

Prediction of Epileptic EEG Signal Based on SECNN-LSTM

Jian Qiang Wang¹, Wei Fang^{1,2,*} and Victor S. Sheng³

¹School of Computer and Software, Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, 210044, China

²State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, 100081, China

³Department of Computer, Texas Tech University, Lubbock, TX 79409, USA

*Corresponding Author: Wei Fang. Email: Fangwei@nuist.edu.cn

Received: 10 January 2022; Accepted: 24 March 2022

Abstract: Brain-Computer Interface (BCI) technology is a way for humans to explore the mysteries of the brain and has applications in many areas of real life. People use this technology to capture brain waves and analyze the electroencephalograph (EEG) signal for feature extraction. Take the medical field as an example, epilepsy disease is threatening human health every moment. We propose a convolutional neural network SECNN-LSTM framework based on the attention mechanism can automatically perform feature extraction and analysis on the collected EEG signals of patients to complete the prediction of epilepsy diseases, overcoming the problem that the disease requires long time EEG monitoring and analysis by manual, which is a large workload and relatively subjective, and improving the prediction accuracy of epilepsy diseases by adding the attention mechanism module. Through experimental tests, the algorithm of SECNN-LSTM can effectively predict the EEG signal of epilepsy disease, and the correct recognition rate is improved. The experiment has some reference value for the subsequent research of EEG signals in other fields in deep learning.

Keywords: EEG signal; SECNN-LSTM; feature analysis; epilepsy

1 Introduction

At the current level of research, brain-computer interface technology uses certain patterns of EEG signals generated by the brain for analysis. Using this phenomenon, we can interpret and process the signals through certain algorithms that can eventually achieve recognition and classification. The application of brain-computer interface technology in all aspects of society will advance the development of society with far-reaching implications. Theoretically, any EEG signal can be used to control a BCI system [1]. In medicine, BCI can assist in the treatment of some diseases, like mental disorders, Parkinson's, epilepsy, stroke, spinal cord injury, and other diseases, as well as provide help and reference for doctors' decision making by observing the patient's brain wave signals [1,2]. In daily life, many scholars have made achievements in BCI technology. For example, some scholars have used BCI technology to design intelligent wheelchairs, which facilitate the travel problems of disabled



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

people [3]; some scholars have used BCI technology to achieve wheelchair steering [4]; some scholars have used BCI technology to achieve EEG feature recognition of emotions [5]. Some scholars achieve manipulation of robot arm by BCI technology [6]. Many scholars in the world are still continuing to research BCI technology, a complete BCI structural system [7], as shown in Fig. 1.

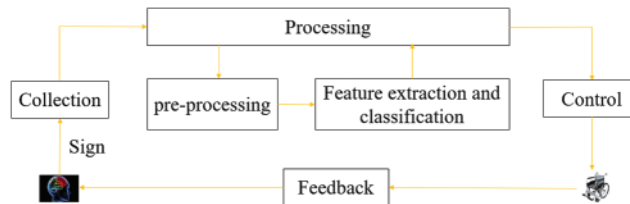


Figure 1: BCI structural system

Among them, scholars have put most of their efforts into the signal processing stage. In the preprocessing stage, EEG signals are usually preprocessed using various filters [8,9]. The feature extraction and classification stage mainly utilizes classifiers, and the main classifiers are classified as: linear classifiers, neural networks, nonlinear Bayesian classifiers, nearest neighbor classifiers, classifier combinations, and adaptive classifiers [9,10]. With the development of the society, more and more people are focusing on a new research area: deep learning. Deep learning is also an effective problem solving method in the BCI signal processing stage. Deep learning methods have been applied to classify emotions in brain waves. Compared to traditional machine learning with k-nearest neighbor (KNN) classifiers and support vector machine (SVM), deep learning significantly outperforms other methods in practical applications. These examples offer a variety of possibilities for the application of brain-computer interface technology in various fields [11].

