A 3D-ResNet Combined with BRNN: Application in the Auxiliary Diagnosis of ADHD

Chuanli Yang[a*], Huawang Qin[a], Siyuan Hu[b]

[a]Nanjing University of Information Science and Technology, Nanjing, 210044, China

[b]Northeast Forestry University, Harbin, 150006, China

*Correspondence: ycl254384@qq.com

Abstract. Attention Deficit/Hyperactivity Disorder (ADHD) is a common mental disorder that exhibits a high incidence rate in children and adolescents, and it is also observed in adults. Currently, there is a lack of objective diagnostic methods for ADHD. Therefore, a three-dimensional residual network (3D-ResNet) deep learning method based on feature extraction from rs-fMRI images for assisting in the diagnosis of ADHD based on resting-state functional magnetic resonance imaging (rs-fMRI) and deep learning models was proposed in this paper. Taking into consideration the temporal characteristics of rs-fMRI, we constructed a 3D-ResNet model based on four-dimensional image. The model utilized TimeDistributed to encapsulate residual blocks which allowed the model to extract spatial features from rs-fMRI while preserving its temporal sequence information. We constructed four different hierarchical structures of 3D-ResNet which are subsequently combined with two different bidirectional recurrent neural networks (BRNNs) to extract sequence features. And BRNNs includes bidirectional long short-term memory (Bi-LSTM) and bidirectional gated recurrent unit (Bi-GRU). The proposed method utilized the ADHD-200 Consortium's public dataset for training and was validated by 5-fold cross-validation. The experimental results indicated that the proposed method in this study demonstrated superior performance on the dataset compared to traditional methods (Accuracy: 76.56%, Sensitivity: 80.16%, Specificity: 90.22%). Therefore, adopting this method can further enhance the accuracy of assisting in the diagnosis of ADHD.

Keywords: rs-fMRI, ADHD, BRNNs, 3D-ResNet, diagnosis

1. Introduction

Attention Deficit/Hyperactivity Disorder (ADHD) is a highly prevalent neurodevelopmental disorder, the onset of the condition typically occurs before the age of 12, and it is characterized by persistent hyperactivity, excessive impulsivity, or inability to concentrate[1-3]. ADHD can be classified into three subtypes, namely Attention Deficit/Hyperactivity Disorder-Combined Type (ADHD-C), Hyperactive-Impulsive Type (ADHD-HI), and Inattentive Type (ADHD-I).

To mitigate the challenges associated with image fusion, this study constructed a deep learning model based on rs-fMRI image data. The paper introduced a network architecture named 3D-ResNet, which is employed to extract spatial features from rs-fMRI. Subsequently, it combines Bidirectional Recurrent Neural Networks (BRNNs) to extract temporal features. Traditional Recurrent Neural Networks (RNNs) face challenges such as gradient vanishing and exploding when dealing with sequential data, these problems make it difficult to capture long-range dependencies. BRNNs capture dependency relationships in sequential data by combining information from both the forward and backward directions. Unlike RNNs, which only consider past information[4-6], BRNNs simultaneously take into account both past and future information[7]. This ability helps the model to comprehensively understand the context within the sequence and consequently improve its classification accuracy. The paper combines two different types of BRNN—Bidirectional Long Short-Term Memory (Bi-LSTM) and Bidirectional Gated Recurrent Unit (Bi-GRU)—to find the optimal network composition, Bi-LSTM is an extended form of Long Short-Term Memory (LSTM), and Bi-GRU is a bidirectional recurrent neural network based on Gated Recurrent Unit (GRU). Unlike LSTM and GRU, Bi-LSTM and Bi-GRU can simultaneously consider past and future information at each time step, and that enables better capture of long-term dependencies in time series. Compared to traditional methods, the main contributions of our method are as follows:

Using scaled rs-fMRI image data as input avoids the cumbersome preprocessing steps associated with multimodal image fusion, and it reduces the need for extensive feature engineering.

A deep learning model based on four-dimensional image data was innovatively constructed. This model focuses on extracting spatial features while preserving fMRI time series, and that enhances the correlation of features in both spatial and temporal dimensions[8].

Four different structures of 3D-ResNet were designed, and they were combined with two types of BRNNs. Through ablation experiments, the optimal model combination was identified[9].

