Multi-layer convolutional autoencoder for recognizing three-dimensional patterns in attention deficit hyperactivity disorder using resting-state functional magnetic resonance imaging

Zarina Begum, Kareemulla Shaik
School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

Article Info	ABSTRACT
	Attention deficit hyperactivity disorder (ADHD) is a neurological disorder that develops over time and is typified by impulsivity, hyperactivity, and attention deficiency. There have been noticeable changes in the patterns of brain activity in recent studies using functional magnetic resonance imaging (fMRI). Particularly in the prefrontal cortex. Machine learning algorithms show promise in distinguishing ADHD subtypes based on these neurobiological signatures. However, the inherent heterogeneity of ADHD complicates consistent classification, while small sample sizes limit the generalizability of findings. Additionally, methodological variability across studies contributes to inconsistent results, and the opaque nature of machine learning models hinders the understanding of underlying mechanisms. We suggest a novel deep learning architecture to overcome these issues by combining spatio-temporal feature extraction and classification through a hierarchical residual convolutional noise reduction autoencoder (HRCNRAE) and a 3D convolutional gated memory unit (GMU). This framework effectively reduces spatial dimensions, captures key temporal and spatial features, and utilizes a sigmoid classifier for robust binary classification. Our methodology was rigorously validated on the ADHD-200 dataset across five sites, demonstrating enhancements in diagnostic accuracy ranging from 1.26% to 9.6% compared to existing models. Importantly, this research represents the first application of a 3D Convolutional GMU for diagnosing ADHD with fMRI data. The improvements highlight the efficacy of our architecture in capturing complex spatio-temporal features, paving the way for more accurate and reliable ADHD diagnoses.

Corresponding Author:

Kareemulla Shaik
School of Computer Science and Engineering, VIT-AP University
Andhra Pradesh 522237, India
Email: kareemulla.shaik@vitap.ac.in

1. INTRODUCTION

One common neurodevelopmental condition that affects children is attention deficit hyperactivity disorder (ADHD), with a significant proportion of cases persisting into adulthood. Research indicates that approximately 65% of individuals diagnosed in childhood continue to experience symptoms into adulthood, resulting in substantial economic and psychological ramifications for both patients and their families [1]. Clinical judgment is the primary method of diagnosis, which introduces subjectivity and variability into evaluations. Timely intervention depends on early identification, but the diagnostic statistical manual interpretive flexibility makes the diagnostic procedure difficult [2]. Although a thorough assessment by qualified experts is required, inconsistencies are exacerbated by disparities in training. Like autism spectrum

disorder, ADHD has recognizable patterns of brain activity that facilitate diagnosis. New developments in machine learning (ML) and medical imaging present encouraging paths for more objective diagnosis. Brain data is being analyzed using methods including structural magnetic resonance imaging (sMRI) and functional magnetic resonance imaging (fMRI), which improve diagnostic accuracy [3]. For instance, Yilin *et al.* [4] used graph convolutional networks (GCN) to examine functional connectivity data, which improved the classification of ADHD by emphasizing the distinctions between peers who are usually developing and those who have ADHD. A similar model, attention-based spatial temporal network, was presented by Qiu *et al.* [5] using resting-state fMRI data with time division multiplexing to capture temporal fluctuations and adaptive functional connectivity generation (AFCG) for spatial correlation analysis. Through the integration of spatial and temporal data, adaptive spatial-temporal neural network (ASTNet) surpasses earlier approaches, opening the door to more precise and impartial diagnosis of ADHD.

