tag{1}$$

where $\boldsymbol{W}_\mathrm{o}$ and $\boldsymbol{b}_\mathrm{o}$ respectively represent the connection weight and bias vector related to the output gate[18].

The structure diagram of CNN composite LSTM is shown in Fig.4. The convolution layer uses one-dimensional convolution. The number of convolution cores is 40, the core length is 1×20, the stride is 4, and the defined input format is (n, 3, 1 536), where n is the number of trials, 3 is the number of characteristic leads, and 1 536 is the number of sampling points of a single trial, which is determined by $fs \times t$ (256×6). The rectified linear unit (Relu) activation function is adopted. A dropout layer is added after the convolution layer to prevent overfitting of data. The extracted data features are standardized, specifically, the sample mean is calculated first, then the variance is computed, and finally the data are standardized. The standardization formula of the data in this paper is as follows

$$\boldsymbol{y}_\mathrm{i} \leftarrow \gamma \hat{\boldsymbol{x}}_\mathrm{i} + \boldsymbol{\beta} \equiv BN_{\gamma,\boldsymbol{\beta}(\boldsymbol{x}_\mathrm{i})}, \tag{2}$$

where $\boldsymbol{x}_\mathrm{i}$ is input, and $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are the parameters obtained by network learning[19].

The kernel length used in the pooling layer is 1×4 and the stride is set at 4, so as to reduce the number of training parameters, reduce the dimension of the feature vector produced by the convolution layer, and lessen overfitting. Then, the results generated from the features and the preceding data in current order are input into the LSTM model. The flattening layer serves to convert multidimensional data into one-dimensional output. The flattening layer is situated between convolution layer and fully-connected layer to allow a smooth transition so that the size of batch is not affected.

The fully-connected layer uses softmax function to conduct two-category classification[20].

As the experiment was repeated in two sets, the experimental data of both sets were spliced to obtain a total of 60 trials data samples. The data samples were preprocessed to obtain EEG data of two types of tasks, one being upper limb movement and the other being lower limb movement. 80% of the offline collected data was established as the training set and the other 20% as the test set. In order to expand the training samples, 5-fold cross validation was used to compare the results of pure CNN, the combination of SST and CNN (SST + CNN), and CNN merged with SST and LSTM (SST + LSTM + CNN)[21,22].

