Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 21(4): 5803–5825.
DOI: 10.3934/mbe.2024256
Received: 30 December 2023
Revised: 13 March 2024
Accepted: 11 April 2024
Published: 28 April 2024

Research article

Unveiling critical ADHD biomarkers in limbic system and cerebellum using a binary hypothesis testing approach

Ying Chen[1], Lele Wang[1], Zhixin Li[1], Yibin Tang[2] and Zhan Huan[1,*]

[1] School of Microelectronics and Control Engineering, Changzhou University, Changzhou 213159, China

[2] College of Information Science and Engineering, Hohai University, Changzhou 213200, China

* **Correspondence:** Email: hzh@cczu.edu.cn.

Abstract: Attention deficit hyperactivity disorder (ADHD) is a common childhood developmental disorder. In recent years, pattern recognition methods have been increasingly applied to neuroimaging studies of ADHD. However, these methods often suffer from limited accuracy and interpretability, impeding their contribution to the identification of ADHD-related biomarkers. To address these limitations, we applied the amplitude of low-frequency fluctuation (ALFF) results for the limbic system and cerebellar network as input data and conducted a binary hypothesis testing framework for ADHD biomarker detection. Our study on the ADHD-200 dataset at multiple sites resulted in an average classification accuracy of 93%, indicating strong discriminative power of the input brain regions between the ADHD and control groups. Moreover, our approach identified critical brain regions, including the thalamus, hippocampal gyrus, and cerebellum Crus 2, as biomarkers. Overall, this investigation uncovered potential ADHD biomarkers in the limbic system and cerebellar network through the use of ALFF realizing highly credible results, which can provide new insights for ADHD diagnosis and treatment.

Keywords: ADHD; binary hypothesis; biomarker detection; limbic system; cerebellum

1. Introduction

Attention deficit hyperactivity disorder (ADHD) is a common neurodevelopmental disorder that is diagnosed in childhood. It is characterized by impulsivity, inattention, and hyperactivity [1], which have a detrimental impact on children's learning, emotions, and social relationships. In recent years, pattern recognition has been applied to neurological disease diagnosis, and significant progress has been made in ADHD diagnosis through the develpopment of various classification approaches that utilize machine learning (ML) and deep learning (DL) techniques [2, 3]. Nowadays, improving

diagnosis accuracy remains a practical challenge for current ADHD classification methods, primarily due to complex factors such as limited data size and noise disturbance in sampled data. Furthermore, the identification of discriminative biomarkers for ADHD represents an essential application requirement. These biomarkers can serve as keys to uncover the mechanisms underlying ADHD and facilitate more convenient and accurate diagnosis and treatment. Accomplishing this task requires achieving satisfactory accuracy and effectively addressing the challenge of interpreting features obtained through the use of learning algorithms.

As we know, the type of biosignal is crucial in ADHD classification and biomarker detection. Suitable biosignals can extremely improve ADHD classification accuracy and yield highly credible biomarker results. Here, the magnetic resonance imaging (MRI) technique provides plenty of metrics, such as cortical thickness, gray matter probability, regional homogeneity (ReHo), amplitude of low-frequency fluctuation (ALFF), and functional connectivity (FC), to elucidate the brain status of patients. Voxel-level metrics describe the detailed brain state at a high spatial resolution. For example, complementary features can be extracted from voxel-level functional MRI (fMRI) and structural MRI (sMRI) data to characterize differences among subjects [4], while binary voxel-level ALFF feature maps are employed and input to an attention-based DL classification network [5]. However, the extraction of features from voxel-level data has significant problems, as learned features often suffer from noise interference due to the limited consideration of relationships among these voxel-level data. As a result, the performance of voxel-based ADHD classification methods is poor, greatly reducing the reliability of biomarker detection in these studies. Region-level metrics can be obtained by matching the original resolution fMRI data with a certain brain template, which preserves the functional similarity of the voxels within the region and effectively elevates the classification accuracy, while providing better interpretability for ADHD disease. Therefore, region-level metrics are preferred. For example, region-level FC has been frequently utilized for accurate ADHD classification and biomarker analysis [6–8]. However, FC is a numerical measure that assesses the correlation between different brain regions. Consequently, when utilized, the identified biomarkers are dispersed in practice [7], making it difficult to in focus on specific brain regions and undermining the interpretability of the obtained outcomes. Studies have shown that spontaneous low-frequency (< 0.08 Hz) fluctuations are highly synchronized between the motor cortices, while ALFFs imply spontaneous neuronal activity within the region [9]. Therefore, region-level ALFF has better functional aggregation, making it more suitable for the detection of biological markers of ADHD. We employed region-level ALFF for ADHD classification and biomarker detection in this study, where abnormal ALFF values serve as indicators of changes in ADHD-related brain regions.

The effectivity of biomarker detection also depends on the detection approaches. As far as we know, there exist two major ways to elucidate ADHD biomarkers. 1) One way is statistical analysis, which compares the differences between ADHD group and control group data. Here, the characteristics associated with credible brain region differences were found by various statistical methods including the volumes of hippocampus and amygdala [10] and the activity intensity of the cerebellum area [9]. Unfortunately, these statistical conclusions have limited capacity in ADHD individual diagnosis, since the ADHD condition is different for each subject. 2) The other way is to employ feature selection or learning strategies during the ADHD classification procedure. For example, high-score FC vectors are determined using by a specific feature ranking method [6], and the learned convolution kernels are mapped to three-dimensional space via reconstruction techniques

to determine the location of highlighted brain regions [11]. But, it is highly challenging to design a satisfactory classification method that is also capable of reliable biomarker detection. On one hand, existing classification methods with ML are good at employing feature selection to find typical features from the input raw data through the various effective algorithms, such as a support vector machine with recursive feature elimination (SVM-RFE) [12], the least absolute shrinkage and selection operator (LASSO) [13, 14], and the elastic net [15]. However, ML-based methods frequently use linear classifiers for ADHD prediction, and they cannot fully describe the complex relationship between features and ADHD disease. Consequently, it has low accuracy [16] and does not provide reliable biomarkers. On the other hand, DL methods exhibit remarkable classification performance, as their flexible feature learning capability facilitates the learning of potential ADHD-related features. For example, three-dimensional [4] and four-dimensional [17] convolutional neural networks (CNNs) have been directly applied to sMRI and fMRI medical image data, exploring the time and spatial patterns of ADHD features in MRI data. Several attention mechanisms have been adopted to adaptively emphasize the learned discriminative features and thus improve the classification performance [18, 19]. In addition, autoencoder (AE) networks have also been widely used in ADHD classification. A deep variational AE (DVAE) network and its advanced version, the spatiotemporal attention AE (STAAE) network, have yielded impressive classification results with an accuracy of over 93% [20, 21]. However, in these cases, the learned features from DL-based methods are viewed as high-level features, which have poor interpretability and cannot meet the biomarker detection requirement. In summary, both methods for biomarker detection have limitations. However, it is an interesting idea to combine the two, as it would introduce the prior knowledge obtained from statistical analysis into the classification method to guide the classification and enhance the interpretability of related derived biomarkers.

Attracted by the superior performance of the AENet-based binary hypothesis testing (BHT) framework, we have employed this model to elucidate the related biomarkers from regional ALFF data. The main contributions of the study are described as follows:

1) We use prior knowledge to guide the classification procedure. Since ADHD children suffer in executive functions, cognitive and emotional control, we applied ALFF to the data on the 50 related brain regions in the limbic system and cerebellum. Moreover, using the attributes of ALFFs as a basis, we modified the existing AENet to be more suitable for feature learning on ALFF data. Considering that the ALFF data on limited brain regions cause the AENet to exhibit unstable feature learning and thus degenerate the classification performance, we have also introduced an ensemble learning strategy to enhance the accuracy. As a result, we achieved remarkable results on multiple datasets with an average accuracy of 93.3%.

2) Regarding high-level classification performance, biomarker detection was effectively carried out. Several potential biomarkers were identified from the selected features by implementing an SVM-RFE algorithm in the BHT framework. Then, we employed a two-sample t-test between groups and correlation analysis based on the ADHD symptom scores to verify the rationality of these biomarkers. We found that brain regions, including the thalamus, hippocampus, and cerebellar lobule X, were most discriminative in our experiments, which is in line with the existing statistical analysis reports. It further demonstrates that the biomarkers obtained via classification and statistical analyses exhibit consistency in the limbic system and cerebellum.

2. Related work

2.1. ADHD effect on limbic system and cerebellum

The limbic system involves a group of regions in the paleocortex that support various functions related to emotion regulation and motivation [22]. In detail, it mainly consists of the amygdala, hippocampus, striatum, and thalamus, which are known to be implicated in ADHD. Several studies have confirmed significant statistical differences between ADHD and healthy control (HC) subjects in the limbic system. For example, structural integrity is impaired in brain regions such as the thalamus [23], caudate nucleus [24], and amygdala [25, 26] in children with ADHD. Additionally, abnormal volumes have been observed in these above regions, as well as the hippocampus [27, 28]. Moreover, FC analysis has revealed significant variations in neural connections involving the thalamus and hippocampus regions [29, 30].