Recent advancements in ML and deep learning (DL) have significantly enhanced the diagnosis of ADHD and other psychiatric disorders through the integration of neuroimaging data. For example, Hatami *et al.* [6] showed how to improve the diagnosis of major depressive disorder by using fMRI data in conjunction with the MobileNet V2 model and the data processing and analysis for brain imaging toolbox. Liu *et al.* [7] developed the multimodal generative fusion framework, integrating fMRI and sMRI data through multi-task learning to generate paired data, improving diagnostic accuracy. Alsharif *et al.* [8] developed an ML-based decision system that achieved 91% accuracy on standard ADHD datasets, lowering subjectivity in traditional evaluations; and Agarwal *et al.* [9] used a dual approach, combining image-based DL models and graph-based networks to analyze fMRI connectivity matrices, showing that different brain atlas and connection matrix selections improve classification accuracy. In contrast to conventional ReHo techniques, Gülhan and Özmen [10] shown that maintaining spatial information improves classification by using 3D convolutional neural networks (CNNs) to evaluate fractional amplitude of low frequency fluctuation data. In order to improve brain activity investigation in ADHD cases, Saurabh and Gupta [11] reshaped 4D pictures and utilized a modified bidirectional long short-term memory (BLSTM) model to interpret resting-state fMRI data. In order to get high computational efficiency, Salah *et al.* [12] combined fMRI and optical amplification data in their residual learning layer-based ADHD screening model. In an effort to create "explainable AI," Amado-Caballero *et al.* [13] evaluated demographic parameters influencing ADHD diagnosis using CNN visualization techniques including occlusion maps. Using a transformer-based model using ADHD-200 fMRI data, Qin *et al.* [14] combined phenotypic and fMRI data to obtain 74.5% classification accuracy, surpassing earlier methods. To facilitate effective feature extraction, a 4D CNN model was also applied, capturing both spatial and temporal dimensions with an observed accuracy of 71.3% [15]. Around 800 people from various universities contributed sMRI and resting-state fMRI data to the ADHD-200 consortium, which was created by the INDI in 2011. This allowed researchers worldwide to improve ADHD classification algorithms by using consistent datasets [16]. This information made it easier to create sophisticated models. For example, Mao *et al.* [15] achieved improved classification accuracy by combining 3D CNNs for spatial extraction with an long short-term memory (LSTM) model to capture temporal relationships. In order to evaluate demographic aspects influencing ADHD diagnosis, Liu *et al.* [17] investigated CNN visualization approaches, such as occlusion maps. This work contributed to the developing topic of "explainable AI," which improves the interpretability of DL models in clinical practice. Finally, efforts to enhance diagnostic accuracy have been exemplified by Qin *et al.* [12] who employed a transformer-based model using ADHD-200 fMRI images, combining fMRI and phenotypic data to achieve a classification accuracy of 74.5%, surpassing multiple advanced approaches. To facilitate effective feature extraction, a 4D CNN model was also applied, capturing both spatial and temporal dimensions with an observed accuracy of 71.3%. This study improves the detection of ADHD by introducing three major contributions: i) a convolutional gated memory unit (GMU) to capture spatial and temporal information; ii) a multi-layer convolutional denoising auto encoder network to recognize 3D spatial patterns in resting-state functional magnetic resonance imaging (rs-fMRI) data; and iii) a validated framework evaluated across various sites. The overview of the algorithm, its conceptual underpinnings, the experimental setting, and its consequences and future directions are covered in the paper.

2.   ALGORITHM

The algorithm for classifying ADHD from resting-state fMRI data extracts spatial features using an hierarchical residual convolutional noise reduction autoencoder (HRCNRAE), which is then strengthened by residual connections to preserve significant information. During training, the model reconstructs the input while measuring reconstruction loss, and a convolutional GMU records temporal dependency. Then, using global average pooling, people are categorized as either ADHD-positive or not, offering a quick and effective way to examine intricate neuroimaging data.

Algorithm 1: Spatio-temporal feature extraction and ADHD classification
Input:X_{input} :4D resting-state fMRI(rs-fMRI) data.
Step 1: Data Regularization
Add Gaussian noise for regularization:
$$Input\ X_{input} = X_{input} + N_{noise}$$
Step 2: Feature Encoding with HRCNRAE
    1.   Pass X_{noisy} through CP block 1: $F_1 = H_{cp}(X_{noisy})$
    2.   Process through CP block 2: $F_2 = H_{cp}(F_1)$
    3.   Process through CP block 3: $F3 = H_{cp}(F_2)$
Step 3: Incorporate Residual blocks
    Combine features: $F_{residual\_outer} = F_3 + H_{skip}(F_1, F_2)$
Step 4: Feature Decoding
    1.Decode 1: $D_1 = H_{ud}(F_{residual\_outer})$
    2.Decode 2: $D_2 = H_{ud}(D_1)$
    3.Final output: $Y_{reconstructed} = H_{ud}(D_2)$
Step 5: Compute Loss Function: calculate reconstruction loss
Step 6: Spation-Temporal Feature Extraction with GMU
    For each time step t
    1.Reset gate: $r_t = \Sigma(W_r . h_{t-1} + U_r .x_t)$
    2.Update gate: $z_t = \Sigma(W_z .h_{t-1} + U_z .x_t)$
    3.Candidate hidden state: $\tilde{h}_t = \tanh(W_h . (r_t \Theta h_{t-1}) + U_h . x_t)$
    4.Final hidden state: $h_t = (1-z_t)\Theta h_{t-1} + z_t \Theta \tilde{h}_t)$
Step 7: Classification:using Global Average Pooling $O=GAP(h_T)$
Output: Predicted ADHD label(1 for ADHD,0 for non-ADHD).