We tried to use deep learning algorithms to process EEG signals, and CNN, long and short term memory (LSTM) networks [12] have been used extensively. Privacy protection has applied a hybrid deep learning framework [13], CNN networks can be implemented to classify and detect moths [14] and also to classify and detect breast cancer [15], a dual LSTM model based on attention mechanism is used to classify depression [16], and a CNN-LSTM model has been applied to motor imagery EEG signals [17], while the model has also been used to predict PM2.5 concentration [18]. And we designed a CNN-LSTM model for epileptic EEG signals based on this model, however, we found that the results were not satisfactory enough to test whether the model is applicable to epileptic class of EEG signals and to test whether there is room for improvement. We analyzed that the model has the problem of channel weights in the convolution process, which can lead to inaccurate feature extraction, so we changed the current model and also added the attention mechanism network to form a new model SECNN-LSTM for automatic feature extraction and classification. In particular, the attention mechanism module is a channel attention mechanism module designed based on Squeeze-and-Excitation (SE) [19] and embedded into the existing network architecture.

For our experiments, we chose public EEG datasets for experimental analysis. Firstly, Our aims are to add the attention mechanism module to the mixed model to verify the feasibility of applying the mixed model network to EEG signals. Secondly, to verify that the effect of adding the attention mechanism module to the mixed model is better than the former one, and to achieve the purpose of improving the effect. Finally, to provide reference basis and ideas for subsequent model innovation and application in signal processing of other EEG signals.

2 Related Work

2.1 CNN-LSTM Deep Learning Classifier

Our model is inspired by the CNN-LSTM deep learning classifier. This classifier was applied by the authors to low-invasive and low-cost BCI headband for motion image EEG detection [17]. The principle of the model is to use CNN instead of RNN encoder to extract represent image features using RNN decoder [20]. Based on the existence of multiple paradigms for the implementation of brain-computer interface systems, our experimental paradigms can be divided into several aspects based on evoking different EEG signals: motor imagery, P300, steady-state visual evoked potentials (SSVEP), auditory, emotional and other paradigms [21]. The emergence of deep learning is very effective in solving the problems that exist in the process of feature extraction and classification [22–25]. However, different paradigms analyze different EEG signals and the model structure has to be changed in relation to the specific EEG signal.

2.2 Determine the Mechanism for Adding Attention

BCI technology is a multidisciplinary technology and there are many scholars from different disciplines studying it, but the starting point of the research is different. In the signal processing stage, classification algorithms are the most important research again. Nowadays, the more popular classification algorithms based on EEG signals are mainly divided into adaptive classifiers, matrix and tensor classifiers, migration learning and deep learning. As for deep learning methods, the current results are still unsatisfactory [9]. However, with the development of technology, the application of deep learning on BCI is becoming more and more widespread. Deep learning classification of EEG signals using convolutional neural networks to detect epilepsy [26]. Motor imagery task classification using convolutional neural networks [27]. Automatic detection of Parkinson's disease using convolutional neural networks [28]. Motor image EEG detection for BCI headband using CNN-LSTM deep learning classifier [17].

However, the above work does not take into account the problem of classification accuracy due to the channel weighting problem. During the convolution process, there is indeed a problem of feature map channel weight consistency, such that the subsequent convolution process will bring about a loss of accuracy. In contrast, our model introduces the attention mechanism module, which is able to reassign weights to each feature map during the convolution process, providing some help for feature extraction in the subsequent convolution.

2.3 Channel Attention Mechanism

Attention is currently used in a wide range of applications. A solid step has been taken in natural language processing (NLP), image description, machine translation, convolutional neural network aspects [29–33], and these approaches have increased the diversity of predictions with good results. However, different attention mechanisms are applied for scenarios that do not use them. In context, this paper focuses on the channel attention mechanism. More specifically, this work attempts to reassign weights to each feature map and make its desired results more predictable by proposing to add a channel attention mechanism after feature extraction.

3 Methods

Currently, neural network-based research in EEG signal recognition is insufficient, and the application of deep learning is to be improved due to the specificity of EEG datasets. In this section, we propose the mixed SECNN-LSTM model, which is based on modifying the original network by

adding the attention mechanism module. In this experiment, our experimental results using the mixed model of CNN-LSTM have the problems of low accuracy, and low stability and long training time. From the experimental data structure, we analyze that the model has a channel weight assignment problem, so the model adds the attention mechanism module to solve the channel weight problem. We explain the model in two parts, first, an introduction to the model we built; second, an explanation of the attention mechanism module we introduced. The overall structure of the epilepsy prediction model we designed is shown in Fig. 2.

