Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.38304

EEG Signal Classification for ADHD Based on Bidirectional GRU and Forward Channel Selection

Qingyi Lyu

College of Life Sciences and Technology, Xinjiang University, Urumqi, China
Lvqingyi@stu.xju.edu.cn

Abstract. Attention Deficit Hyperactivity Disorder (ADHD) is a common neurodevelopmental disorder in children, and many scholars have used Electroencephalography (EEG) to study it and achieved good results. However, current models still have the problems of too many channels and weak spatial dependence, which may increase computational complexity and affect practical applications. In this study, a new classification model is constructed, which can not only analyze the spatial relationship but also introduce the forward feature selection method. Using Pearson correlation and random forest, the independent diagnostic ability of all channels is analyzed, and the baseline model is constructed to extract temporal features. In the study, the forward feature selection strategy is used to screen a 10-channel subset from the original 19 channels. The experimental results show that the test accuracy of the 10 core electrodes is relatively high, which can significantly reduce the false positive rate in diagnosis and provide effective support for the screening of ADHD.

Keywords: Attention deficit hyperactivity disorder, EEG, Channel selection

1. Introduction

ADHD is a common neurodevelopmental disorder in children [1]. If no intervention is taken, the core symptoms of inattention, hyperactivity, and impulsivity will be carried forward to adulthood, which will have a negative impact on the child's life and study, and also have a negative impact on the child's social interaction [2]. Therefore, it is necessary to make a correct diagnosis as soon as possible. In the process of clinical diagnosis of this disease, mental health workers should make a subjective evaluation of children's behavior and make a comprehensive evaluation of children's behavior and interview results. This is the routine diagnostic approach, but it takes a long time, is easily disturbed by subjective factors, and the diagnosis is not accurate, which is easy to misdiagnose [3, 4], and the problem is more prominent in areas where medical resources are relatively scarce. Therefore, it is necessary to introduce advanced auxiliary diagnostic tools to improve the accuracy of diagnosis.

In order to meet the needs of more and more people, artificial intelligence technology has been used to identify ADHD. In the early stage of research, the algorithm of machine learning was introduced into the study of Electroencephalography (EEG) signals, including decision tree, support vector machine and random forest [5, 6]. The common feature of these methods is that they can

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.38304

extract features by hand, and describe the characteristics of signals from the time domain and frequency domain.These methods have achieved certain success, but the features obtained are too subjective and can not reflect the characteristics of the signal.

With the rapid development of deep learning, more and more scholars are willing to use end-to-end architecture in their research, which has strong automatic feature extraction ability. Kwak et al. constructed CNN model [7], which can extract spatial and temporal features and play an important role in EEG classification. Du et al. constructed ATDD-LSTM model [8], which can describe the non-linear relationship between multi-channel EEG signals, extract important information from multiple channels, and play a leading role in EEG recognition. Song et al. constructed DGCNN model [9], which can fully grasp the relationship between multiple channels of EEG and describe the topological relationship. Bi-GRU model has strong ability to extract spatial features and can extract long-term temporal features from sequential EEG signals, which can play an important role in temporal feature extraction. These deep learning models have improved the accuracy of ADHD diagnosis.

Although there are many innovations in the current research, there are still many challenges that need to be solved. First, most of the current deep learning models use the International 10-20 system to obtain EEG signals. Most of the methods only use the concatenation strategy to extract features, but do not consider the relationship between adjacent electrodes and the spatial structure of the cerebral cortex. Second, multi-channel EEG signals are redundant, some signals are not related to ADHD, which will increase the computational load and introduce noise, and it is difficult to achieve lightweight development of diagnostic equipment.

In order to make up for the shortcomings of the above research, this paper constructs a robust framework, which can integrate multidimensional evaluation and forward feature selection, and can also evaluate the diagnostic capabilities of various channels. This framework uses the random forest classifier to evaluate the diagnostic capabilities of various channels, and uses Bi-GRU to extract temporal features. The forward feature selection strategy is introduced to screen the optimal subset of channels. The model performance is evaluated by a variety of indicators, including test accuracy, recall, F1 score, specificity and so on. The experimental results show that the framework built in this paper can reduce the number of channels from 19 to 10 (T4, Cz, Fp2, F3, O1, O2, Fz, P4, C4, F7) , which can not only ensure the accuracy of diagnosis, but also provide a reliable solution for ADHD screening.