3.   PROPOSED METHOD

DL has been used extensively for a variety of applications because it is effective at extracting features from huge datasets. It is distinguished by its complexity and more accurate predictions compared to conventional machine learning methods. This research presents a network architecture that combines a convolutional GMU with an HRCNRAE to use rs-fMRI data to obtain joint spatiotemporal properties, which are then applied to the classification of ADHD. As illustrated in Figure 1, the encoder part of the auto encoder first extracts high-level spatial characteristics from the rs-fMRI. The auto encoder incorporates a residual network to further capture deeper spatial features. The convolutional GMU then organizes and processes these spatial characteristics in a temporal manner, capturing both spatial and temporal dynamics at the same time. The GMU-extracted characteristics are subjected to GAP [18] before being subjected to a sigmoid classifier for the ultimate ADHD classification.
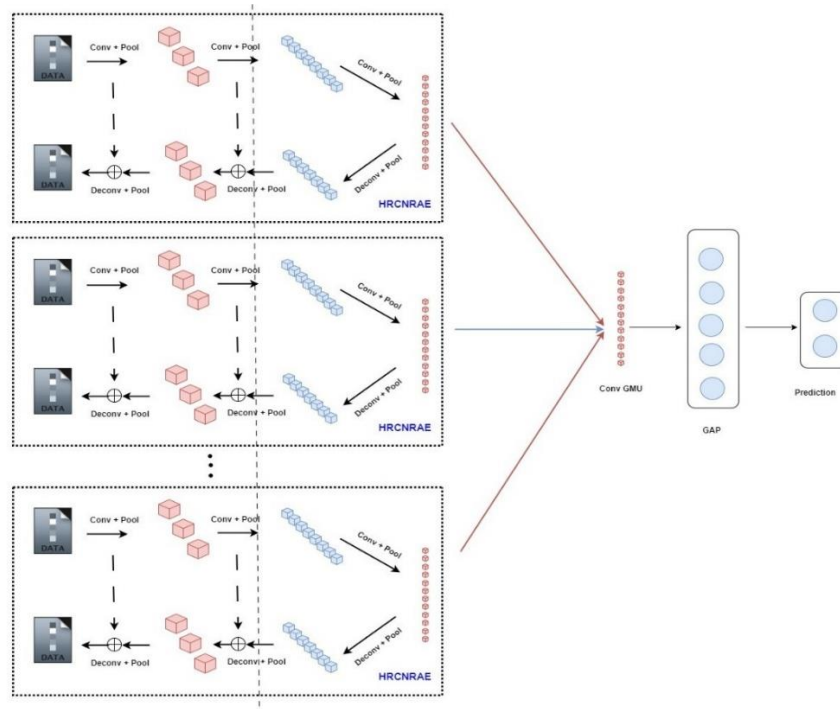

Figure 1. Flowchart of the proposed classification algorithm

3.1. Hierarchical residual convolutional noise reduction autoencoder

The HRCNRAE is a new method that combines residual networks with convolutional denoising autoencoders (CDAE) to extract spatial information from unlabeled rs-fMRI data [19]. Its main goal is to recover high-level spatial characteristics while lowering dimensionality so that the convolutional GMU can more easily extract spatiotemporal features. HRCNRAE uses skip connections to preserve high-level feature extraction while using the residual learning technique first presented by He *et al.* [20] to avoid the vanishing gradient issue in deep networks. The encoder, decoder, and residual blocks make up the HRCNRAE architecture, as seen in Figure 2. There are three CP segments with data filtering and reduction layers in the input processing unit, and three unpooling deconvolution (UD) segments with reverse filtering and data expansion in the output processing unit. An external bypass with two CP and two UD segments and an internal bypass with one CP and one UD segment is the two bypass architectures that guarantee effective data flow. The system performance is improved by these bypasses, which allow input and output units to be trained simultaneously.
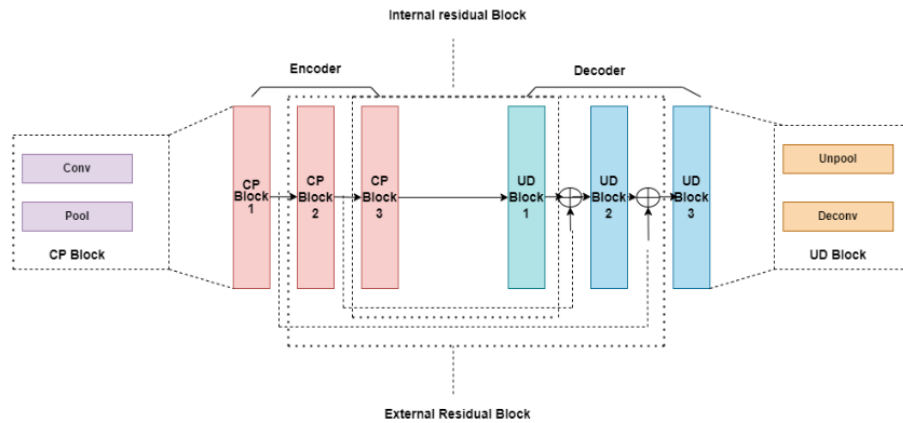