Initially believed to be primarily involved in motor learning and coordination, the cerebellum is now recognized to play a significant role in cognition and emotion, thus making a vital contribution to the pathophysiology of ADHD [31]. Numerous studies have shown that the cerebellum undergoes the most noticeable structural and functional changes for ADHD [32]. Children with ADHD also exhibit a significant decrease in the ALFF signal in the bilateral cerebellum [9], accompanied by impaired structural integrity [33, 34].

In summary, statistical differences associated with ADHD have been found in these brain regions, which means that these findings highlight the potential of the limbic system and cerebellum to serve as a source of biomarkers for ADHD detection and assist in its classification process. Hence, using this prior knowledge, the limbic system and cerebellum can be used as input brain regions. In this way, highly correlated input data can be obtained, and the interpretability of the intermediate features generated via the classification method is also enhanced.

2.2. FC-based BHT classification framework

To the best of our knowledge, two commonly employed classification frameworks exist in the field of ADHD diagnosis: the training-test framework and the hypothesis-test framework [7]. The training-test framework dominates ADHD classification, involving the learning of features from training data their comparison with those of test data to predict labels. However, this strategy proves inadequate for small-sized datasets, leading to unsatisfactory performance. These learned training-data features fail to fully encompass the characteristics of test data, resulting in a significant hindrance to accuracy improvement due to the limitation of sample number [7]. In contrast, the hypothesis-test framework follows a semi-supervised approach, where we have successfully designed a BHT framework for ADHD classification [35–37]. The core idea behind the BHT framework is to incorporate the test data, along with an assumed label, into the feature learning process for the training data. In scenarios in which the assumed label is incorrect, this feature learning process becomes disrupted because the test data introduces noise-like elements that are not present in the training data. As a result of comparing the training features learned under different assumptions, predicting the labels of the test data becomes easier. Therefore, the BHT framework is suitable for research on the ADHD-200 dataset due to its anti-noise ability and ability to overcome the limitation of sample size.

Figure 1 depicts the flowchart of the existing BHT framework, which utilizes FC as the raw data

input (more details can be found in reference [7]). The test data were initially categorized as either HC ($\mathcal{H}_0$) or ADHD ($\mathcal{H}_1$) data. For feature selection, the SVM-RFE approach is employed to choose the N most characteristic connections for both the training and test data, wherein only the connections from the training data are retained to form two typical feature sets, $X^{\mathcal{H}_0}$ and $X^{\mathcal{H}_1}$. Subsequently, a feature extraction process is performed by using a modified AE network to process these feature sets and generate the corresponding high-level feature sets, $\tilde{X}^{\mathcal{H}_0}$ and $\tilde{X}^{\mathcal{H}_1}$. During the ADHD decision step, the inter-class and intra-class variability scores, $D^{\mathcal{H}_0}$ and $D^{\mathcal{H}_1}$, are computed for the high-level feature sets and compared. The assumption with a smaller variability score is identified as the true hypothesis, $\tilde{\mathcal{H}}_{true}$, with the corresponding label assigned to the test data.

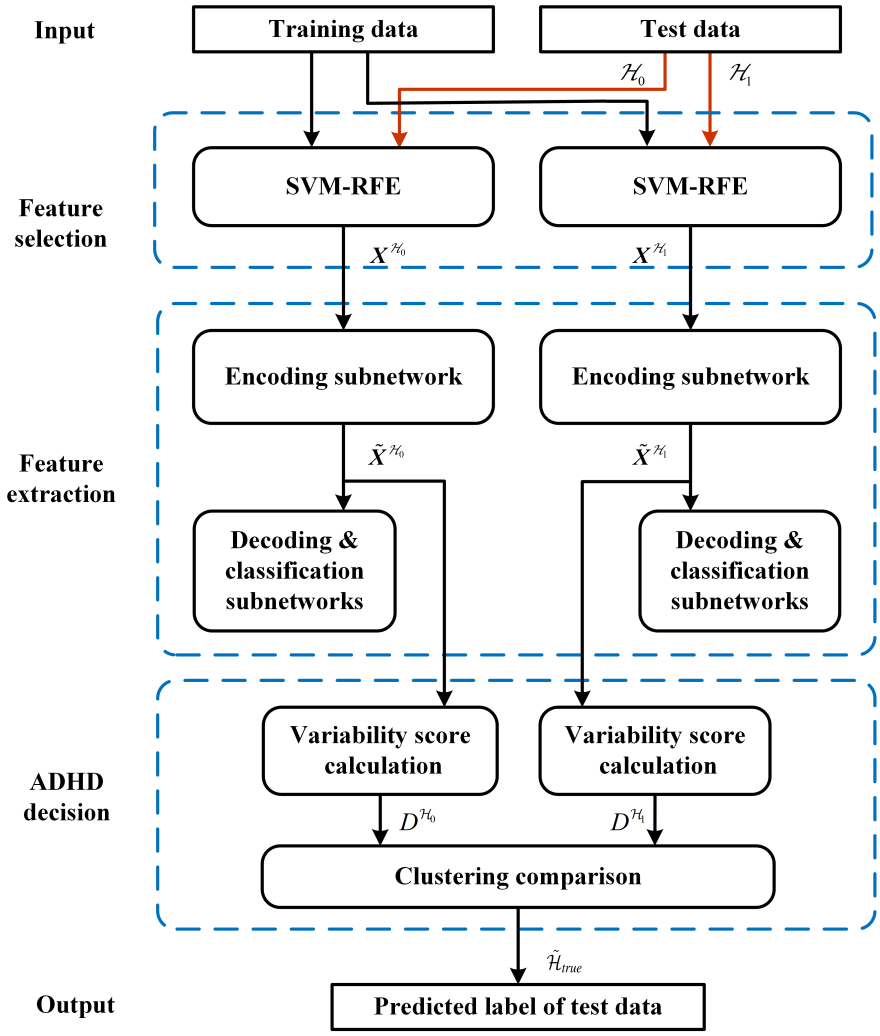

Figure 1. Flowchart of BHT framework. $\mathcal{H}_0$ and $\mathcal{H}_1$ correspond to different assumptions. Typical feature sets ($X^{\mathcal{H}_0}$ and $X^{\mathcal{H}_1}$) and high-level feature sets ($\tilde{X}^{\mathcal{H}_0}$ and $\tilde{X}^{\mathcal{H}_1}$) are obtained in the feature selection and feature extraction stages, respectively. Then, the variability scores ($D^{\mathcal{H}_0}$ and $D^{\mathcal{H}_1}$) are calculated in the decision stage for ADHD and the correct hypothesis ($\tilde{\mathcal{H}}_{true}$) is selected based on the score.

It should be noted that this FC-based model suffers from two problems in biomarker detection. Firstly, during feature selection, a total of 4005 connections are used, resulting in scattered characteristic connections. This severely undermines the meaningfulness of the identified biomarkers, as we believe that biomarkers should be localized in specific regions. Second, no prior knowledge is considered during feature selection. As a result, this purely data-driven framework may introduce bias and compromise the reliability of the biomarkers. To overcome these problems, we propose using a limited number of regional ALFFs in the limbic system and cerebellum for biomarker detection.

3. Methodology

3.1. Data preprocessing

All the data used in this study were taken from the Athena pipeline in the ADHD-200 preprocessed dataset of the (http://preprocessed-connectomes-project.org/adhd200/), which contains data of seven sites from the ADHD-200 competition dataset. Note that the University of Pittsburgh and Washington University in sT. Louis sites only contain HC samples, while there is relatively little research on the Oregon Health & Science University sites. Thus, we ultimately chose the remaining four sites for the experiment, including New York University Medical Center (NYU), Peking University (PU, comprising three PU sub-sites), Kennedy Krieger Institute (KKI), and NeuroIMAGE (NI). Detailed information on the samples from these four datasets is recorded in Table 1.

Table 1. Summary of several ADHD-200 sites.

Site	Age	Female/Male	Control	ADHD	Total
NYU	7–18	76/140	98	118	216
PU	8–17	52/142	116	78	194
KKI	8–13	37/46	61	22	83
NI	11–22	17/31	23	25	48

Table 2. Used brain regions in limbic system and cerebellum.

Region name	Region index	Region name	Region index
Olfactory cortex	21–22	Amygdala	41–42
Insula	29–30	Caudate	71–72
Anterior cingulate	31–32	Putamen	73–74
Middle cingulate	33–34	Pallidum	75–76
Posterior cingulate	35–36	Thalamus	77–78
Hippocampus	37–38	Cerebellum	91–108
Parahippocampal	39–40	Vermis	109–116

We utilized a meticulous approach to extract regional ALFFs from the downloaded time course values of blood-oxygen-level-dependent (BOLD) signals recorded from the subjects in our study. To derive the voxel-based ALFFs, we implemented several standard operations, including filtering (band-pass 0.009–0.08 Hz) to eliminate noise, performing a fast Fourier transform, and calculating the power spectrum proportionally. Subsequently, we employed the automated anatomical labeling

(AAL-116) atlas to divide the brain into distinct regions. Finally, regional ALFF values were obtained by averaging the voxel-based ALFF values within the regions of interest (ROIs), where a total of 50 ROIs are provided in Table 2 with their brain indices. But please note that there might be a slight deviation for the limbic system, which contains the entire cingulate gyrus including the anterior, middle, and posterior cingulate gyri.