2. Methodology

2.1. ADHD EEG dataset and preprocessing

In this study, the experimental data set is obtained from IEEE DataPort and is a public benchmark [10]. There are 121 children in this data set, aged 7-12, including 60 healthy children and 61 ADHD patients.The data in this data set are collected by using 19 channels, in accordance with the 10-20 system, and the sampling frequency is 128 Hz. In order to input the data into the network, it is necessary to use the standard preprocessing pipeline, and the missing data and zero data are deleted, and the time series are unified to 128 time steps, so as to eliminate the impact of impedance differences.The data are divided into three parts: training set, test set and validation set, with a ratio of 8:1:1.

DOI: 10.54254/2755-2721/2026.38304

2.2. Bidirectional GRU classification network

In order to grasp the temporal characteristics of EEG signals, this paper introduces BiGRU model, which is a variant of RNN model [11]. The biggest advantage of this model is that it can deal with sequential data and avoid the problem of gradient disappearance, which has been widely used in the process of classifying EEG signals. In the process of classification, the model is obviously different from LSTM, with fewer gates. In this model, a gate can not only play the role of forgetting gate, but also play the role of memory gate, which can improve the computational efficiency.In the process of classification, the model can deal with large-scale data samples, and the bidirectional structure can provide information for the network in both forward and reverse directions, so that the model can describe the data more accurately, and the fitting performance is relatively ideal.This model can process the information of forward and reverse sequences at the same time, and its ability to extract time series features is obviously stronger than that of traditional RNN models [12].

The network structure constructed in this study has three layers: (1) Input layer: the shape of the tensor is (128, 19), which can accept 3D tensor; (2) Feature extraction layer: 128 hidden units are set in this layer, forward and reverse temporal outputs are combined, and the effectiveness of BiGRU is reflected; (3) Classification output layer: the Softmax function and full connection layer are used to make predictions.The model uses cross entropy loss function in the training process, the learning rate is, and the batch size is 16. The specific structure of the model is shown in Figure 1.

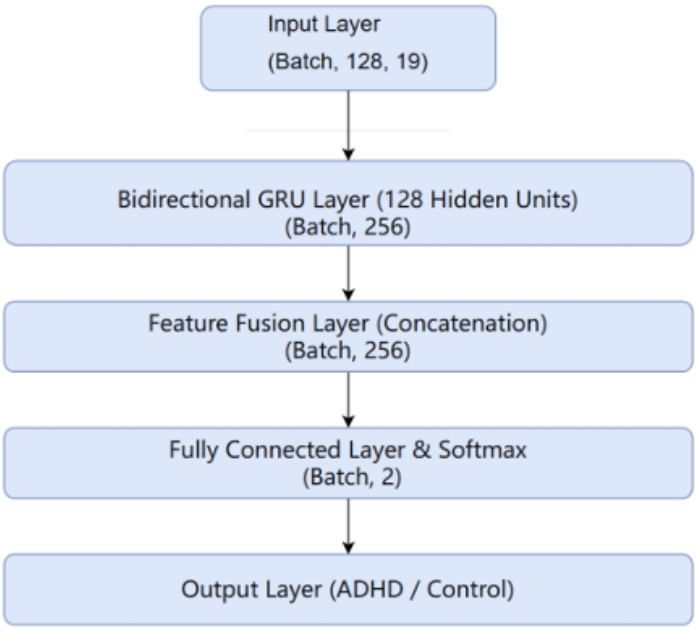

Figure 1. Structural diagram of the proposed Bidirectional GRU (Bi-GRU) classification model (picture credit: original)

2.3. Experimental design and channel evaluation strategy

In order to eliminate redundant features and find the most valuable electrodes, the three-stage channel screening method is introduced.

A 19-channel baseline experiment is carried out, and the Bi-GRU model is trained for 30 epochs, and the model with the highest test accuracy is saved as a checkpoint. The test accuracy, specificity,

DOI: 10.54254/2755-2721/2026.38304

F1-score, precision, and other indicators are calculated to obtain the confusion matrix.

In this stage, the correlation between the signals of the two channels is analyzed, and the Pearson coefficient is calculated to obtain the relationship between the signals of each channel and the diagnosis label, so as to grasp the linear dependency between the disease and the electrode.

In this study, the forward feature selection algorithm is used to select the most appropriate subset from the channel. The subset is empty at the beginning, and the candidate channel is added to the subset in the iterative process, so as to maximize the test accuracy of the model.The iteration is stopped when the model performance reaches saturation. In order to improve the reliability of the subset, the best 10-channel combination is selected and trained independently for 30 times, and the evaluation index is compared with the baseline model.