Figure 2. Organization of HRCNRAE

Regarding the provided rs-fMRI data:

$$x = [x_1 x_p](x_1 \in R^{60 \times 72 \times 60}) \tag{1}$$

where the time period, denoted by p, varies depending on the location.

The input of HRCNRAE is first obtained by adding random noise, which is regularized by following a conventional normal distribution. As a result, we can get

$$\overline{x} = [\overline{x_1 x_p}], (\overline{x_i} \in R^{60 \times 72 \times 60}) \tag{2}$$

To extract the enhanced pattern h, the input processing section of HRCNRAE transforms the original data into a condensed representation, which is

$$h = f(\overline{x}) \tag{3}$$

The interim pattern h is reconstructed into y through HRCNRAE's output processing section, which is

$$y = g(h) \tag{4}$$

Since the goal of HRCNRAE's training is to reduce the discrepancy between the cleaned and rebuilt data, the loss is described as (5).

$$loss = |y - x|^2 \tag{5}$$

The maximum value [21] that can be produced within the pooling window is the goal of the max pooling window. Let $s_{k,i} \in Q^{l \times m \times n}$ be the previous layer's i-th feature map, and let $s_{k-1,i} \in Q^{l \times m \times n}$ represent the i-th feature map in the k-th layer. The i-th feature map of the k-th layer's components can be calculated as (6).

$$s_{k,i}^{x_1,x_2,x_3} = max\left(Q_{k,i}^{x_1,x_2,x_3}\right) \tag{6}$$

To perform deconvolution, the deconvolution kernel slides across the features, whereas the unpooling process zero-fills the latent features to return the feature map to its original picture size [22]. Three CP blocks make up the encoder, and the CP block's i-th output is

$$F_i = H_{cp} (F_{i-1}) \ i=1, 2, 3, \tag{7}$$

where specific composite functions are used, such pooling, rectified linear unit (ReLU), and convolution, are represented by HCP [23]. For i = 1, Fi-1 represents the input noise image I, and the final result resulting from the subsequent CP block for all other values. The skip-connection's output is

$$S_{1\ or\ E} = ReLU\left(\sum F_i * w_j + b_j\right) \tag{8}$$

The internal residual block is represented when S_I or $_E$ = I and i = 2, and the outer residual block is represented where i = 1 and SI or E = E. The i-th convolution kernel is indicated by ω_j. B_j is a representation of the bias. The ReLU activation adds a tiny degree of sparsity to the trained network when j is the number of channels, which is shown as (9).

$$\mathrm{Re}\,LU\left(x\right) = \max\left(0, x\right) \tag{9}$$

Three symmetric UD blocks make up the decoder component. The result of the first UD block is displayed as (10):

$$G_0 = H_{UD}\left(F_3\right) \tag{10}$$

The following is the outcome of the i-th UD block:

$$G_i = H_{UD}\left(G_{i-1} + S\right) i = 2, 3 \tag{11}$$

where HUD is for the composite operation that combines sampling, deconvolution, and ReLU, and S stands for the output from the shortcut convolution layer. It should be noted that the deconvolution up sampling blocks belong to the decoder, whereas the convolution pooling blocks indicated above are part of the encoder. Two nested residual blocks in the residual structure connect the encoder and decoder. In the beginning, the internal residual block creates a skip-connection between the encoder and decoder, which is established by (12)

$$R_1 = S_1 + G_1 \tag{12}$$

The outer residual block's output is shown as (13):

$$R_E = S_E + G_2 \tag{13}$$

3.2. Gated convolutional memory unit

The GMU is a sophisticated, parameter-efficient variant of the traditional recurrent neural network (RNN) [24] that successfully resolves long-sequence dependence and gradient vanishing problems. In this study, spatial information must be incorporated in order to extract spatiotemporal characteristics from rs-fMRI data using convolutional GMU. Convolutional GMU processes spatial and temporal features at the same time, in contrast to conventional GMU [25]. The information flow is managed by means of reset and update gates, where the reset gate eliminates some previous data and the update gate regulates the amount of historical data that is kept. Both gates to process the preceding output (h_{t-1}) at time t and the current input (x_t) use a sigmoid activation function. The update gate (z_t) regulates the amount of the prior data that is retained, while the reset gate (r_t) establishes how much of it is considered. Convolution with x_t multiplies r_t by h_{t-1} to yield the candidate hidden state $\tilde{h}_t$. The previous hidden state and the candidate-hidden state are combined to determine the final hidden state h_t.