3.2. ALFF-based BHT framework

We introduce the ALFF-based BHT framework for ADHD classification, wherein regional ALFFs replace FC data as the input raw data. Similar to the FC-based framework, we focus on enhancing the feature extraction step within the utilized AE network. The architecture of our ALFF-based AE network is illustrated in Figure 2. Within this diagram, the typical features (i.e., characteristic ALFF) of the training data are shown to undergo encoding via an encoder subnetwork to derive their high-level representations. These representations are then passed through a decoder subnetwork for reconstruction. Simultaneously, a classification subnetwork supervises the labeling of high-level features acquired through the encoder subnetwork. This strategy aims to retain category information related to the training data within the high-level features.

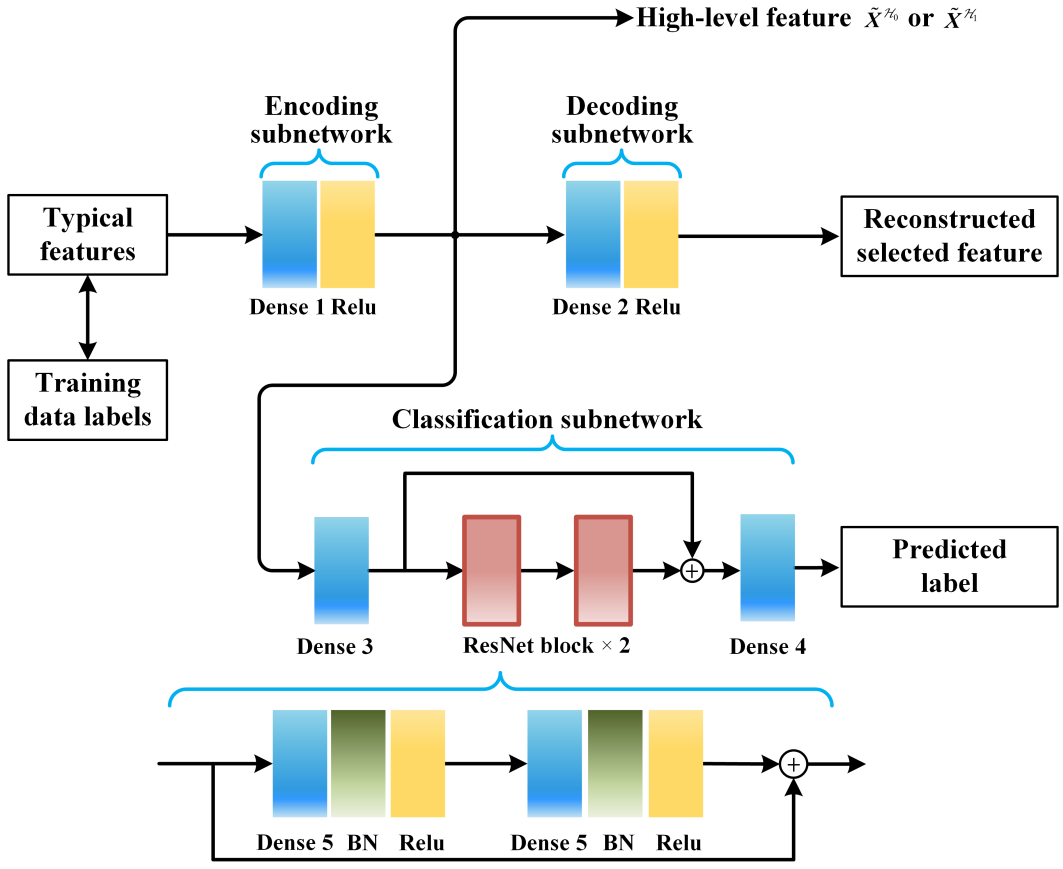

Figure 2. Structure of ALFF-based AE network. The proposed AE network includes encoding, decoding, and classification subnetworks. At the same time, the details of ResNet block in the classification subnetwork are also given.

In contrast to the FC-based AE network [7], we have introduced several targeted modifications to our network architecture, making it specifically tailored to harness the attributes of ALFF. Notably, we have incorporated a subsequent ReLU into the dense layer within the decoding subnetwork. This adjustment ensures that the reconstructed ALFF maintains non-negative values. Furthermore, in practical implementation, our AE network employs the ALFF-based selected features with an output dimension (i.e., the output dimension of SVM-RFE) that has been reduced to 25, as opposed to the 50 dimensions pf the FC-based features. This conscious reduction mitigates the risk of overfitting that might arise when directly inputting these features into the FC-based AE network. Consequently, we have optimized the architecture of the classification subnetwork. Presently, the classification subnetwork employs just two residual network (ResNet) blocks, designed to process an input vector with dimensions of 8×1. This departure from the 20×1 input used in the FC-based AE network results in the generation of more efficacious high-level features. Lastly, we have meticulously detailed the parameters for each dense unit within our AE network in Table 3, note that the size of each dense layer is determined by grid search. The loss functions applied for the reconstructing of the selected features and prediction of the corresponding labels closely align with those employed in a previous study [7], ensuring methodological consistency.

Table 3. Size of dense layers.

	Layer size*
Dense 1	(25, 10)
Dense 2	(10, 25)
Dense 3	(10, 8)
Dense 4	(8, 8)
Dense 5	(8, 2)

* The parameters (a, b) describe a dense unit with input size of a and output size of b.

3.3. Enhanced classification with ensemble learning

Although adjusting the AE network to accommodate ALFF data represents a positive step, a prominent challenge persists. It centers around the inherent instability of the AE network, which directly impacts the final prediction outcomes. Throughout network training, the initial configuration of learned parameters wields substantial influence. Optimal initial values possess the capacity to expedite the training process and guide the network toward a state of stability. However, when handling a small-sized dataset, the effect of these initial values is disproportionately amplified, thereby compromising the network's robustness and introducing an element of uncertainty. In practical implementation, our ALFF-based AE network is susceptible to this uncertainty. Even when holding the ALFF input constant for both training and test data, the resulting ADHD prediction outcomes may exhibit minor variances with a low probability. This uncertainty significantly impedes the pursuit of refined ADHD classification accuracy, consequently undermining the efficacy of biomarker detection endeavors.

In this study, we have employed ensemble learning as a powerful tool to address this challenge. Ensemble learning stands as a classical strategy that is renowned for mitigating issues related to data imbalance, model robustness, and uncertainty estimation. Its basic idea is to integrate several weak

classifiers and build a fortified classifier, which engenders heightened reliability within the classification outcomes. The existing examples involve the collection of classifiers from the fields of ML and DL for application in various medical diagnostic scenarios, such as Parkinson's disease classification [38], anticancer peptide prediction [39], and virulence factor detection [40]. This strategy was also successfully utilized for ADHD diagnosis, where solid decision-making can be realized based on multimodal data [41].

The proposed ensemble classification method is illustrated in Figure 3. For a given test data, the process involves generating the typical features sets ($X^{\mathcal{H}_0}$ and $X^{\mathcal{H}_1}$) of training data based on opposite test label assumptions in the feature selection step. Then, multiple pairs of AE networks are utilized, each initialized with random parameters, to acquire multiple pairs of high-level feature sets, $\tilde{X}^{\mathcal{H}_0}$ and $\tilde{X}^{\mathcal{H}_1}$. Subsequently, pairwise high-level features sets are compared to determine the true hypothesis and corresponding hypothesis labels (L_{true}), which compares the intra-class and inter-class distances ratios of each feature set. The hypothesis label is considered as the prediction result of a baseline classifier. Finally, a hard voting strategy is then applied to these hypothesis labels. By assessing the frequency of label values (0 for ADHD and 1 for HC), class labels of the data are assigned higher frequency label values.