2. Experiment

This paper trains on rs-fMRI data that has been preprocessed using the Athena pipeline. The samples are sourced from the ADHD-200 Global Competition dataset. The Athena pipeline provides information on 973 preprocessed subjects, including rs-fMRI scans, T1-weighted structural scans, and preprocessed script files. The preprocessing steps primarily include operations such as slice timing correction, head motion correction, smoothing, and filtering[10]. To mitigate the impact of age differences and the imbalanced distribution of positive and negative samples on model training, after stage exclusion, the remaining rs-fMRI data from 430 subjects is used as input. ADHD subtypes are ignored, and all subclasses are labeled as 1. The average age of participants is 12.62, with an equal proportion of ADHD to Typically Developing Control (TDC) subjects at a ratio of 1:1. For a detailed composition of the dataset, refer to Table 1.

Table 1: The detailed composition of multi-site samples.

	Pittsburgh	Peking	Total
ADHD	0	78	215
TDC	49	88	215
Total	49	166	430

3. Results

This section primarily presents the experimental results of training four different residual networks combined with two types of BRNNs, and compares them with existing models. In certain research reports focusing on classification tasks, the majority often rely solely on the accuracy metric to assess their methods. However, this alone is inadequate to substantiate the feasibility of their approaches, that's because high classification accuracy may be a result of imbalanced sample distribution, and it will lead the model to exhibit bias towards predicting a specific class in extreme cases. For instance, in a binary classification task where positive samples constitute only 10% of the entire dataset, if the model predicts all samples as negative, the accuracy can reach 90%. However, for the 10% positive samples, the model's ability to accurately predict is uncertain. In this case, the high accuracy is superficial and lacks practical significance. Therefore, to accurately assess the model performance, this paper introduces specificity and sensitivity. Specificity represents the false positive rate, and a high specificity indicates a low number of misdiagnosed samples. In simple terms, it reflects the model's ability to correctly identify TDC.

3.1 Results of combining Residual Networks with Bi-LSTM

Figure 1: The experimental results of the combination of 3D-ResNet and Bidirectional LSTM model

In Figure 1, the experimental results of training with the combination of four Residual Networks and Bi-LSTM are presented. The data in the table represents the average results as the paper utilized five-fold cross-entropy validation. From the experimental results, it can be observed that the performance of these four combined models is quite close, with the main differences manifesting in terms of accuracy and sensitivity.

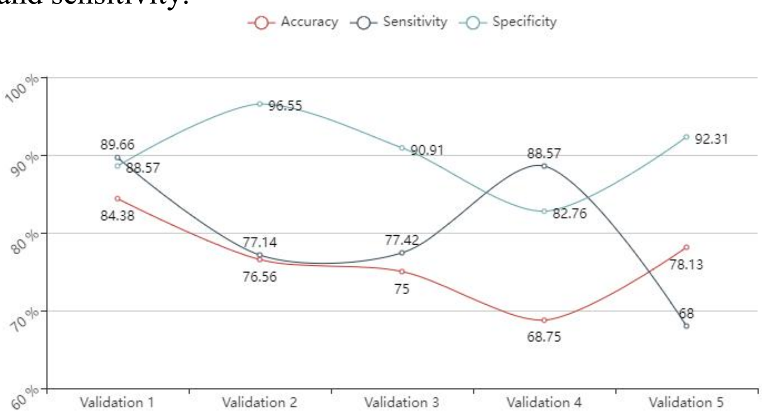

Figure 2: The performance of the combination of Residual Network C and Bi-LSTM using 5-fold cross-entropy validation

In the case of combining with Bi-LSTM, the accuracy of Residual Networks A and B is inferior to that of Residual Networks C and D. The combined model of Residual Network C achieved the highest sensitivity of 80.16% and the highest accuracy of 76.56%. During the training process, the models of Residual Networks C and D exhibit a faster convergence rate compared to those of Residual Networks A and B. Figure 2 illustrates the performance of this model using 5-fold cross-entropy validation[10]. As shown in the figure, except for validation set 4, the accuracy of other validation sets is greater than 75%. The comprehensive performance of sensitivity and specificity indicates that this combination has good adaptability and can fit the model's classification curve well.