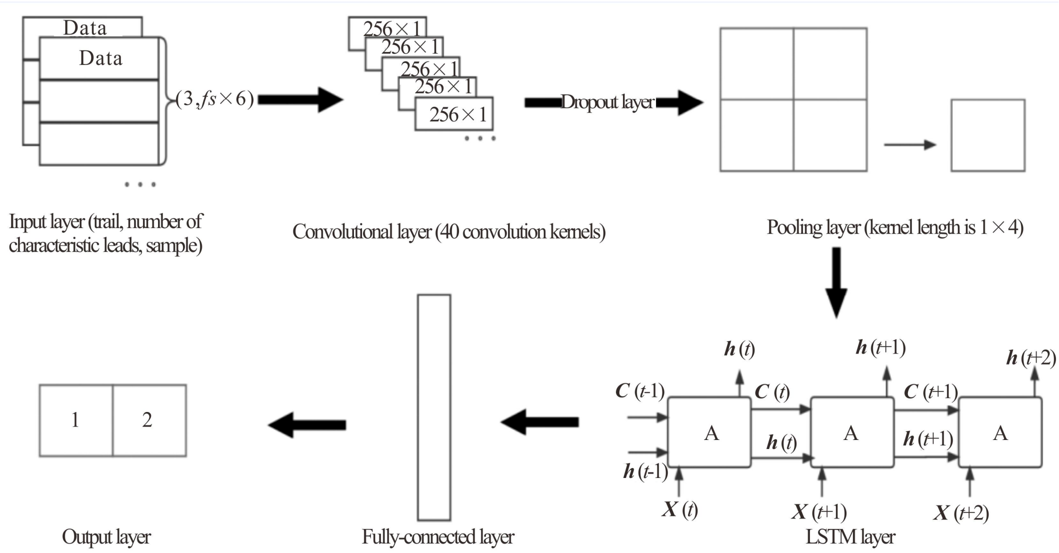

Fig.4 Structure diagram of the CNN composite LSTM

EEG data are processed through the forward and inverse transformation of SST, which can ensure that the data structure of CNN and SST + CNN is the same before being sent to CNN. After that, the data label is extracted and converted into one hot-coding. The data are changed to $N\times D\times P$ format, where N is the number of trials, D is the number of leads, and P is the number of sampling points of a single trial. All the data are enlarged by 106 times to avoid overfitting due to too weak data features. Then the data are sent into the neural network for training. CNN training includes two stages. First, the signal is transmitted through some neural network layers, then the forward propagation of the output is obtained in the output layer. Finally, the back propagation is used to calculate the error between the actual output and the expected output, and then the gradient is continuously renewed to train toward smaller gradient to gradually reduce the error, using cross entropy as the loss function. Finally, the function of TensorFlow is adopted to store the offline data training model into the local for the next experiment. The results of the 10 subjects (S1—S10) are shown in Tab.1.

Tab.1 Classification results about 10 subjects

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
CNN	73.3%	75.6%	80.7%	70.2%	76.7%	83.3%	73.3%	74.7%	71.6%	78.3%
SST+CNN	81.7%	85.3%	83.3%	75.4%	80.2%	78.3%	71.6%	76.2%	73.3%	81.7%
SST+LSTM+CNN	87.6%	85.7%	86.7%	78.3%	83.3%	84.3%	85.7%	75.8%	75%	86.7%

Except for subjects S6 and S7, the accuracy of the method of combining SST with CNN is higher than that of pure CNN, indicating that SST has a positive impact on improving the accuracy of the network. And except for S8, the accuracy of CNN incorporating SST and LSTM is higher than that of the other two methods, indicating that this method has the best effect.

Tab.2 shows a comparison among the accuracy of the experimental results analyzed by the three algorithms about the 10 subjects. From the data in the table, it can be seen that the accuracy of each subject analyzed by SST + LSTM + CNN algorithm is higher than that of the other two, implying that this method has evident advantages over the traditional methods.

In order to verify the repeatability of this method, the authors used the data of the public datasets "BCI Competition IV Datasets 2a"[23] with the 9 subjects to verify the algorithm. The comparison method is consistent with Tab.2. The results are shown in Tab.3.

Tab.2 Accuracy comparison of the three algorithms about 10 subjects

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
WPD+SVM	70.9%	73.4%	81.4%	69.7%	74.7%	82.1%	71.3%	72.5%	71.7%	77.1%
CSP+SVM	72.3%	74.7%	80.6%	71.1%	75.6%	78.9%	70.7%	73.9%	72.1%	77.4%
SST+LSTM+CNN	87.6%	85.7%	86.7%	78.3%	83.3%	84.3%	85.7%	75.8%	75%	86.7%

Tab.3 Algorithm verification in BCI Competition IV Datasets 2a

	S1	S2	S3	S4	S5	S6	S7	S8	S9	Average
WPD+SVM	83.5%	62.6%	84.5%	68.8%	57.8%	48.5%	84.9%	84.1%	77.9%	72.5%
CSP+SVM	88.6%	65.2%	81.4%	73.9%	71.4%	56.6%	86.4%	83.8%	87.2%	77.2%
SST+LSTM+CNN	98.1%	98.6%	95.5%	96.7%	87.5%	92.1%	93.9%	93.1%	98.8%	94.9%

From the data in Tab.3, it can be seen that the accuracy of each subject analyzed by SST + LSTM + CNN algorithm is also higher than that of the other two, which means that the method in this paper is repeatable.