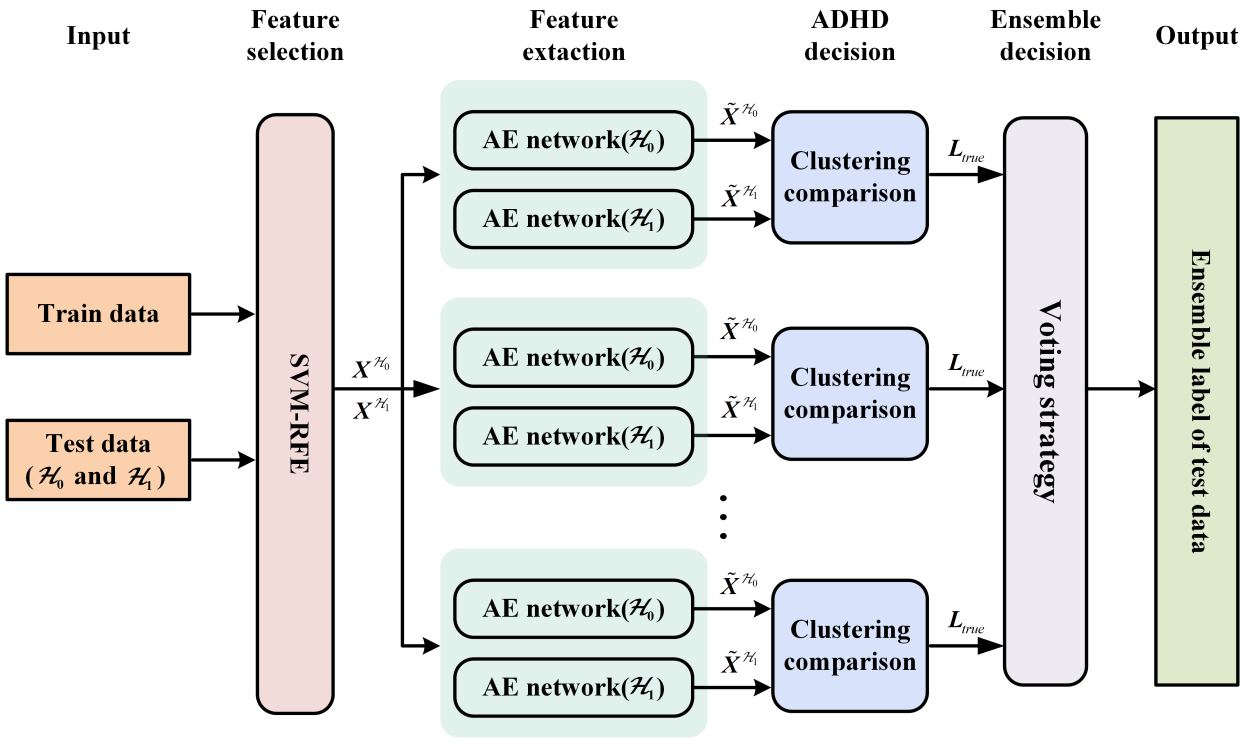

Figure 3. ALFF-based BHT framework. Different from the existing BHT framework, the voting strategy is applied to the output label (L_{true}) of the ADHD decision stage, which improves the reliability of results. The ensemble label obtained by voting is regarded as the final prediction label of test data.

3.4. Extraction strategy for biomarkers

As described for the BHT framework in Section 2.2, typical feature sets were selected by the SVM-RFE algorithm in the feature selection stage. Meanwhile, reliability weights W_{ji} of typical feature sets can be obtained during this process. Specifically, the SVM-RFE algorithm trains a linear SVM in one iteration to obtain a weight vector that fits to the input feature set. Then, the square of the weight vector is used as the criterion for judging the usefulness of features, and the feature with the smallest-valued criterion is removed. After removing a certain number of features over multiple iterations, a typical feature set with the expected dimensions is obtained. The square of the fitted linear SVM weight vector on this feature set is considered to comprise the reliability weights of the features.

Reliability weights not only measure the contribution of features to classification they also serve as a criterion for subsequently extracting brain biomarkers. However, there are differences in the values and ranks of reliability weights at different sites. Moreover, under the binary hypothesis framework, the typical feature subset generated under the correct hypothesis is more valuable. Therefore, the weighted average of the feature reliability weights were designed to be from the correct hypotheses for the four sites, where the weighted value is the product of the number of people and the classification accuracy for the site. The weighted average result is seen as a more comprehensive measure to evaluate each feature, and it is called the feature score S_i for each ROI. The feature score can be defined as follows:

$$S_i = \frac{\sum_{j=0}^{4} Acc_j \times N_j \times W_{ji}}{\sum_{j=0}^{4} Acc_j \times N_j}, \tag{3.1}$$

where Acc_j and N_j respectively denote the classification accuracy after ensemble learning and data size of the j-th dataset, W_{ji} represents the reliability weight for the i-th ROI of the j-th dataset. This feature score provides enough convenience for finding ADHD biomarkers.

4. Experimental result

4.1. Classification performance

The classification performance of our BHT framework with ALFF-based AE network is presented in Table 4, where the accuracies with and without ensemble learning are provided. Specifically, each site was subjected to 50 leave-one-out cross-validation (LOOCV) trials, and the average accuracy without ensemble learning strategies was obtained. The results with ensemble learning were derived from the average of 1000 hard voting trials, where each trial entailed randomly selecting seven hypothesis labels (L_{true}) from the 50 LOOCV trials. During the AE network learning, an Adam optimizer is utilized to optimize the whole network. Hyperparameters of each site were determined through grid search, including the learning rate, training epoch, and the rate of dropout. After a certain number of epochs, the training loss converges and becomes stable. In addition, ablation experiments were designed to verify the role of components of the ALFF-based AE network, as well as under the BHT framework. 1) Default-classifier network: the classifier subnetwork was hidden, only the complete AE network was retained, and the entire network was trained by using the reconstruction loss. 2) Default-decoder network: the decoder subnetwork was hidden, the encoder degenerated into a dense layer, and the whole network was trained by using the cross-entropy loss. Later, ADHD decisions were made based on the encoder output features of the default-classifier network and the output features of the first

dense layer of the default-decoder network, respectively. The results of ablation experiments are shown in Table 5, which shows that the hyperparameter and experimental details were identical to those of the ALFF-based AE network. The source code is available at https://github.com/BiolabHHU/ALFF-based-BHT.

In Table 4, one can see that an evident distinction emerges in the classification performance when comparing the results in the absence of ensemble learning to those in its presence, and it is manifested in the differences in metrics, including accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), F1 score (F1), and apply capitalization correlation coefficient (MCC). The absence of ensemble learning resulted in notably inferior performance, as indicated by the average accuracy of 87.0%. Notably, when confronted with the PU dataset, the accuracy decreased to an unsatisfactory 80.3%. This decline is attributed to a high prevalence of comorbidity disorders among the ADHD children within this dataset, affecting 44 out of 78 cases. Conversely, the application of ensemble learning resulted in a marked enhancement in the accuracy metrics. The average accuracy significantly increased, reaching 93.3%. This positive impact is particularly evident in the case of the PU dataset, where accuracy improved dramatically from the aforementioned 80.3% to 88.8%. This result effectively substantiates the ability of ensemble learning to augment the potency of weaker classifiers and make them robust. Moreover, Table 4 also presents the results of measurements of the area under the curve. Specifically, the average AUROC and AUPRC values were both around 0.93, underscoring the excellent performance of the ALFF-based BHT framework on the ADHD-200 dataset. Remarkably, this heightened level of accuracy was achieved through the utilization of a mere 25 selected features from 50 regional ALFFs. This outcome serves to reinforce the presumption that our approach facilitates the biomarker detection endeavors. Additionally, we provide the receiver operating characteristic (ROC) curves for the used datasets in Figure 4, which shows that the curve for the KKI site had the largest area among these datasets.

Table 4. Classification performance for various datasets.

Site	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUROC	AUPRC	MCC	F1
Without ensemble learning							
NYU	91.94	90.59	93.55	0.9207	0.8913	0.8396	0.9239
PU	80.32	78.87	81.29	0.8008	0.7984	0.5994	0.7697
KKI	88.19	87.45	88.46	0.8796	0.8817	0.7445	0.8256
NI	87.38	88.40	86.26	0.8733	0.8296	0.7488	0.8786
Average	86.96	86.33	87.39	0.8686	0.8503	0.7331	0.8495
With ensemble learning							
NYU	96.20	96.88	95.38	0.9616	0.9400	0.9250	0.9660
PU	88.79	96.67	83.50	0.8982	0.9074	0.7801	0.8719
KKI	95.05	98.90	93.66	0.9656	0.9815	0.9033	0.9276
NI	93.23	98.32	87.70	0.9302	0.9174	0.8667	0.9374
Average	93.32	97.69	90.06	0.9389	0.9366	0.8688	0.9257

Table 5. Comparison of classification accuracy of the three networks.

Network	NYU (%)	PU (%)	KKI (%)	NI (%)	Average (%)
Without ensemble learning					
Default-classifier network	69.00	53.69	42.82	67.67	58.29
Default-classifier network	89.83	80.26	86.84	86.79	85.93
ALFF-based AE network	**91.94**	**80.32**	**88.19**	**87.38**	**86.86**
With ensemble learning					
Default-classifier network	75.01	52.38	35.51	75.49	59.60
Default-classifier network	95.14	85.45	91.29	92.39	91.07
ALFF-based AE network	**96.20**	**88.79**	**95.05**	**93.23**	**93.32**

To further validate the utilization of ensemble learning, we constructed a box-and-whisker plot to show the accuracy distribution, as shown in Figure 5. It is evident that the adoption of ensemble learning leads to an enhancement in the average accuracy. While this learning strategy may not entirely eliminate the fluctuations in accuracy owing to the inherent uncertainty of the AE network, can be ascertained from Table 4, there was a substantial reduction in the standard deviation of accuracy. Especially, the enhancement of accuracy are obviously disclosed on the PU dataset. These findings prove the effectiveness of ensemble learning as a tool to enhance the robustness of our ALFF-based BHT framework.

Table 5 shows a comparison of the classification accuracy of the three networks. The accuracy of the default-classifier network was only 42–76%, which may have been caused by the lack of information learned from the label based on the guidance of the cross-entropy loss; thus, the extracted high-level features had no inter-class discrimination. The classification accuracy of the default-decoder network was 1–4% lower that of proposed ALFF-based AE network, indicating that the lack of reconstruction loss may reduce the representation ability of high-level features.