3.2 Results of combining Residual Networks with Bi-GRU

Following the same method as the previous section, four Residual Networks were individually combined with Bi-GRU for training and validation. Figure 3 illustrates the performance of the models after 5-fold cross-entropy validation. The main difference in the current combination method is reflected in sensitivity, while the four Residual Networks show similar performance in accuracy and specificity. From the figure, it can be observed that the combined model of Residual

Network C has the best overall performance across these three metrics, and its accuracy, sensitivity, and specificity are 71.25%, 70.97%, and 89.13%. When the model's accuracy and specificity exhibit similar performance, sensitivity becomes a key indicator representing the performance differences among the four models.

The data in the figure indicates that except for validation set 2, the accuracy of the model combining Residual Network C with Bi-GRU is less than 75%. The accuracy on validation sets 4 and 5 is even less than 70%. Compared to the model combined with Bi-LSTM, the combination of Residual Network C and Bi-GRU is not stable enough, especially in terms of sensitivity. It can't fit the classification curve well and only performs well in certain specific intervals, and the results lack generality. Therefore, by comparing the data, it can be concluded that the model combining Residual Network C with Bi-LSTM exhibits the best performance in the task of ADHD classification recognition.

Figure 3: The experimental results of the combination of 3D-ResNet and Bidirectional GRU model

4. Conclusion

A three-dimensional residual network named 3D-ResNet which was combined with BRNNs was introduced in this study. Compared to techniques that involve fusing fMRI with MRI, the method proposed in this paper eliminated the need for complex image preprocessing; And compared to methods that extract low-level features from fMRI, this model retained spatial correlations while extracting features. This paper constructed four different structures of residual networks, and through ablation experiment, it demonstrated that the model combining Residual Network C with Bi-LSTM has the best performance. Under the 5-fold cross-entropy validation method, the average accuracy, sensitivity, and specificity are 76.56%, 80.16%, and 90.22%. Compared to existing methods, there is a significant improvement in accuracy when performing classification tasks on the multi-site ADHD-200 dataset. This result indicated that combining 3D-ResNet with BRNNs for assisting in the diagnosis of ADHD is feasible. What is even more promising is that this technology can be applied to the classification and diagnosis of other neurological disorders. It holds considerable prospects in studies based on rs-fMRI.

References

[1] Zhou, Dingfu et al. “Deep Learning Enabled Diagnosis of Children's ADHD Based on the Big Data of Video Screen Long-Range EEG.” Journal of healthcare engineering vol. 2022 5222136. 4 Apr. 2022.

[2] Cherkasova, Mariya V et al. “Review: Adult Outcome as Seen Through Controlled Prospective Follow-up Studies of Children With Attention-Deficit/Hyperactivity Disorder Followed Into Adulthood.” Journal of the American Academy of Child and Adolescent Psychiatry vol. 61,3 (2022): 378-391.

[3] Wiku L K ,Pingan W ,Hyun-Ho N , et al.Airborne hyperspectral imaging for early diagnosis of kimchi cabbage downy mildew using 3D-ResNet and leaf segmentation[J].Computers and Electronics in Agriculture,2023,214

[4] Shiqi W ,Kankan W ,Tingping Y , et al.Improved 3D-ResNet sign language recognition algorithm with enhanced hand features[J].Scientific Reports,2022,12(1):17812-17812.

[5] Li Darui, Hu Wanjun, Liu Guangyao, et al. Prediction of preoperative grading of glioma by 3D-ResNet101 deep learning model based on multi-center MRI [J]. Journal of Magnetic Resonance Imaging, 2019,14(05):25-30.

[6] Sun Z N, Wang KX, Huang W P, et al. Automatic screening of prostate multi-parameter MR Scan sequences by 3D ResNet deep learning model: a multi-center study [J]. Chinese Journal of Medical Imaging Technology, 2019,40(05):769-773.

[7] Kuang, Deping et al. “Discrimination of ADHD Based on fMRI Data with Deep Belief Network.” International Conference on Intelligent Computing (2014).

[8] Hao, A.J.; He, B.L.; Yin, C.H.: 'Discrimination of ADHD children based on Deep Bayesian Network', IET Conference Proceedings, 2015, p. 6 .-6 .

[9] Khater A W ,Madeed A S .Using 3D-VGG-16 and 3D-Resnet-18 deep learning models and FABEMD techniques in the detection of malware[J].Alexandria Engineering Journal,2024,8939-52.

[10] Niu Weihua, Zhai Ruibing. Research on Human behavior recognition in video based on improved 3D ResNet [J]. Computer Engineering and Science, 2019,45(10):1814-1821.