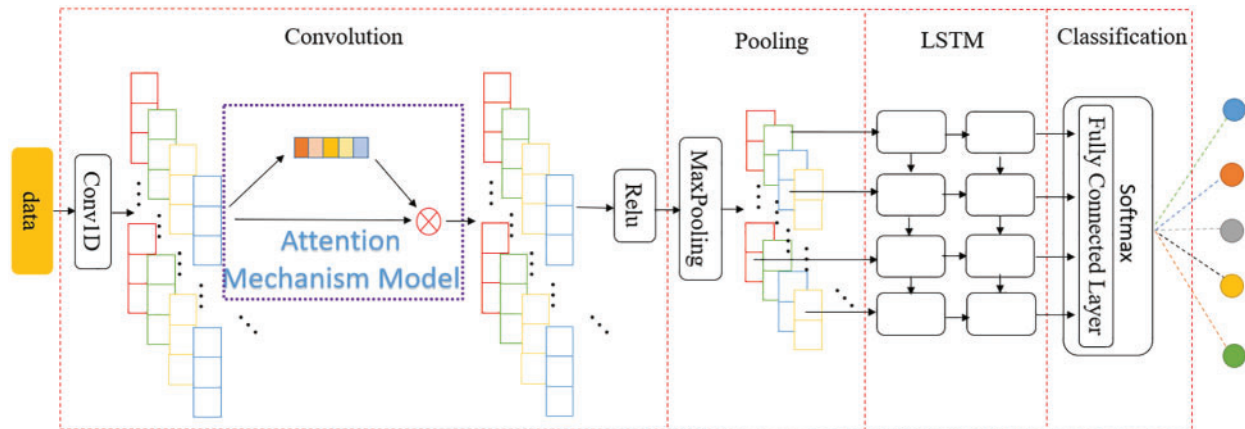


Figure 2: Model structure diagram

3.1 The Convolution Module

CNN-LSTM has many applications, but the structure has to be redesigned with specific scenarios and data to achieve good results, and our model is mainly designed in the convolution module. The diagram of SECNN-LSTM model incorporating attention mechanism is shown in Fig. 3.

In our designed convolutional module, the attention mechanism module is placed in the middle of two convolutional kernels, three modules form a part of the convolutional module, and four partial convolutional modules form a complete convolutional module. Analyzing from the data side, due to the special characteristics of the brain structure and the limitation of the acquisition equipment, the collected data need to be pre-processed, and the data quality will be affected by the influence from the human body and the surrounding environment. The preprocessed data is different from image data, which uses two-dimensional convolution to extract features [34]. We can observe that the analysis of processed brainwave epilepsy data resembles a text sequence, a time series. Therefore, it was decided to use one-dimensional convolution for this experiment. After the one-dimensional convolution, it needs to go through normalization with the activation function. The purpose of using normalization is to make the input of each layer of the neural network keep the same distribution during the training of the deep neural network, and the activation function is used to increase the nonlinear expressiveness of the network. At this moment, the expressiveness of the neural network is greatly improved, but there are some defects.

The output of our neural network finally has to realize the classification task, and the brainwave data will produce the corresponding number of feature maps after one-dimensional convolution. In the traditional one-dimensional convolution process, each channel of the feature map is equally important by default, which will have a certain impact on the accuracy of the classification results. Therefore, our model adds a channel attention mechanism module here, which reassigns weights to each channel of

the feature map to solve the loss problem caused by the different importance of different channels of the feature map in the convolution process, and makes the expression of the neural network more accurate. The feature map is reassigned with weights and then convolved again with feature extraction. The neural network is improved in feature extraction after reassigning weights. Finally, the neural network again undergoes normalization, activation function, maximum pooling and dropout operations to complete the partial convolution module process.

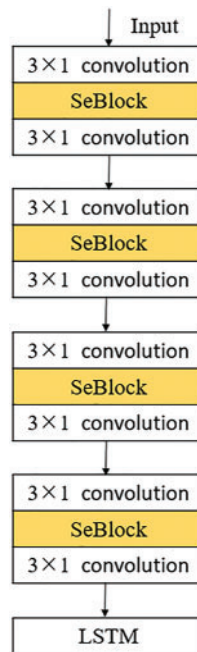


Figure 3: Graph of SECNN-LSTM model with fused attention mechanism

Our model partial convolution process is designed with 4 parts, each with the same structure, and the number of filters is doubled for each part from front to back. Because there will be a channel attention mechanism in each partial module, the more the later part, the better the accuracy of the convolution extraction will be compared to the former one. So the number of filters will be increased in the later part compared to the former.