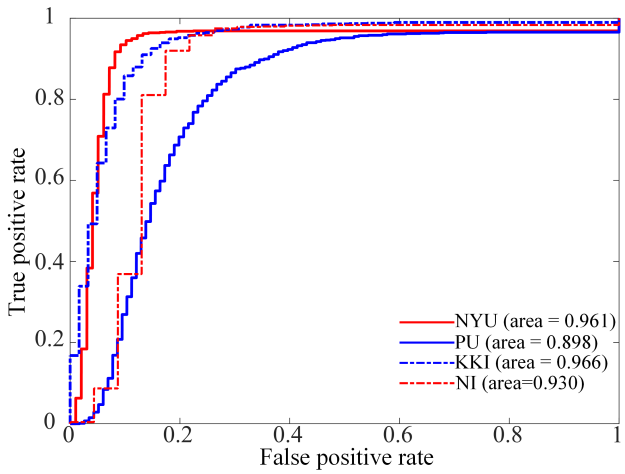

Figure 4. ROC curves for various data sites.

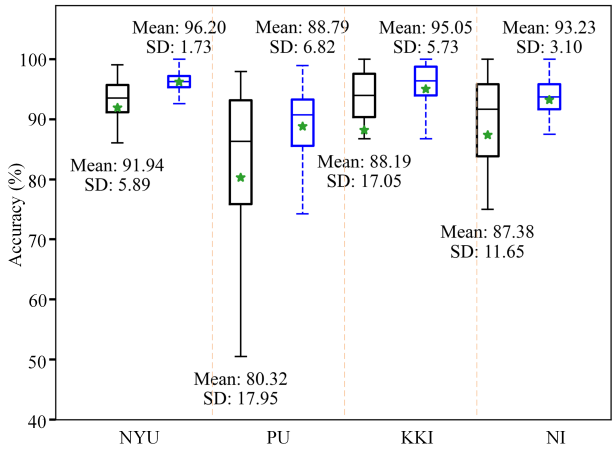

Figure 5. Accuracy distribution with and without ensemble learning, presented in the form of a box-and-whisker plot. Each dataset is tagged with its accuracy's mean and the standard variation (SD) value. The absence of ensemble learning is colored with black, whereas the application of ensemble learning is in blue.

We tested our method against other state-of-the-art methods the results are presented in Table 6. The compared methods include ML-based methods such as R-Relief [42], L1BioSVM [43], and Fusion fMRI [44], as well as DL-based methods such as the 3D CNN [4], DeepfMRI [45], CDAE [46], DVAE [20], STAAE [21] and data augmentation [47]. In addition, our previous work under the BH framework is also included in this comparison. These methods are referred to as SP-BH [37] and SP-$l_{2,1}$-BH [36], which use subspace learning and $l_{2,1}$-norm subspace learning for ADHD classification, respectively.

From Table 6, it is observed that the existing ML methods had the lowest accuracies, stemming from their limited exploration of potential features for classification. Meanwhile, most examined DL methods are confined to the traditional training-test framework, which causes the learned features from the training data to inadequately representing the features of the test data. Among these methods, the STAAE approach stands out because of its remarkable performance achievements and a commendable biomarker detection capability. Its success benefits from the integration of an attention module that captures temporal patterns from fMRI spatiotemporal features. Based on its accurate classification, STAAE extends its application by projecting these temporal patterns onto corresponding brain regions, effectively fulfilling the biomarker detection task. The listed methods within the BHT framework all exhibited impressive classification performance, facilitated by the utilization of test data information (without seeing its label). Among these methods, AENet leverages DL to effectively represent high-level features, making it superior to the SP-BH, SP-$l_{2,1}$-BH approaches. As an FC-based BHT method, AENet incorporates a great number of connections, consequently leading to a widespread selection of features across the entire brain. This, however, renders the identified biomarkers indiscriminate. In contrast, our ALFF-based method achieved an average accuracy of 93.3%, even though it was slightly inferior to that of AENet. A distinctive trait lies in the substantial reduction of both the used raw and selected features. This implies that the features used and selected by our algorithm carry more significance as potential ADHD biomarkers. Moreover, the selected features were tightly localized within the limbic system and cerebellum, rendering our ALFF-based method exceedingly interpretable.

Table 6. Accuracy comparison between our method and state-of-the-art methods.

	NYU (%)	PU (%)	KKI (%)	NI (%)	Average (%)	Material	Raw features	Selected features*	Biomarker detection
ML									
Fusion fMRI (2018)	52.7	-	86.7	72.9	70.8	FC	4005	-	No
L1BioSVM (2018)	-	81.1	81.3	-	81.2	FC	6670	-	No
R-Relief (2019)	70.7	68.6	81.8	76.0	74.3	fALFF	$31 \times 37 \times 31$	-	No
DL									
3D CNN (2017)	70.5	63.0	-	72.8	68.8	MRI image	$90 \times 117 \times 100$	1024	No
DeepfMRI (2020)	73.1	62.7	-	67.9	67.9	BOLD	90	32	No
CDAE (2021)	73.2	70.6	81.7	79.0	76.1	MRI image	$60 \times 72 \times 60$	$5 \times 6 \times 5$	No
DVAE (2021)	62.4	67.0	78.1	68.8	69.1	BOLD	28,546	80	No
STAAE (2022)	93.5	92.7	90.4	91.7	92.1	BOLD	28,546	100	Yes
Data augmentation (2023)	75.6	76.5	76.0	-	76.0	FC	13,456	1856	No
BHT framework									
SP-BH (2019)	96.2	95.8	86.7	91.6	92.6	FC	4005	100	No
SP-$l_{2,1}$-BH (2020)	99.5	96.3	100	95.8	97.9	FC	4005	50	No
AENet (2022)	99.8	99.6	99.8	99.3	99.6	FC	4005	50	Yes
Our	96.2	88.8	95.1	93.3	93.3	ALFF	50	25	Yes

* The number of feature selected in the ML method is not fixed. And the DL method selects the learned high-level features, while the BHT method selects typical features from the raw features.

4.2. Biomarker detection

Table 7 presents the top 10 brain regions obtained with the highest feature scores S_i; their locations are visualized in Figure 6. These identified regions are considered to be potential ADHD biomarkers. To substantiate our findings, we conducted a correlation analysis between the ALFF data from these regions and the symptom scores provided by the ADHD-200 dataset; the results are shown in Table 7. Since the NI site did not give symptom assessment data, we performed correlation analysis for the PU, KKI and NYU sites. Notably, only individuals with ADHD-simplex were included in these analyses to mitigate the influence of comorbid disorders. In Table 7, it is evident that the majority of detected biomarkers had a robust correlation with symptom scores from the PU dataset, demonstrating significant values above 0.2, under the 95% confidence interval. This finding underscores the potential of our identified biomarkers to capture ADHD-related characteristics. However, an intriguing observation arose from the results obtained for the other two datasets. In this case, the correlation between biomarkers and symptoms was notably weaker, especially for the NYU site, as reflected by the higher P values over 0.05. This discrepancy can be attributed to the utilization of distinct symptom assessment measures across these datasets. To elaborate, although three datasets were derive from symptom measures based on questionnaires administered to parents and teachers, the specific scales employed differ. The PU dataset employs the ADHD Rating Scale-IV (ADHD-RS), whereas the KKI and NYU datasets utilize Conners' Parent Rating Scale-Revised, Long version (CPRS-LV). Despite the shared origin in parental and teacher reports, we posit that ADHD-RS offers

advantages in terms of its ability to abnormalities within the limbic system and cerebellum. This assertion underscores the potential for ADHD-RS to provide insights into these specific regions.

Figure 7 supplements our analysis and visualizes the results of correlation analysis for ADHD-RS scales, encompassing regions the left hippocampal gyrus, left and right thalamus, and left cerebellum Crus II gyri, which are the four regions with the highest correlation coefficients. It further supports that these regions exhibit significant relevance to ADHD, thereby warranting consideration as potential biomarkers.

Table 7. Detected biomarkers and their symptom score correlations for the PU, KKI and NYU datasets.

Brian region		Ranking order	PU (ADHD-RS)		KKI (CPRS-LV)		NYU (CPRS-LV)	
Name	Abbrev		Corr	P-value	Corr	P-value	Corr	P-value
Middle cingulate gyrus (R)	DCG.R	1	**0.235**	**0.007**	**0.358**	**0.003**	-0.035	0.660
Cerebellum IX (R)	CRBL9.R	2	0.161	0.065	0.155	0.213	0.038	0.630
Amygdala (R)	AMYG.R	3	**0.198**	**0.023**	0.057	0.648	-0.005	0.953
Thalamus (L)	THA.L	4	**0.245**	**0.005**	0.150	0.229	0.044	0.572
Cerebellum X (R)	CRBL10.R	5	**0.244**	**0.005**	0.241	0.051	0.021	0.787
Cerebellum Crus II (L)	CRBL Crus2.L	6	**0.245**	**0.005**	**0.259**	**0.036**	0.128	0.102
Cerebellum Crus II (R)	CRBL Crus2.R	7	**0.214**	**0.014**	0.236	0.057	0.123	0.115
Thalamus (R)	THA.R	8	**0.270**	**0.002**	0.184	0.139	0.043	0.580
Hippocampus (L)	HIP.L	9	**0.274**	**0.001**	**0.273**	**0.027**	0.021	0.786
Caudate (R)	CAU.L	10	**0.235**	**0.007**	0.215	0.083	0.034	0.667

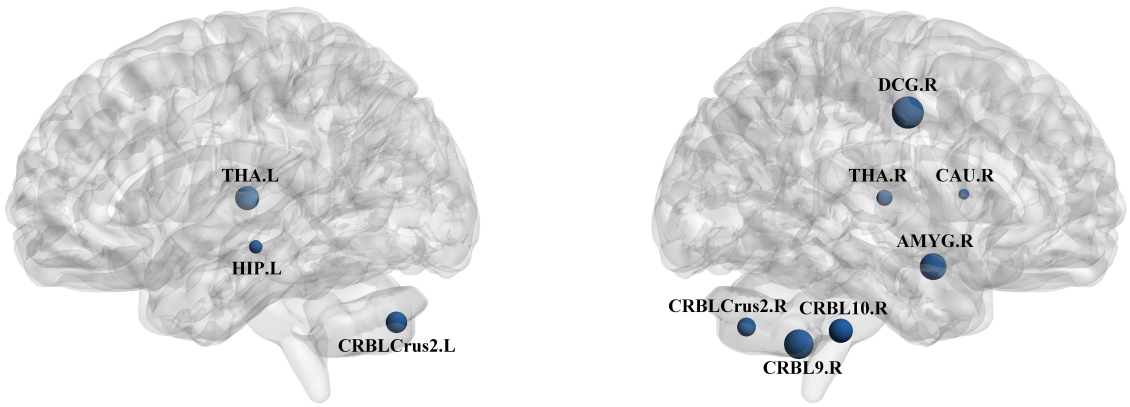

Figure 6. Visualization results for the top 10 regions. Here, the larger the node diameter, the higher region feature score achieved.