$$r_t = \sigma\left(W_r * [x_t, h_{t-1}]\right) \tag{14}$$

$$Z_t = \sigma\left(W_z * [x_t, h_{t-1}]\right) \qquad (15)$$

$$\overline{h_t} \tanh\left(W * [x_t, r_t \Theta h_{t-1}]\right) \qquad (16)$$

The convolution operation is represented by *, the Hadamard product by $\odot$, and the sigmoid activation function by σ. The weights that correspond to r_t, z_t, and $\tilde{h}_t$ are W_r, W_z, and W.

3.3. Experimental setup
3.3.1. Data description

        The developed method is analyzed and processed using the Python simulation platform. The publicly available ADHD-200 database, which contains fMRI images, is collected and examined for this study in order to carry out the experiment. dataset comprised of imaging and phenotypic data (gender, age, IQ, handedness) from 973 people spread over eight worldwide sites. Of these, 585 had normally developing typical developing (TD) profiles, 26 had an unidentified diagnosis, and 362 had ADHD. Three sites—Brown, Pittsburgh, and Washington University—were disqualified because of missing diagnostic data, whereas data from five sites—Peking, Neuroimaging, KKI, OHSU, and NYU—were examined. The data processing assistant for resting-state fMRI (DPARSF) tools, which include noise reduction, spatial smoothing, normalization to MNI space, slice-timing correction, and head motion correction, were applied to pre-process rs-fMRI data. Images containing objects and participants who moved their heads excessively were disqualified. 93,650 frames were kept for training the HRCNRAE model as shown in Table 1 after pre-processing.

Table 1. Building data using the ADHD-200 dataset

		NYU	OHSU	Peking	N image	KKI	Total
Training set	ADHD	105	24	61	12	20	222
	Control	90	33	108	16	53	300
	Total	195	57	169	28	73	522
Testing set	ADHD	29	6	24	5	3	67
	Control	12	28	27	14	8	89
	Total	41	34	51	19	11	156

3.3.2. Training of models

        The encoder for the HRCNRAE model is made up of three CP blocks with convolution kernel sizes of 3×3×3, 3×3×3, and 2×2×2, and feature mappings set to 16, 32, and 96. The stride for all convolution layers is 1×1×1, however the pooling layers have the same padding but kernel sizes and strides of 2×2×2, 3×3×3, and 2×2×2. Its structure is mirrored in the decoder. The learning rate was set at 0.001 and decreased exponentially at 0.9 during the 220 epochs of training, which had a batch size of 50. To improve every network component, an overview of the model's complex design is shown in Table 2. The Adam optimizer was employed. A learning rate of 0.00001 and a batch size of 8 were used to refine the convolutional GMU. The fully connected layer was subjected to a dropout rate of 0.5 and L2 regularization (0.3) in order to avoid overfitting.

Table 2. The network's specifics

Layer type	Output size	Filter size, stride
Input layer	62×70×62	------------
CP Block 1	16, 30×36×30	3×3×3 Conv, stride 1
External residual block	16, 1×1×1	1×1×1 Conv, stride 1
CP Block 2	32, 10×12×10	2×2×2 Conv, stride 1
		3×3×3 max pool, stride 3
Internal residual block	32, 1×1×1	1×1×1 Conv, stride 1
CP Block 3	96, 5×6×5	3×3×3 Conv, stride 1
		2×2×2 max pool, stride 2
UD Block 1	96, 10×12×10	2×2×2 DeConv, stride 1
		2×2×2 Un Pool, stride 2
Internal residual block	32, 1×1×1	------------
UD Block 2	32, 30×36×30	3×3×3 DeConv, stride 1
		3×3×3 Un Pool, stride 3
External residual block	16, 1×1×1	------------
UD Block 3	16, 60×72×60	3×3×3 DeConv, stride 1
		2×2×2 Un pool, stride 2

4.   RESULTS AND DISCUSSION

We evaluated the proposed model's performance using the test set derived from the dataset ADHD-200. Since generalization ability is essential when using machine learning to diagnose disorders, the model's efficacy was evaluated using classification performance across several sites. The next sections present the findings from a variety of experimental testing.