3.2 Channel Attention Mechanism Module

The channel attention mechanism module added to the model we designed solves the problem that each channel has the same weight in the tradition by reassigning weights to the feature maps of each channel. The individual model diagram of the fused channel attention mechanism module is shown in [Fig. 4](#).

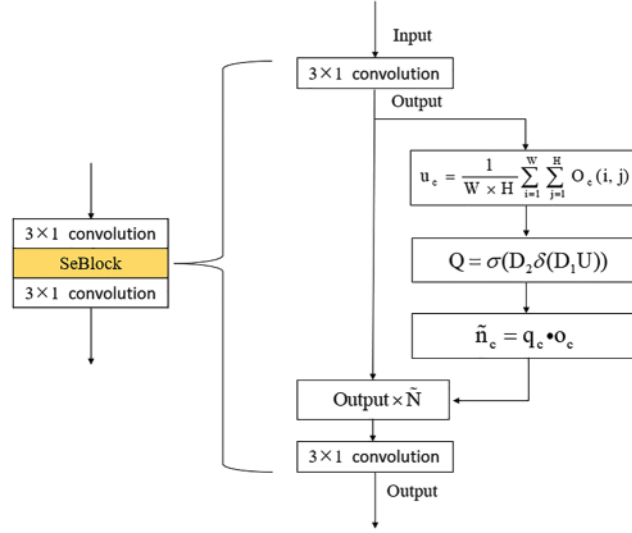


Figure 4: Single model diagram of the attention mechanism of the fusion channel

For convenience, we will denote Input as $I \in \mathbb{R}^{H' \times W' \times C'}$, Output is denoted as $O \in \mathbb{R}^{H \times W \times C}$, First, I is converted into the O by convolution, and we can also write the Output after convolution as $O = [O_1, O_2, \dots, O_C]$, At this time, by the formula:

$$O_c = f_c * I = \sum_{s=1}^{C'} f_c^s * i^s \quad (1)$$

Here $*$ represents the convolution, and the set of filter kernels is denoted by $F = [f_1, f_2, \dots, f_c]$, f_c denotes the parameter of the c -th filter, $f_c = [f_c^1, f_c^2, \dots, f_c^{C'}]$ and $I = [i^1, i^2, \dots, i^{C'}]$ are expressed in a simplified way, f_c^s represents a 1D convolution kernel, so it represents a f_c channel and acts on the channel corresponding to I .

The convolution is followed by two steps. First, Output waits along the straight line with reassigned weights to do the operation; second, the solution of the weights is performed along the branch. In the branching path, firstly, by the formula:

$$u_c = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H O_c(i, j) \quad (2)$$

Because each filter is a convolution operation performed in the local range, there is no way to combine each cell of O after convolution with information outside that region. To solve this problem, we compress the global information to the channel for representation, which is implemented in this paper using global average pooling (GAP), where $U \in \mathbb{R}^c$ and u_c is an element in U .

Then, we needed to make the neural network flexible and simple and able to learn important features ignoring the unimportant ones, so a gating mechanism with two fully connected layers was designed. The computational process at this point is as follows:

$$Q = \sigma(D_2 \delta(D_1 U)) \quad (3)$$

where δ represents the ReLU function, $D_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $D_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ represent the weight matrices of the two fully connected layers, respectively, r is the number of hidden nodes in the middle layer,

and σ represents the Sigmoid activation function. The two fully connected layers are used to reduce the complexity and generalization ability of the model. The first fully connected layer plays the role of dimensionality reduction, then the ReLU function is used to activate, the second fully connected layer plays the role of restoring the original dimensionality, and finally the Sigmoid function is used to obtain the vector Q of the gating unit.

Finally, the calculation process is as follows:

$$\tilde{n}_c = q_c \cdot o_c \quad (4)$$

The set of feature maps after reassigning the weights \tilde{N} , where $\tilde{N} = [\tilde{n}_1, \tilde{n}_2, \dots, \tilde{n}_c]$, q_c represent one of the scalar values of the gating unit Q . At this point, the reassignment of the weights of the feature map has been completed. The result at this point only needs to be combined with the Output of the first step again.