3. Results and discussion

3.1. Performance comparison between the full 19-channel and proposed 10-channel models

After training the selected 10 channels, the comparison between the two models is carried out in Table 1. It is worth noting that the baseline model has a high recall rate (99.40%) and can avoid missed diagnosis, but the specificity is only 82.43% and the precision is 87.69%.Compared with it, the 10-channel model can also achieve good results in terms of test accuracy (91.77%) and F1-score (92.67%), but the precision reaches 91.93% and the specificity reaches 89.68%.The specificity is continuously improved, which can effectively avoid false alarms for healthy people.

Table 1. Quantitative performance comparison between the baseline and proposed 10-channel model

Evaluation Metrics	Full 19-channel Baseline Model	Proposed Optimal 10-channel Model
Test Accuracy	91.89%	91.77%
Precision	87.69%	91.93%
Recall / Sensitivity	99.40%	93.43%
Specificity	82.43%	89.68%
F1-Score	93.18%	92.67%

3.2. Analysis of confusion matrices

In order to better explain the above conclusions, a confusion matrix is drawn. It can be seen from Figure 2 that the number of false positives of the baseline model is 418, which is obviously too high.In contrast, the 10-channel model can reduce the number of false positives to 232. Although the number of false negatives has increased from 11 to 186, the number of misdiagnosed healthy people has been significantly reduced, and the number of false positives has decreased significantly, which is enough to improve the specificity and accuracy, and the elimination of redundant features can make the 10-channel model have a strong ability to suppress false positives and maintain high sensitivity.

DOI: 10.54254/2755-2721/2026.38304

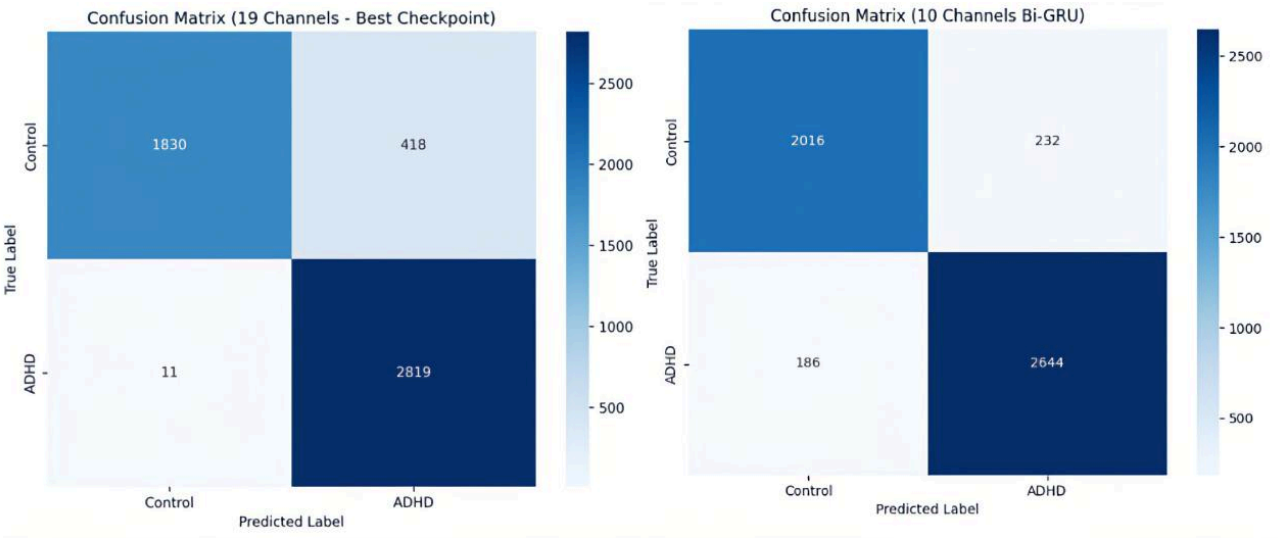

Figure 2. Comparison of confusion matrices between the full 19-channel baseline and the proposed 10-channel model (picture credit: original)

3.3. Analysis of channel correlation and independent evaluation

By analyzing the channel correlation matrix (Figure 3), it is known that the similarity between some adjacent electrodes is relatively high, and the correlation coefficient can reach 0.8, which means that the information of 19 channels has a high degree of redundancy.In addition, the evaluation of a single channel found that the classification accuracy of channels covering the midline frontal and occipital lobes is significantly higher than that of other brain regions, which is enough to show that the number of channels should be reduced.