4.1. Visualization results

To evaluate the model's performance, feature maps and convolution weights from CP blocks were examined. Figure 3 [26] presents 16 neural weights from the initial CP block, initialized using the Xavier method. The variation in grayscale shades highlights how the trained weights capture richer and more diverse feature representations. Figure 4 emphasizes the extraction of spatial information; and Figure 5, which demonstrates notable variations between people with ADHD and those with TD.

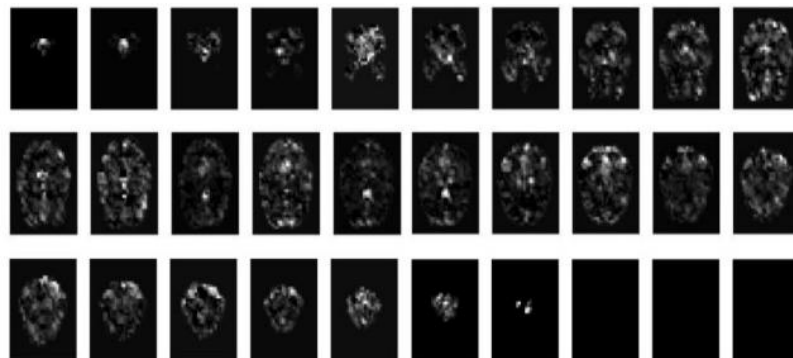Figure 3. Weights assigned to the convolution layer in the first CP block

Figure 4. The maps with features that were produced using the initial CP block

Figure 5. The first filter was used to create differential feature maps for the third block of CP

4.2. Choosing the regularization parameters

The convolutional GMU's accuracy under various regularization settings is displayed in Figure 6, underscoring the important influence that parameter choice has on classification outcomes. As the regularization parameter rises, accuracy increases and reaches a top of 0.3. But when the parameter goes beyond 0.3, performance suffers, most likely as a result of overfitting. Smaller numbers, on the other hand, result in excessive features, redundancy and lower performance.
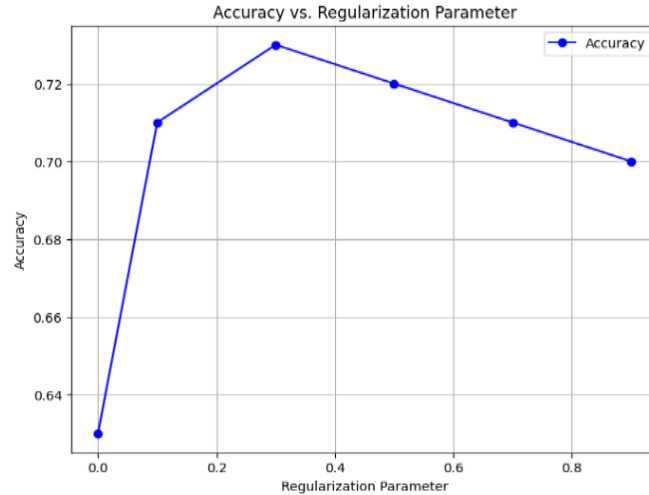
Figure 6. The precision associated with distinct regularization settings

4.3. Model ablation
4.3.1. The residual block

The HRCNRAE model shows improved classification performance compared to the convolutional denoising autoencoder (CDAE), largely due to the integration of a residual block. This residual block enhances feature preservation across layers, allowing the model to learn patterns that are more complex without losing important information. As shown in Figure 7, evaluation metrics such as the area under the curve (AUC) and receiver operating characteristic (ROC) clearly demonstrate the effectiveness of this architectural improvement.
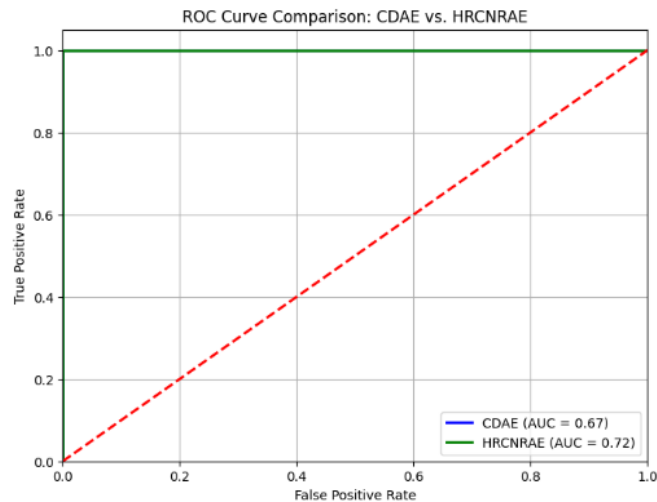