Besides the comparison with traditional methods, another comparison is made between this SST + LSTM + CNN algorithm and a similar novel method, CNN combined with discrete wavelet transform (DWT) and power spectral density (PSD)[24]. DWT is used for energy analysis in order to obtain the data required by PSD. Without automatic band selection, the average accuracy of DWT + PSD + CNN is 94.0% as stated in the relevant paper, while this SST + LSTM + CNN algorithm gives an average accuracy of 94.9% when processing the same public datasets (BCI Competition IV Datasets 2a). It suggests that the effect of SST feature extraction is better than that of PSD. In future research, we will explore this related method.

From the perspective of feature extraction, compared with WPD, SST has higher resolution and is more conducive to feature extraction. Similarly, compared with CSP, SST is not easily disturbed by noise and only needs single channel data, which means it has higher operational efficiency and is convenient for practical application. From the perspective of classification method, compared with SVM, LSTM + CNN has better fuzzy information processing ability and resolution, which can better improve the classification accuracy. To sum up, the CNN algorithm combined with SST and LSTM has the best effect in processing EEG data in O-MI paradigm among the three methods, and can better identify motor intention.

In this paper, SST + LSTM + CNN algorithm is proposed to overcome the shortcomings of traditional CNN algorithm in EEG signal data processing. The improvement mainly lies in the following two aspects. The original data are no longer sent directly to CNN for feature extraction. SST is used for more accurate data decomposition according to frequency domain characteristics. The LSTM network, which excels in modeling time series information, is used to optimize the classification results. In order to study the performance of SST + LSTM + CNN algorithm, the EEG signals of 10 subjects conducting two tasks of motor imagery in O-MI paradigm were collected. SST + LSTM + CNN algorithm is stronger in electrophysiological signal recognition than pure CNN and SST + CNN, and it also significantly surpasses traditional algorithms (WPD + SVM and CSP + SVM). The limitation of this paper is that this research is still at an exploratory stage, hence the number of subjects is quite limited and no participants with upper or lower limb impairments were involved. In future research, the data size used for verification will be expanded, and the potential effect of this therapy on stroke patients' limb function rehabilitation with also be studied in depth. Due to the complexity of electrophysiological signals, for future study it's necessary to extract more effective MI related features from multiple electrophysiological signals in motor imagery recognition algorithms.

Statements and Declarations

The authors declare that there are no conflicts of interest related to this article.

References

[1] MANE R, CHOUHAN T, GUAN C. BCI for stroke rehabilitation：motor and beyond[J]. Journal of neural engineering, 2020, 17(4)：041001.

[2] KANEKO N, SASAKI A, YOKOYAMA H, et al. Effects of action observation and motor imagery of walking on the corticospinal and spinal motoneuron excitability and motor imagery ability in healthy participants[J]. Plos one, 2022, 17(4)：e0266000.

[3] STOLBKOV Y K, GERASIMENKO Y P. Cognitive motor rehabilitation：imagination and observation of motor actions[J]. Human physiology, 2021, 47(1)：104-112.

[4] STOLBKOV Y K, GERASIMENKO Y P. Observation of motor actions as a tool for motor rehabilitation[J]. Neuroscience and behavioral physiology, 2021, 51(7)：1018-1026.

[5] TORRISI M, MAGGIO M G, COLA M, et al. Beyond motor recovery after stroke：the role of hand robotic rehabilitation plus virtual reality in improving cognitive function[J]. Journal of clinical neuroscience, 2021, 92(9859)：11-16.

[6] DAUBECHIES I, LU J, WU H T. Synchrosqueezed wavelet transforms：an empirical mode decomposition-like tool[J]. Applied and computational harmonic analysis, 2011, 30(2)：243-261.

[7] MANDHOUJ B, CHERNI M A, SAYADI M. An automated classification of EEG signals based on spectrogram and CNN for epilepsy diagnosis[J]. Analog integrated circuits and signal processing, 2021, 108(1)：101-110.