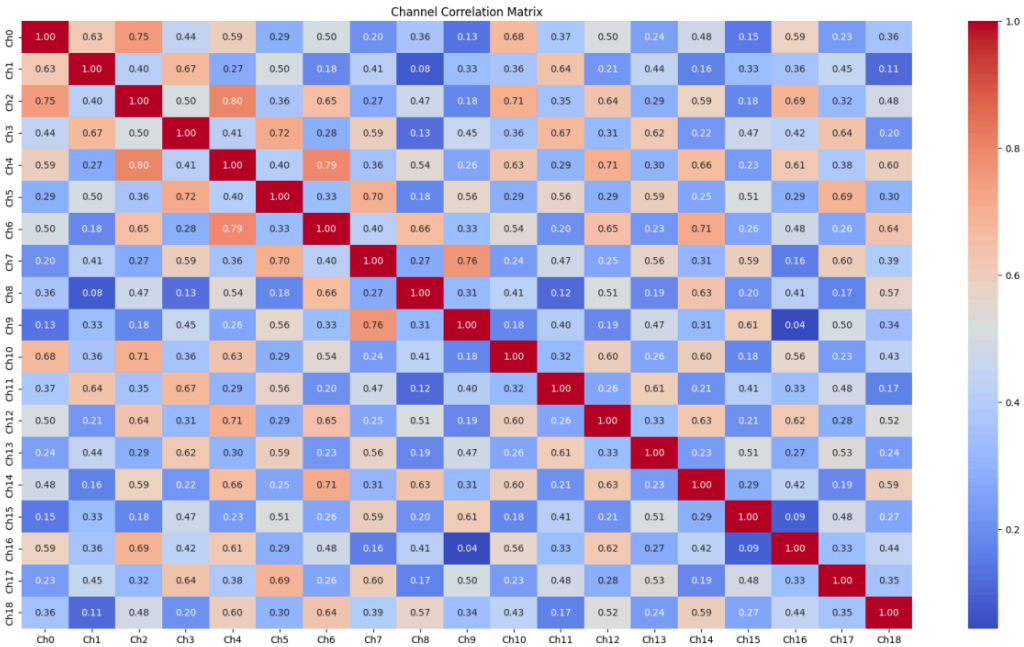

Figure 3. Channel correlation matrix of the 19 EEG Electrodes (picture credit: original)

DOI: 10.54254/2755-2721/2026.38304

3.4. Feature forward selection and neuroanatomical validation

After 19 rounds of forward iterative screening, the model performance reaches saturation when the number of selected channels is 10. The algorithm ultimately identifies the optimal 10-channel subset: T4, Cz, Fp2, F3, O1, O2, Fz, P4, C4, and F7. To verify the biological rationality of this selection, these electrodes are mapped to their respective brain regions. These ten channels cover the prefrontal lobe (Fp2, Fz) [13, 14], temporal lobe (T4) [15], central region (Cz, C4) [16], parietal lobe (P4) [17], occipital lobe (O1, O2) [18-20] and inferior frontal lobe (F3, F7) [21]. According to existing neuroanatomical literatures, these regions correspond closely to the executive control network, visual attention network, and sensorimotor network, which have been proven to be consistently abnormal in ADHD patients. This strongly validates the neuropathological rationality of the forward selection result.

4. Conclusions

In this study, the number of EEG electrodes is reduced from 19 to 10, and the forward channel selection strategy is used to identify the 10 core channels, which are consistent with the anatomical regions related to ADHD. The model with 10 channels has a robust accuracy of 91.77% and a high F1-score of 92.67%.Compared with the baseline model with 19 channels, the model with 10 channels has higher accuracy, specificity and other indicators, which can reduce computational complexity and reduce false positives. The method put forward in this study can provide a good reference for the development of home EEG screening equipment for ADHD.

In this study, there are two obvious shortcomings. First, the data set does not include age, gender and other demographic subgroups, which makes it difficult to carry out a more detailed population-based ablation analysis. Secondly, only Bi-GRU architecture is adopted, but CNN, Transformer and other advanced spatial-temporal networks are not introduced, and other channel selection algorithms are not compared in this study. In the next stage of research, Future work should introduce more data sets, use embedded equipment to carry out hardware deployment, and make the model more flexible.