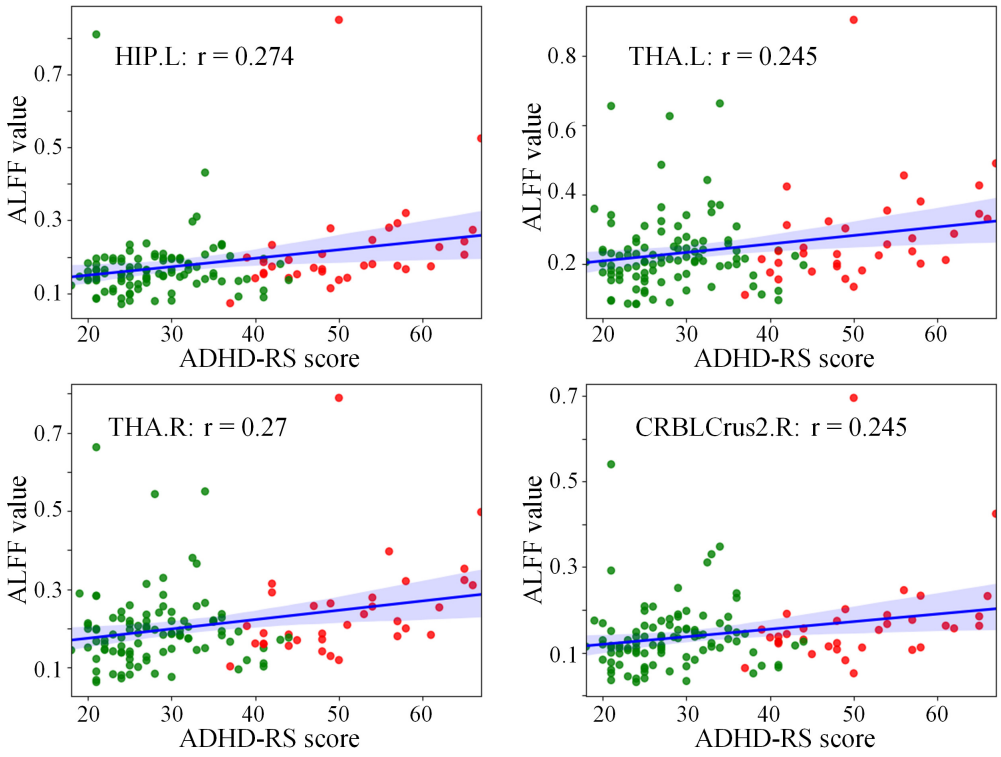

Figure 7. Relationship between ALFF value and symptom score for ADHD-RS on the PU dataset. ADHD and HC subjects are colored with red and green, respectively.

5. Biomarker discussion

We shall present some biological explanations for our biomarkers. Table 7 presents the findings, highlighting significant associations within regions such as the amygdala, caudate nucleus, hippocampus, and thalamus gyri. These regions are well-established in their role in shaping and comprehending human emotions, as extensively documented in the existing literature [48, 49]. Notably, ADHD patients have consistently exhibited reduced amygdala and hippocampal volumes in anatomical control experiments. Moreover, developmental delays and degenerative changes within these regions have been unveiled through lifespan exploratory modeling [10]. Concurrently, the caudate nucleus has emerged as a consistent focal point, with volumetric differences consistently noted in ADHD research [50–52]. Particularly, a volume asymmetry analysis demonstrated noteworthy links between caudate asymmetry and cumulative severity ratings of inattentive behaviors in ADHD-afflicted children [24]. As we delve into the thalamus, its pivotal role in information transmission, regulation, and participation in cognitive and behavioral processes such as attention, emotion, and motor control cannot be understated. Morphological aberrations in the thalamus [23] have been discovered and strongly correlated with ADHD symptom scores [53], further substantiating our identified biomarkers within the limbic system.

Furthermore, our biomarkers extend to the cerebellum, particularly in the right inferior posterior

lobe of the cerebellar hemisphere, encompassing cerebellum 9, 10, and Crus II. A longitudinal case-control study revealed that ADHD participants with worse clinical outcomes exhibited a gradual decrease in overall cerebellar volume, primarily caused by abnormalities in the posterior inferior cerebellar hemisphere [54]. Moreover, a smaller volume in the cerebellar lobule X has been found in children [55, 56]. Our findings align harmoniously with these reports. The lateral hemisphere of the cerebellum has emerged as a pivotal player in executive functions [57, 58], spatial cognition, and language processing [59]. Consequently, the identified abnormalities within the posterior inferior cerebellar hemisphere may indeed contribute to the executive function deficits that constitute a prominent symptomatology of ADHD.

Overall, the aforementioned evidence firmly supports the presence of our delineated ADHD biomarkers within the limbic system and cerebellum. Here, the right middle cingulate gyrus is also established as a biomarker, but there is also a consensus that it is part of the salience network. Furthermore, when it comes to the cognitive function and emotional processing of ADHD patients, the middle cingulate gyrus is considered to play an important role [60, 61]. Some existing reports have confirmed that the right middle cingulate gyrus is abnormal in ADHD patients. For example, a sex difference study on the factors related to the symptoms of common mental disorders in adolescence found that more prominent symptoms of hyperactivity/inattention were associated with lower grey matter volume of the bilateral anterior and midcingulum among boys [62]. Another study showed that compared with normal children, the increased degree centrality values for the right middle cingulate gyrus indicated differences in functional network connectivity for ADHD children [63].

6. Limitations

Our research has two limitations. First of all, biomarker detection only involves the linear relationship between features and does not take into account the nonlinear components. Our attempt to validate the discrimination of these biomarkers through a two-sample t-test experiment did not yield statistically significant results. This indicates no inter-group differences in the ALFF features of a single brain region. However, the aggregate of brain region features still contributes to the construction of excellent high-level feature space, which is due to the powerful feature extraction ability of DL. Unfortunately, the contribution here is difficult to quantify, as it involves the amount of classification information held by the brain region features and the details of non-linear fitting performed by the neural network. This issue limits further exploration of biomarkers; thus, in the future, we will attempt to explore this limitation by using other brain region features such as the voxel-based morphometry and ReHo, as well as introduce the interpretability theory of neural networks. In addition, our algorithm achieved good classification performance on limited brain regions of the limbic system and cerebellum, indicating the importance of these brain regions for ADHD. Moreover, we conducted biomarker detection, which revealed abnormal brain regions that are consistent with previously reported anatomical abnormalities. However, it may be insufficient to explore brain regions related to the biological mechanisms of ADHD solely through ALFF data. In future studies, we will consider multi-modal features for to characterize brain regions.