[8] ONO Y, WADA K, KURATA M, et al. Enhancement of motor-imagery ability via combined action observation and motor-imagery training with proprioceptive neurofeedback[J]. Neuropsychologia, 2018, 114(1)：134-142.

[9] XIE J, PENG M, LU J, et al. Enhancement of event-related desynchronization in motor imagery based on transcranial electrical stimulation[J]. Frontiers in human neuroscience, 2021, 15(1)：141-147.

[10] NAGAI H, TANAKA T. Action observation of own hand movement enhances event-related desynchronization[J]. IEEE transactions on neural systems and rehabilitation engineering, 2019, 27(7)：1407-1415.

[11] KINOSHITA T, FUJIWARA K, KANO M, et al. Sleep spindle detection using RUSBoost and synchrosqueezed wavelet transform[J]. IEEE transactions on neural systems and rehabilitation engineering, 2020, 28(2)：390-398.

[12] SHEYKHIVAND S, MOUSAVI Z, REZAII T Y, et al. Recognizing emotions evoked by music using CNN-LSTM networks on EEG signals[J]. IEEE access, 2020, 8(1)：139332-139345.

[13] SHOEIBI A, SADEGHI D, MORIDIAN P, et al. Automatic diagnosis of schizophrenia in EEG signals using CNN-LSTM models[J]. Frontiers in neuroinformatics, 2021, 15(1)：1-7.

[14] XU G, REN T, CHEN Y, et al. A one-dimensional CNN-LSTM model for epileptic seizure recognition using EEG signal analysis[J]. Frontiers in neuroscience, 2020, 14(1)：1253-1259.

[15] SIMAR C, PETIEAU M, CEBOLLA A, et al. EEG-based brain-computer interface for alpha speed control of a small robot using the MUSE headband[C]//2020 International Joint Conference on Neural Networks, July 19-24, 2020, Glasgow, UK. New York：IEEE, 2020：1-4.

[16] DAI G, ZHOU J, HUANG J, et al. HS-CNN：a CNN with hybrid convolution scale for EEG motor imagery classification[J]. Journal of neural engineering, 2020, 17(1)：016025.1-016025.11.

[17] HWAIDI J F, CHEN T M. Classification of motor imagery EEG signals based on deep autoencoder and convolutional neural network approach[J]. IEEE access, 2022, 10(1)：48071-48081.

[18] KHADEMI Z, EBRAHIMI F, KORDY H M. A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals[J]. Computers in biology and medicine, 2022, 143(1)：105288.

[19] AMIN S U, ALSULAIMAN M, MUHAMMAD G, et al. Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion[J]. Future generation computer systems, 2019, 101(1)：542-554.

[20] SINGHAL V, MATHEW J, BEHERA R K. Detection of alcoholism using EEG signals and a CNN-LSTM-ATTN network[J]. Computers in biology and medicine, 2021, 138(1)：104940.

[21] LI H, DING M, ZHANG R, et al. Motor imagery EEG classification algorithm based on CNN-LSTM feature fusion network[J]. Biomedical signal processing and control, 2022, 72(1)：103342.

[22] JEONG J H, SHIM K H, KIM D J, et al. Brain-controlled robotic arm system based on multi-directional CNN-BiLSTM network using EEG signals[J]. IEEE transactions on neural systems and rehabilitation engineering, 2020, 28(5)：1226-1238.

[23] KAI K A, ZHANG Y C, ZHANG H, et al. Filter bank common spatial pattern (FBCSP) in brain-computer interface[C]//2008 IEEE International Joint Conference on Neural Networks, June 1-8, 2008, Hong Kong, China. New York：IEEE, 2008：2390-2397.

[24] MA X, WANG D, LIU D, et al. DWT and CNN based multi-class motor imagery electroencephalographic signal recognition[J]. Journal of neural engineering, 2020, 17(1)：016073.