Figure 7. The contrast between HRCNRAE and CDAE

4.3.2. A set of CP units

We also ran two other schemes using two blocks and one block for contrast in order to get the ideal number of CP blocks. Table 3 displays the findings. The classification assessment indices' accuracy, sensitivity, specificity and are examples of metrics used in evaluation. The probability that the model would accurately categorize TD and ADHD is known as accuracy, and it is described as (17)

$$\text{Accuracy} = TP + TN/TP+TN+FP+FN \qquad (17)$$

The likelihood of correctly identifying ADHD is represented by sensitivity, which could be computed as (18)

$$Sensitivity = TP/TP+FN \qquad\qquad (18)$$

Specificity is a measure of the model's ability to accurately identify TD and is calculated as (19)

$$Specificity = TN/TN+FP \qquad\qquad (19)$$

The study used TP, TN, FP, and FN indicators to assess ADHD categorization performance. Using 10,000 bootstrapped samples, accuracy, sensitivity, and specificity were computed with 90% confidence intervals. Table 3 illustrates how additional CP blocks increased accuracy, with three blocks yielding 72.44% accuracy, 70.22% sensitivity, and 74.18% specificity. Computational constraints prevented testing of more blocks.

Table 3. Comparative analysis of many convolutions pooling blockshe network's specifics

S.No	Accuracy	Sensitivity	Specificity
1	69.39%	71.14%	70.88%
2	72.24%	70.42%	72.12%
3	72.44%	70.22%	74.18%

With its residual connections and hierarchical structure, HRCNRAE provides faster convergence, increased stability, and improved noise suppression, making it perfect for complicated workloads. The simpler and more effective CDAE, on the other hand, could have trouble with complex noise patterns and deeper architecture. Figure 7 demonstrates that HRCNRAE is more resilient than CDAE, with a higher AUC of 0.72 compared to a lower AUC of 0.67.

4.4. Choosing techniques for extracting spatiotemporal features
4.4.1. Contrast between convolutional LSTM and convolutional GMU

Using the same datasets, the efficacy of convolutional GMU and convolutional LSTM was evaluated. Table 4 demonstrates that convolutional GMU surpassed convolutional LSTM in classification accuracy and specificity, despite having a marginally lower sensitivity. GMU had superior overall performance and stability, presumably as a result of LSTM's complexity and larger number of variables, which make training more challenging. LSTM attained 72.14% accuracy, 72.24% specificity, and 68.44% sensitivity.

Table 4. Convolutional LSTM vs. convolutional GMU comparison

Methodology	Accuracy	Specificity	Sensitivity
Convolutional LSTM	68.44%	72.14%	72.24%
Convolutional GMU	72.44%	75.26%	70.15%

4.4.2 GMU and convolutional GMU comparison

The effectiveness of the convolution operation is assessed by comparing the efficiency of convolutional GMU with standard GMU in order to classify ADHD. The same testing set is used to evaluate both models, and the same training set is used to train them. Table 5 shows the experiment's results.

Table 5. Comparison of convolutional GMU and convolutional GMU

Methodology	Accuracy	Sensitivity	Specificity
GMU	69.87%	66.26%	70.81%
Convolutional GMU	73.22%	70.14%	73.14%

4.5. Comparison with other methods

The suggested algorithm for classifying ADHD was contrasted with five cutting-edge techniques: CDAE-AdaDT [23], MKL [27], MDS-SVM [28], 3D-CNN [19], and 4D-CNN [22]. In accuracy, specificity, and sensitivity, it outperformed MDS-SVM, 3D-CNN, MKL, and 4D-CNN, as seen in Table 6. However, in specificity, it performed somewhat worse than CDAE-AdaDT, most likely as a result of variations in generalization across different sites. Figure 8 shows a graphic comparison of these findings.

Table 6. Results of classification on the ADHD-200 dataset using different methods