4 Experiments

In this section, we use the publicly available EEG epilepsy dataset to test the effectiveness of the model. Correctness, loss value, and gradient trend were used as evaluation criteria, using the adam optimizer [35] with the learning rate set to 0.001.

4.1 Epilepsy Dataset

Since brainwave data acquisition is demanding in terms of experimental conditions equipment, external environment, and people involved in the experiment, we used public datasets as experimental data to ensure data stability.

The experimental data were obtained from the University of Bonn Epilepsy EEG database. This dataset is a preprocessed and reconstructed version of a very commonly used dataset for seizure detection.

The original dataset consists of 5 different folders, each containing 100 files, each recording 23.6 s of brain activity. The corresponding time series were sampled to 4097 data points. Each data point is an EEG value recorded at a different time point, and we have a total of 500 individuals with 4097 data points each.

We split and shuffle every 4097 data points into 23 blocks, each block contains 178 data points for 1 s, and each data point is the EEG value recorded at different time points. The data has 11500 rows of information, each containing 178 data points, and the last column represents the label $y \in \{1, 2, 3, 4, 5\}$, as shown in Tab. 1. This version of the dataset was created to simplify access to the data by creating a .csv version.

Table 1: Description of data labels

y label	Status
1	Seizure
2	Region where tumor is located
3	Healthy brain region
4	Closed eyes
5	Open eyes

4.2 Experimental Comparisons

We conducted experiments using the processed epilepsy dataset and the results are shown in Figs. 5–8 and Tab. 2.

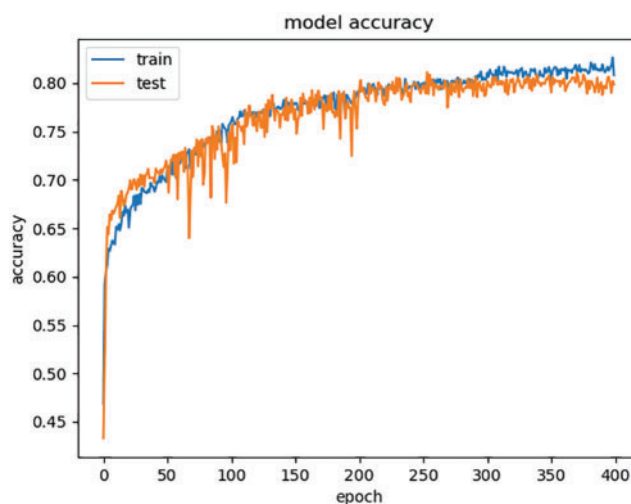


Figure 5: Accuracy of CNN-LSTM model

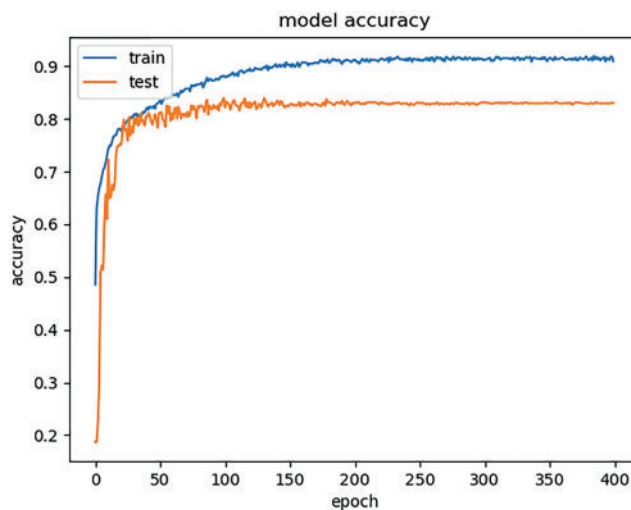


Figure 6: Accuracy of SECNN-LSTM model

As we can see in the above figure, the correctness aspect is improved by the model with the addition of the channel attention mechanism module, and the accuracy is better than the original model for both the experimental set and the data set. The experimental set improves the correct rate by about 8% in the improved model, and the test set improves by about 0.04% in the improved model. In terms of curve trends, after about 50 training sessions, the improved model has stabilized, the original model is still fluctuating, and the new model is able to reach stability in a much shorter time. From the loss value, the loss value is higher compared to the original model, but the stability is better and within the acceptable range.