7. Conclusions

This study primarily combined existing statistical prior knowledge and pattern recognition methods to elucidate critical ADHD biomarkers. Specifically, the ALFF in the limbic system and cerebellum, which is highly related to ADHD in statistical analysis, was used as input for classification, contributing to reduce the range of biomarker detection. The BHT framework and ensemble learning methods were employed to ensure high accuracy in the classification results. Consequently, we performed a highly credible biomarker detection task that achieved an average accuracy of 93% on the ADHD-200 datasets. Several brain regions such as the thalamus, hippocampus, amygdala, and cerebellum IX were extracted from the results of the SVM-RFE algorithm as biomarkers. We validated them by analyzing the correlation among symptom scores. Moreover, these findings extend previous findings and align with the existing reports on the neurobiological contributions to ADHD, which demonstrates the effectiveness of our method.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Author Contributions

Ying Chen: Conceptualization, Methodology, Writing–original draft; Lele Wang: Conceptualization, Software, Writing–original draft; Zhixin Li: Validation, Writing–review & editing; Yibin Tang: Software, Methodology, Writing–original draft; Zhan Huan: Supervision.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grants: 62201093.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. G. Polanczyk, P. Jensen, Epidemiologic considerations in attention deficit hyperactivity disorder: A review and update, *Child Adolesc. Psychiatr. Clin. N. Am.*, **17** (2008), 245–260. https://doi.org/10.1016/j.chc.2007.11.006
2. Z. Zhang, G. Li, Y. Xu, X. Tang, Application of artificial intelligence in the MRI classification task of human brain neurological and psychiatric diseases: A scoping review, *Diagnostics*, **11** (2021), 1402. https://doi.org/10.3390/diagnostics11081402

3. M. Quaak, L. Mortel, R. M. Thomas, G. V. Wingen, Deep learning applications for the classification of psychiatric disorders using neuroimaging data: Systematic review and meta-analysis, *Neuroimage Clin.*, **30** (2021), 102584. https://doi.org/10.1016/j.nicl.2021.102584
4. L. Zou, J. Zheng, C. Miao, M. J. Mckeown, Z. J. Wang, 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI, *IEEE Access*, **5** (2017), 23626–23636. https://doi.org/10.1109/ACCESS.2017.2762703
5. L. Su, S. I. Kamata, ADHD classification with low-frequency fluctuation feature map based on 3D CBAMe, in *Proceedings of the 7th International Conference on Biomedical Signal and Image Processing*, ACM, (2022), 74–79. https://doi.org/10.1145/3563737.3563749
6. M. Chen, H. Li, J. Wang, J. R. Dillman, N. A. Parikh, L. He, A multichannel deep neural network model analyzing multiscale functional brain connectome data for attention deficit hyperactivity disorder detection, *Radiol. Artif. Intell.*, **2** (2019), e190012. https://doi.org/10.1148/ryai.2019190012
7. Y. Tang, J. Sun, C. Wang, Y. Zhong, A. Jiang, G. Liu, et al., ADHD classification using auto-encoding neural network and binary hypothesis testing, *Artif. Intell. Med.*, **123** (2022), 102209. https://doi.org/10.1016/j.artmed.2021.102209
8. Y. Sun, L. Zhao, Z. Lan, X. Jia, S. Xue, Differentiating boys with ADHD from those with typical development based on whole-brain functional connections using a machine learning approach, *Neuropsychiatr. Dis. Treat.*, **16** (2020), 691–702. https://doi.org/10.2147/NDT.S239013
9. Y. Zang, Y. He, C. Zhu, Q. Cao, M. Sui, M. Liang, et al., Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI, *Brain Dev.*, **29** (2007), 83–91. https://doi.org/10.1016/j.braindev.2006.07.002
10. M. Hoogman, J. Bralten, D. P. Hibar, M. Mennes, M. P. Zwiers, L. S. J. Schweren, et al., Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: A cross-sectional mega-analysis, *Lancet Psychiatry*, **4** (2017), 310–319. https://doi.org/10.1016/S2215-0366(17)30049-4
11. L. Liu, S. Tang, F. X. Wu, Y. P. Wang, J. Wang, An ensemble hybrid feature selection method for neuropsychiatric disorder classification, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **19** (2021), 1459–1471. https://doi.org/10.1109/TCBB.2021.3053181
12. J. B. Colby, J. D. Rudie, J. A. Brown, P. K. Douglas, M. S. Cohen, Z. Shehzad, Insights into multimodal imaging classification of ADHD, *Front. Syst. Neurosci.*, **6** (2012), 59. https://doi.org/10.3389/fnsys.2012.00059
13. M. Wang, B. Jie, W. Bian, X. Ding, W. Zhou, Z. Wang, et al., Graph-kernel based structured feature selection for brain disease classification using functional connectivity networks, *IEEE Access*, **7** (2019), 35001–35011. https://doi.org/10.1109/ACCESS.2019.2903332
14. Y. Zhao, H. Chen, R. T. Ogden, Wavelet-based weighted LASSO and screening approaches in functional linear regression, *J. Comput. Graphical Stat.*, **24** (2015), 655–675. https://doi.org/10.1080/10618600.2014.925458
15. M. Nunez-Garcia, S. Simpraga, M. A. Jurado, M. Garolera, R. Pueyo, L. Igual, FADR: Functional-anatomical discriminative regions for rest fMRI characterization, in *Machine Learning in Medical Imaging*, Springer, (2015), 61–68. https://doi.org/10.1007/978-3-319-24888-2_8

16. H. W. Loh, C. P. Ooi, P. D. Barua, E. E. Palmer, F. Molinari, U. R. Acharya, Automated detection of ADHD: Current trends and future perspective, *Comput. Biol. Med.*, **146** (2022), 105525. https://doi.org/10.1016/j.compbiomed.2022.105525

17. Z. Mao, Y. Su, G. Xu, X. Wang, Y. Huang, W. Yue, et al., Spatio-temporal deep learning method for ADHD fMRI classification, *Inf. Sci.*, **499** (2019), 1–11. https://doi.org/10.1016/j.ins.2019.05.043

18. T. Zhang, C. Li, P. Li, Y. Peng, X. Kang, C. Jiang, et al., Separated channel attention convolutional neural network (SC-CNN-attention) to identify ADHD in multi-site rs-fMRI dataset, *Entropy*, **22** (2020), 893. https://doi.org/10.3390/e22080893

19. Z. Wang, Y. Zhu, H. Shi, Y. Zhang, C. Yan, A 3D multiscale view convolutional neural network with attention for mental disease diagnosis on MRI images, *Math. Biosci. Eng.*, **18** (2021), 6978–6994. https://doi.org/10.3934/mbe.2021347

20. N. Qiang, Q. Dong, F. Ge, H. Liang, B. Ge, S. Zhang, et al., Deep variational autoencoder for mapping functional brain networks, *IEEE Trans. Cognit. Dev. Syst.*, **13** (2020), 841–852. https://doi.org/10.1109/TCDS.2020.3025137

21. N. Qiang, Q. Dong, H. Liang, B. Ge, S. Zhang, C. Zhang, et al., A novel ADHD classification method based on resting state temporal templates (RSTT) using spatiotemporal attention autoencoder, *Neural. Comput. Appl.*, **34** (2022), 7815–7833. https://doi.org/10.1007/s00521-021-06868-w

22. A. Gyurak, M. S. Goodkind, J. H. Kramer, B. L. Miller, R. W. Levenson, Executive functions and the down-regulation and up-regulation of emotion, *Cognit. Emotion*, **26** (2012), 103–118. https://doi.org/10.1080/02699931.2011.557291

23. C. Fu, S. Chen, A. Qian, R. Zhou, J. Zhou, J. Li, et al., Larger thalamus correlated with inattentive severity in the inattentive subtype of ADHD without comorbidity, *Psychiatry Res.*, **304** (2021), 114079. https://doi.org/10.1016/j.psychres.2021.114079

24. G. W. Schrimsher, R. L. Billingsley, E. F. Jackson, B. D. Moore, Caudate nucleus volume asymmetry predicts attention-deficit hyperactivity disorder (ADHD) symptomatology in children, *J. Child Neurol.*, **17** (2002), 877–884. https://doi.org/10.1177/08830738020170122001

25. T. Frodl, J. Stauber, N. Schaaff, N. Koutsouleris, J. Scheuerecker, M. Ewers, et al., Amygdala reduction in patients with ADHD compared with major depression and healthy volunteers, *Acta Psychiatr. Scand.*, **121** (2010), 111–118. https://doi.org/10.1111/j.1600-0447.2009.01489.x

26. K. Nickel, L. T. Elst, E. Perlov, R. Jitten-Schachenmeier, D. Beier, D. Endres, et al., Manual morphometry of hippocampus and amygdala in adults with attention-deficit hyperactivity disorder, *Psychiatry Res. Neuroimaging*, **267** (2017), 32–35. https://doi.org/10.1016/j.pscychresns.2017.07.001

27. K. J. Plessen, R. Bansal, H. Zhu, R. Whiteman, J. Amat, G. A. Quackenbush, et al., Hippocampus and amygdala morphology in attention-deficit/hyperactivity disorder, *Arch. Gen. Psychiatry*, **63** (2006), 795–807. https://doi.org/10.1001/archpsyc.63.7.795

28. Y. Wang, Q. Xu, S. Li, G. Li, C. Zuo, S. Liao, et al., Gender differences in anomalous subcortical morphology for children with ADHD, *Neurosci. Lett.*, **665** (2018), 176–181. https://doi.org/10.1016/j.neulet.2017.12.006

29. S. B. Hong, Thalamocortical functional connectivity in youth with attention-deficit/hyperactivity disorder, *J. Psychiatry Neurosci.*, **48** (2023), E50–E60. https://doi.org/10.1503/jpn.220109

30. J. Posner, F. Siciliano, Z. Wang, J. Liu, E. Sonuga-Barke, L. Greenhill, A multimodal MRI study of the hippocampus in medication-naive children with ADHD: What connects ADHD and depression, *Psychiatry Res. Neuroimaging*, **224** (2014), 112–118. https://doi.org/10.1016/j.pscychresns.2014.08.006

31. M. M. Bruchhage, M. P. Bucci, E. B. Becker, Cerebellar involvement in autism and ADHD, *Handb. Clin. Neurol.*, **155** (2018), 61–72. https://doi.org/10.1016/B978-0-444-64189-2.00004-4

32. M. V. Cherkasova, L. Hechtman, Neuroimaging in attention-deficit hyperactivity disorder: Beyond the frontostriatal circuitry, *Can. J. Psychiatry*, **54** (2009), 651–664. https://doi.org/10.1177/070674370905401002