References

[1] Polanczyk, G., & Rohde, L. A. (2007). Epidemiology of attention-deficit/hyperactivity disorder across the lifespan. *Current Opinion in Psychiatry, 20*, 386-392.

[2] Moore, S., Paalanen, L., Melymuk, L., et al. (2022). The association between ADHD and environmental chemicals—A scoping review. *International Journal of Environmental Research and Public Health, 19*, 2849.

[3] Michel, C. M., Murray, M. M., Lantz, G., et al. (2004). EEG source imaging. *Clinical Neurophysiology, 115*, 2195-2222.

[4] Parlatini, V., Bellato, A., Gabellone, A., et al. (2024). A state-of-the-art overview of candidate diagnostic biomarkers for attention-deficit/hyperactivity disorder (ADHD). *Expert Review of Molecular Diagnostics, 24*, 259-271.

[5] Gevins, A., Leong, H., Smith, M. E., et al. (1995). Mapping cognitive brain function with modern high-resolution electroencephalography. *Trends in Neurosciences, 18*, 429-436.

[6] Stokes, M. G., Wolff, M. J., & Spaak, E. (2015). Decoding rich spatial information with high temporal resolution. *Trends in Cognitive Sciences, 19*, 636-638.

[7] Kwak, N. S., Müller, K. R., & Lee, S. W. (2017). A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PLOS ONE, 12*, e0172578.

[8] Du, X. B., Ma, C. X., Zhang, G. H., et al. (2020). An efficient LSTM network for emotion recognition from multichannel EEG signals. *IEEE Transactions on Affective Computing, 11*, 1528–1540.

DOI: 10.54254/2755-2721/2026.38304

[9] Song, T. F., Zheng, W. M., Song, P., et al. (2020). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing, 11*, 532–541.
[10] Nasrabadi, A. M., Allahverdy, A., Samavati, M., & Mohammadi, M. R. (2020). EEG data for ADHD/control children. *IEEE DataPort*. https://doi.org/10.21227/rzfh-zn36
[11] Li, L., Yang, L., & Zeng, Y. (2021). Improving sentiment classification of restaurant reviews with attention-based Bi-GRU neural network. *Symmetry, 13*, 1517.
[12] Chung, J. Y., Gulcehre, C., Cho, K. H., et al. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv Preprint*, arXiv:1412.3555. https://arxiv.org/abs/1412.3555
[13] Burgess, G. C., Depue, B. E., Ruzic, L., Willcutt, E. G., Du, Y. P., & Banich, M. T. (2010). Attentional control activation relates to working memory in attention-deficit/hyperactivity disorder. *Biological Psychiatry, 67*, 632-640.
[14] Konrad, K., Neufang, S., Hanisch, C., Fink, G. R., & Herpertz-Dahlmann, B. (2006). Dysfunctional attentional networks in children with attention deficit/hyperactivity disorder: Evidence from an event-related functional magnetic resonance imaging study. *Biological Psychiatry, 59*, 643-651.
[15] Silk, T. J., Vance, A., Rinehart, N., et al. (2009). White-matter abnormalities in attention deficit hyperactivity disorder: A diffusion tensor imaging study. *Human Brain Mapping, 30*, 2757-2765.
[16] Aricò, M., Arigliani, E., Giannotti, F., et al. (2020). ADHD and ADHD-related neural networks in benign epilepsy with centrotemporal spikes: A systematic review. *Epilepsy & Behavior, 112*, 107448.
[17] Zhang, Y., Tang, A. C., & Zhou, X. (2014). Synchronized network activity as the origin of a P300 component in a facial attractiveness judgment task. *Psychophysiology, 51*, 285-289.
[18] Pitcher, D., Dilks, D. D., Saxe, R. R., et al. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage, 56*, 2356-2363.
[19] Braddick, O. (2015). Occipital lobe (visual cortex): Functional aspects. In *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., pp. 127–132). Elsevier, Oxford.
[20] Renier, L. A., Anurova, I., De Volder, A. G., et al. (2010). Preserved functional specialization for spatial processing in the middle occipital gyrus of the early blind. *Neuron, 68*, 138–148.
[21] Mattes, J. A. (1980). The role of frontal lobe dysfunction in childhood hyperkinesis. *Comprehensive Psychiatry, 21*, 358–369.