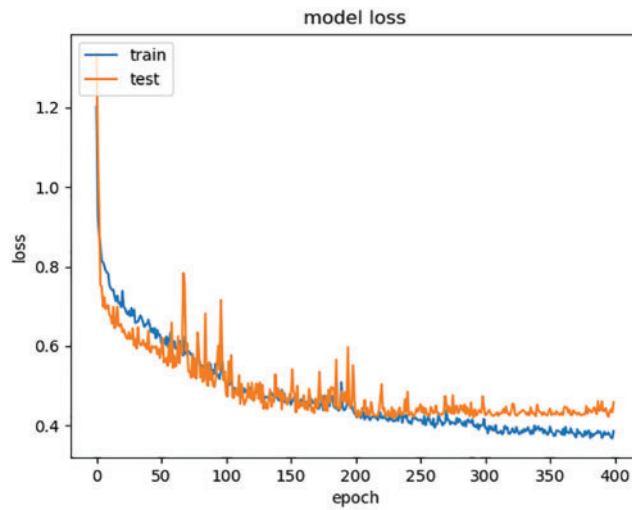


Figure 7: Loss value of CNN-LSTM model

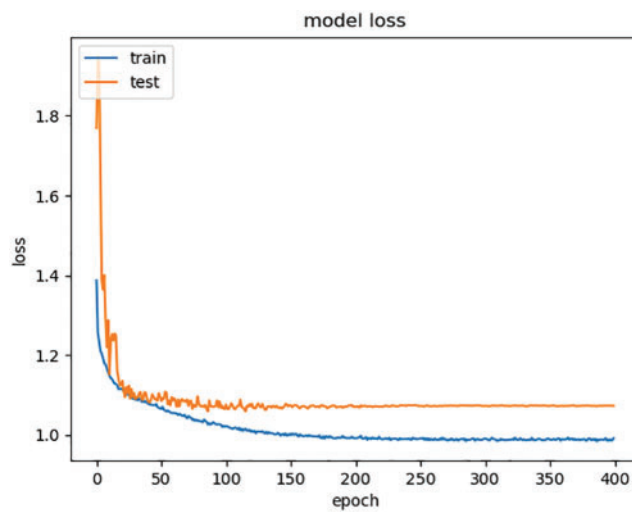


Figure 8: Loss value of SECNN-LSTM model

Table 2: Comparison of model test set training

Epoch	Loss		Accuracy	
	CNN-LSTM	SECNN-LSTM	CNN-LSTM	SECNN-LSTM
50	0.5697	1.0850	0.7200	0.8117
100	0.5362	1.0650	0.7304	0.8396
150	0.4789	1.0737	0.7626	0.8239
200	0.4663	1.0715	0.7817	0.8291
250	0.4439	1.0727	0.7943	0.8313

(Continued)

Table 2: Continued

Epoch	Loss		Accuracy	
	CNN-LSTM	SECNN-LSTM	CNN-LSTM	SECNN-LSTM
300	0.4456	1.0729	0.7887	0.8317
350	0.4369	1.0724	0.8000	0.8300
400	0.4584	1.0727	0.7987	0.8304

The specific accuracy values can be seen in [Tab. 2](#). The accuracy of the improved model is stable after 200 training sessions, and the final accuracy of the training reaches 0.8304%, and the accuracy of the original model finally reaches 0.7987%, which is more intuitive to reflect the improvement of the accuracy of the improved model. The loss value in the improved model has been relatively stable, and the original model has some minor fluctuations.

The experimental results show that the accuracy of the improved model is enhanced and the loss is in the acceptable range. The improved model can reach stability in a short time, which is better than the original model. The experiments also demonstrate the application of neural networks with added attention mechanism in the field of brainwave signal classification.

5 Conclusions

Our improved model SECNN-LSTM, based on CNN-LSTM with channel attention mechanism, has a strong classification capability, but there is some improvement in loss values, which is within the acceptable range. Our model brings improvement in accuracy and stability. Our experiments have made progress in terms of results and have made reference for applications on other EEG signals. There is still a lot of room for growth in deep learning applications in the BCI domain, and through our experiments, we demonstrate that adding attention mechanisms can be done in the BCI domain for the task of signal processing classification. In the future, we can continue to explore more deep learning frameworks combined with other attention mechanisms to accomplish more EEG signal classification.