33. G. Bush, Attention-deficit/hyperactivity disorder and attention networks, *Neuropsychopharmacology*, **35** (2010), 278–300. https://doi.org/10.1038/npp.2009.120

34. L. A. Friedman, J. L. Rapoport, Brain development in ADHD, *Curr. Opin. Neurobiol.*, **30** (2015), 106–111. https://doi.org/10.1016/j.conb.2014.11.007

35. Y. Chen, Y. Tang, C. Wang, X. Liu, L. Zhao, Z. Wang, ADHD classification by dual subspace learning using resting-state functional connectivity, *Artif. Intell. Med.*, **103** (2020), 101786. https://doi.org/10.1016/j.artmed.2019.101786

36. Y. Tang, X. Li, Y. Chen, Y. Zhong, A. Jiang, C. Wang, High-accuracy classification of attention deficit hyperactivity disorder with $l_{2,1}$-norm linear discriminant analysis and binary hypothesis testing, *IEEE Access*, **8** (2020), 56228–56237. https://doi.org/10.1109/ACCESS.2020.2982401

37. Y. Tang, C. Wang, Y. Chen, N. Sun, A. Jiang, Z. Wang, Identifying ADHD individuals from resting-state functional connectivity using subspace clustering and binary hypothesis testing, *J. Atten. Disord.*, **25** (2021), 736–748. https://doi.org/10.1177/1087054719837749

38. N. Liu, X. Li, E. Qi, M. Xu, L. Li, B. Gao, A novel ensemble learning paradigm for medical diagnosis with imbalanced data, *IEEE Access*, **8** (2020), 171263–171280. https://doi.org/10.1109/ACCESS.2020.3014362

39. Q. Yuan, K. Chen, Y. Yu, N. Q. K. Le, M. C. H. Chua, Prediction of anticancer peptides based on an ensemble model of deep learning and machine learning using ordinal positional encoding, *Briefings Bioinf.*, **24** (2023), bbac630. https://doi.org/10.1093/bib/bbac630

40. S. Singh, N. Q. K. Le, C. Wang, Vf-pred: Predicting virulence factor using sequence alignment percentage and ensemble learning models, *Comput. Biol. Med.*, **168** (2024), 107662. https://doi.org/10.1016/j.compbiomed.2023.107662

41. Y. Qin, Y. Lou, Y. Huang, R. Chen, W. Yue, An ensemble deep learning approach combining phenotypic data and fMRI for ADHD diagnosis, *J. Signal Process. Syst.*, **94** (2022), 1269–1281. https://doi.org/10.1007/s11265-022-01812-0

42. B. Miao, L. L. Zhang, J. L. Guan, Q. F. Meng, Y. L. Zhang, Classification of ADHD individuals and neurotypicals using reliable RELIEF: A resting-state study, *IEEE Access*, **7** (2019), 62163–62171. https://doi.org/10.1109/ACCESS.2019.2915988

43. L. Shao, Y. Xu, D. Fu, Classification of ADHD with bi-objective optimization, *J. Biomed. Inf.*, **84** (2018), 164–170. https://doi.org/10.1016/j.jbi.2018.07.011

44. A. Riaz, M. Asad, E. Alonso, G. Slabaugh, Fusion of fMRI and non-imaging data for ADHD classification, *Comput. Med. Imaging Graphics*, **65** (2018), 115–128. https://doi.org/10.1016/j.compmedimag.2017.10.002

45. A. Riaz, M. Asad, E. Alonso, G. Slabaugh, Deepfmri: End-to-end deep learning for functional connectivity and classification of ADHD using fmri, *J. Neurosci. Methods*, **335** (2020), 108506. https://doi.org/10.1016/j.jneumeth.2019.108506

46. S. Liu, L. Zhao, X. Wang, Q. Xin, J. Zhao, D. S. Guttery, et al., Deep spatio-temporal representation and ensemble classification for attention deficit/hyperactivity disorder, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **29** (2020), 1–10. https://doi.org/10.1109/TNSRE.2020.3019063

47. S. Pei, C. Wang, S. Cao, Z. Lv, Data augmentation for fmri-based functional connectivity and its application to cross-site adhd classification, *IEEE Trans. Instrum. Meas.*, **72** (2022), 1–15. https://doi.org/10.1109/TIM.2022.3232670

48. H. Suzuki, K. N. Botteron, J. L. Luby, A. C. Belden, M. S. Gaffrey, C. M. Babb, et al., Structural-functional correlations between hippocampal volume and cortico-limbic emotional responses in depressed children, *Cogn. Affect. Behav. Neurosci.*, **13** (2013), 135–151. https://doi.org/10.3758/s13415-012-0121-y

49. V. Vuontela, S. Carlson, A. M. Troberg, T. Fontell, P. Simola, S. Saarinen, et al., Working memory, attention, inhibition, and their relation to adaptive functioning and behavioral/emotional symptoms in school-aged children, *Child Psychiatry Hum. Dev.*, **44** (2013), 105–122. https://doi.org/10.1007/s10578-012-0313-2

50. E. M. Valera, S. V. Faraone, K. E. Murray, L. J. Seidman, Meta-analysis of structural imaging findings in attention-deficit/hyperactivity disorder, *Biol. Psychiatry*, **61** (2007), 1361–1369. https://doi.org/10.1016/j.biopsych.2006.06.011

51. T. Frodl, N. Skokauskas, Meta-analysis of structural MRI studies in children and adults with attention deficit hyperactivity disorder indicates treatment effects, *Acta Psychiatr. Scand.*, **125** (2012), 114–126. https://doi.org/10.1111/j.1600-0447.2011.01786.x

52. T. Nakao, J. Radua, K. Rubia, D. Mataix-Cols, Gray matter volume abnormalities in ADHD: Voxel-based meta-analysis exploring the effects of age and stimulant medication, *Am. J. Psychiatry*, **168** (2011), 1154–1163. https://doi.org/10.1176/appi.ajp.2011.11020281

53. I. Ivanov, R. Bansal, X. Hao, H. Zhu, C. Kellendonk, L. Miller, et al., Morphological abnormalities of the thalamus in youths with attention deficit hyperactivity disorder, *Am. J. Psychiatry*, **167** (2010), 397–408. https://doi.org/10.1176/appi.ajp.2009.09030398

54. S. Mackie, P. Shaw, R. Lenroot, R. Pierson, D. K. Greenstein, T. F. Nugent, et al., Cerebellar development and clinical outcome in attention deficit hyperactivity disorder, *Am. J. Psychiatry*, **164** (2007), 647–655. https://doi.org/10.1176/appi.ajp.164.4.647

55. F. X. Castellanos, J. N. Giedd, P. C. Berquin, J. M. Walter, W. Sharp, T. Tran, et al., Quantitative brain magnetic resonance imaging in girls with attention-deficit/hyperactivity disorder, *Arch. Gen. Psychiatry*, **58** (2001), 289–295. https://doi.org/10.1001/archpsyc.58.3.289

56. S. H. Mostofsky, A. L. Reiss, P. Lockhart, M. B. Denckla, Evaluation of cerebellar size in attention-deficit hyperactivity disorder, *J. Child Neurol.*, **13** (1998), 434–439. https://doi.org/10.1177/088307389801300904

57. J. D. Schmahmann, Disorders of the cerebellum: Ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome, *J. Neuropsychiatry Clin. Neurosci.*, **16** (2004), 367–378. https://doi.org/10.1176/appi.neuropsych.16.3.367

58. K. P. Schulz, J. Fan, C. Y. Tang, J. H. Newcorn, M. S. Buchsbaum, A. M. Cheung, et al., Response inhibition in adolescents diagnosed with attention deficit hyperactivity disorder during childhood: An event-related FMRI study, *Am. J. Psychiatry*, **161** (2004), 1650–1657. https://doi.org/10.1176/appi.ajp.161.9.1650

59. D. Riva, C. Giorgi, The cerebellum contributes to higher functions during development: Evidence from a series of children surgically treated for posterior fossa tumours, *Brain*, **123** (2000), 1051–1061. https://doi.org/10.1093/brain/123.5.1051

60. D. Dong, Q. Ming, X. Wang, W. Yu, Y. Jiang, Q. Wu, et al., Temporoparietal junction hypoactivity during pain-related empathy processing in adolescents with conduct disorder, *Front. Psychol.*, **7** (2017), 231676. https://doi.org/10.3389/fpsyg.2016.02085

61. G. Bush, Cingulate, frontal, and parietal cortical dysfunction in attention-deficit/hyperactivity disorder, *Biol. Psychiatryry*, **69** (2011), 1160–1167. https://doi.org/10.1016/j.biopsych.2011.01.022

62. F. Biondo, C. N. Thunell, B. Xu, C. Chu, T. Jia, A. Ing, et al., Sex differences in neural correlates of common psychopathological symptoms in early adolescence, *Psychol. Med.*, **52** (2022), 3086–3096. https://doi.org/10.1017/S0033291720005140

63. K. Jiang, Y. Yi, L. Li, H. Li, H. Shen, F. Zhao, et al., Functional network connectivity changes in children with attention-deficit hyperactivity disorder: A resting-state fMRI study, *Int. J. Dev. Neurosci.*, **78** (2019), 1–6. https://doi.org/10.1016/j.ijdevneu.2019.07.003

AIMS Press