Acknowledgement: This work was supported by the National Natural Science Foundation of China (Grant No. 42075007), and the Open Grants of the State Key Laboratory of Severe Weather (No. 2021LASW-B19).

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] K. Belwafi, F. Ghaffari, R. Djemal and O. Romain, "A hardware/software prototype of EEG-based BCI system for home device control," *Journal of Signal Processing Systems*, vol. 89, no. 2, pp. 263–279, 2017.
- [2] M. Pais-Vieira, A. P. Miguel, D. Moreira, D. Guggenmos, A. Santos *et al.*, "A closed loop brain-machine interface for epilepsy control using dorsal column electrical stimulation," *Scientific Reports*, vol. 6, no. 1, pp. 1–9, 2016.

- [3] B. Rebsemen, C. Guan, H. H. Zhang, C. C. Wang, C. Teo *et al.*, “A brain controlled wheelchair to navigate in familiar environments,” *IEEE Transactions on Neural System and Rehabilitation Engineering*, vol. 18, no. 6, pp. 590–598, 2010.
- [4] G. Bartels, L. C. Shi and B. L. Lu, “Automatic artifact removal from EEG-a mixed approach based on double blind source separation and support vector machine,” in *Proc. of Int. conf. of the IEEE Engineering in Medicine and Biology Society*, Buenos Aires, Argentina, IEEE, pp. 5383–5386, 2010.
- [5] G. Zouridakis, U. Patidar, N. S. Padhye, L. Pollonini, A. Passaro *et al.*, “Spectral power of brain activity associated with emotion-apilot MEG study,” in *Proc. of Int. Conf. on Biomagnetism Advances in Biomagnetism*, Dubrovnik, Springer, pp. 354–357, 2010.
- [6] A. P. Garcia, I. Schjolberg and S. Gale, “EEG control of an industrial robot manipulator,” in *IEEE 4th Int. Conf. on Cognitive Infocommunications (CogInfoCom)*, Budapest, Hungary, pp. 39–44, 2013.
- [7] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, J. R. Wolpaw *et al.*, “BCI2000: A general-purpose brain-computer interface (BCI) system,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.
- [8] B. Blankertz, R. Tomioka, S. Lemm *et al.*, “Optimizing spatial filters for robust EEG single-trial analysis,” *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2007.
- [9] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo *et al.*, “A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update,” *Journal of Neural Engineering*, vol. 15, no. 3, pp. 31005.1–31005.28, 2018.
- [10] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi *et al.*, “A review of classification algorithms for EEG-based brain-computer interfaces,” *Journal of Neural Engineering*, vol. 4, no. 2, pp. R1–R13, 2007.
- [11] J. Teo, L. H. Chew and J. T. Chia, “Classification of affective states via EEG and deep learning,” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 5, pp. 132–142, 2018.
- [12] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [13] S. Nithyanantham and G. Singaravel, “Hybrid deep learning framework for privacy preservation in geo-distributed data centre,” *Intelligent Automation & Soft Computing*, vol. 32, no. 3, pp. 1905–1919, 2022.
- [14] S. Lee, “A study on classification and detection of small moths using cnn model,” *Computers, Materials & Continua*, vol. 71, no. 1, pp. 1987–1998, 2022.
- [15] R. Rajakumari and L. Kalaivani, “Breast cancer detection and classification using deep cnn techniques,” *Intelligent Automation & Soft Computing*, vol. 32, no. 2, pp. 1089–1107, 2022.
- [16] A. M. Almars, “Attention-based bi- lstm model for arabic depression classification,” *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3091–3106, 2022.
- [17] F. M. Garcia-Moreno, M. Bermudez-Edo, M. J. Rodríguez-Fórtiz and J. L. Garrido “A CNN-LSTM deep learning classifier for motor imagery EEG detection using a low-invasive and low-cost BCI headband,” in *2020 16th Int. Conf. on Intelligent Environments (IE)*, IEEE, Madrid, Spain, pp. 84–91, 2020.
- [18] X. Shao, “Accurate multi-site daily-ahead multi-step pm2.5 concentrations forecasting using space-shared cnn-lstm,” *Computers, Materials & Continua*, vol. 70, no. 3, pp. 5143–5160, 2022.
- [19] J. Hu, L. Shen S. Albanie, G. Sun, E. Wu, “Squeeze-and-excitation networks,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.
- [20] O. Vinyals, A. Toshev, S. Bengio, D. Erhan *et al.*, “Show and tell: A neural image caption generator,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 3156–3164, 2015.
- [21] H. J. Hwang, S. Kim, S. Choi, C. Im *et al.*, “EEG-based brain-computer interfaces: A thorough literature survey,” *International Journal of Human-Computer Interaction*, vol. 29, no. 12, pp. 814–826, 2013.
- [22] M. Aldayel, M. Ykhlef, A. S. Lih *et al.*, “Deep learning for EEG-based preference classification in neuromarketing,” *Applied Sciences*, vol. 10, no. 4, pp. 1525, 2020.
- [23] Z. Mao, W. X. Yao, Y. Huang *et al.*, “EEG-based biometric identification with deep learning,” in *8th International IEEE/EMBS Conference on Neural Engineering (NER)*, vol. 609, Shanghai, China, pp. 609–612, 2017.

- [24] T. Wilaiprasitporn, A. Dittthapron, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul *et al.*, “Affective EEG-Based Person Identification Using the Deep Learning Approach,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 3, pp. 486–496, 2020.
- [25] P. P. Thoduparambil, A. Dominic, S. M. Varghese *et al.*, “EEG-based deep learning model for the automatic detection of clinical depression,” *Physical and Engineering Sciences in Medicine*, vol. 43, no. 4 pp. 1349–1360, 2020.
- [26] J. Wang G. Yu, L. Zhong, W. Chen, S. Sathish, “Classification of EEG signal using convolutional neural networks,” in *14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Xi’an China, pp. 1694–1698, 2019.
- [27] M. Parvan, A. R. Ghiasi, T. Y. Rezaii, A. Farzamnia *et al.*, “Transfer Learning based Motor Imagery Classification using Convolutional Neural Networks,” in *27th Iranian Conference on Electrical Engineering (ICEE)*, Yazd, Iran, pp. 1825–1828, 2019.
- [28] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvarajet, N. Arunkumar *et al.*, “A deep learning approach for Parkinson’s disease diagnosis from EEG signals,” *Neural Computing and Applications*, vol. 32, no. 15, pp. 10927–10933, 2020.
- [29] B. Zhang, D. Xiong, J. Su *et al.*, “Neural Machine Translation with Deep Attention,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 1, PMLR, pp. 154–163, 2020.
- [30] D. Hu, “An introductory survey on attention mechanisms in NLP problems,” in *Proceedings of SAI Intelligent Systems Conference*, Springer, London, United Kingdom, pp. 432–448, 2019.
- [31] H. Hu, Q. Li, Y. Zhao, and Y. Zhang, “Parallel Deep Learning Algorithms With Hybrid Attention Mechanism for Image Segmentation of Lung Tumors,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2880–2889, 2021.
- [32] K. Wang, J. He, L. Zhang *et al.*, “Attention-Based Convolutional Neural Network for Weakly Labeled Human Activities’ Recognition With Wearable Sensors,” in *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7598–7604, 2019.
- [33] H. Shi, J. Li and Y. Xu, “Double-Head Attention-Based Convolutional Neural Networks for Text Classification,” *International Joint Conference on Information, Media and Engineering (IJCIME)*, Osaka, Japan, pp. 27–31, 2019.
- [34] S. Hussain, M. Abualkibash, S. Tout *et al.*, “A Survey of Traffic Sign Recognition Systems Based on Convolutional Neural Networks,” in *IEEE International Conference on ElectrolInformation Technology (EIT)*, Rochester, MI, USA, pp. 570–573, 2018.
- [35] S. Y. ŞEN and N. ÖZKURT, “Convolutional Neural Network Hyperparameter Tuning with Adam Optimizer for ECG Classification,” *Innovations in Intelligent Systems and Applications Conference (ASYU)*, Istanbul, Turkey, pp. 1–6, 2020.