Method	4D-CNN	MDS-SVM	3D-CNN	CDAE	MKL	Our method
Accuracy	61.54%	62.81%	69.15%	71.30%	72.64%	**74.24%**
Specificity	71.21%	68.12%	---	58.24%	70.28%	**78.26%**
Sensitivity	41.33%	27.27%	---	71.10%	69.24%	**71.15%**
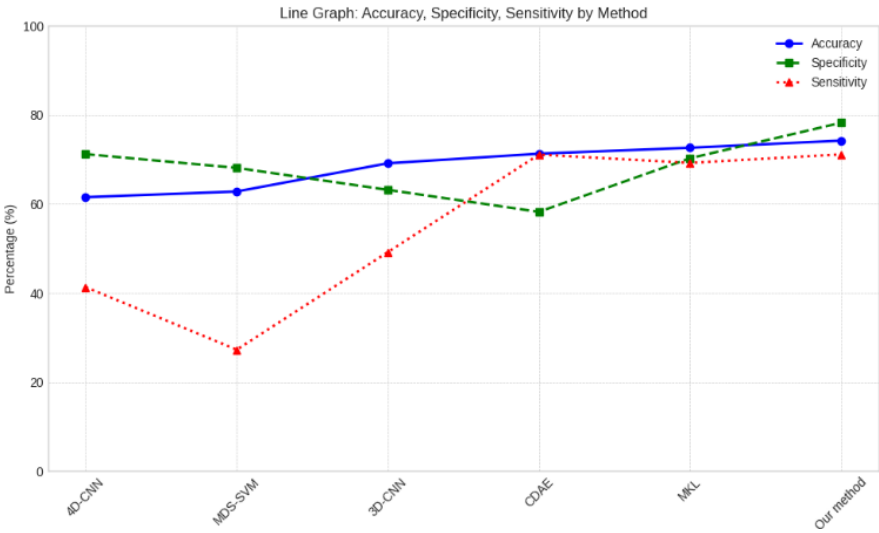
Figure 8. Comparing the results of various methods

4. CONCLUSION

The four-dimensional data recorded using functional magnetic resonance imaging in the resting state (rs-fMRI) incorporates one-dimensional temporal information alongside three-dimensional spatial details. Traditional methods often reduce this 4D data to 2D or 3D formats for categorization, it may cause a substantial loss of information. We provide a novel classification method to address this problem. That utilizes HRCNRAE and convolutional GMU, effectively preserving the comprehensive 4D structure of rs-fMRI images. Through the utilization of the temporal and spatial dynamics present in the data, our approach improves the model's capacity to precisely detect ADHD. The results of the experiments show that the suggested method functions well in cross-site classification tasks, greatly increasing the classification accuracy of ADHD. While this study focuses on ADHD, future research will aim to broaden the application of this methodology to other neurodevelopmental disorders, thereby enhancing its generalizability and potential impact within the field.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Zarina Begum	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Kareemulla Shaik	✓	✓				✓		✓		✓	✓	✓		

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, KS. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1]   J. C. Agnew-Blais, G. V Polanczyk, A. Danese, J. Wertz, T. E. Moffitt, and L. Arseneault, "Young adult mental health and functional outcomes among individuals with remitted, persistent and late-onset ADHD," *British Journal of Psychiatry*, vol. 213, no. 3, pp. 526–534, 2018.

[2]   American Psychiatric Association, "Diagnostic and statistical manual of mental disorders: DSM-5." American psychiatric association, 2013.

[3]   S. Cheekaty and G. Muneeswari, "Exploring sparse Gaussian processes for Bayesian optimization in convolutional neural networks for autism classification," *IEEE Access*, vol. 12, pp. 10631–10651, 2024, doi: 10.1109/ACCESS.2024.3351168.

[4]   Y. Hu *et al.*, "Identifying ADHD-related abnormal functional connectivity with a graph convolutional neural network," *Neural Plasticity*, vol. 2024, 2024, doi: 10.1155/2024/8862647.

[5]   B. Qiu, Q. Wang, X. Li, W. Li, W. Shao, and M. Wang, "Adaptive spatial-temporal neural network for ADHD identification using functional fMRI," *Frontiers in Neuroscience*, vol. 18, May 2024, doi: 10.3389/fnins.2024.1394234.

[6]   A. Hatami, A. Ranjbar, and S. Azizi, "Utilizing fMRI and deep learning for the detection of major depressive disorder: A MobileNet V2 approach," in *HORA 2024 - 6th International Congress on Human-Computer Interaction, Optimization and Robotic Applications, Proceedings*, 2024, pp. 1–5, doi: 10.1109/HORA61326.2024.10550687.

[7]   R. Liu, Z. A. Huang, Y. Hu, Z. Zhu, K. C. Wong, and K. C. Tan, "Attention-like multimodality fusion with data augmentation for diagnosis of mental disorders using MRI," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 6, pp. 7627–7641, 2024, doi: 10.1109/TNNLS.2022.3219551.

[8]   N. Alsharif, M. H. Al-Adhaileh, and M. Al-Yaari, "Diagnosis of attention deficit hyperactivity disorder: A deep learning approach," *AIMS Mathematics*, vol. 9, no. 5, pp. 10580–10608, 2024, doi: 10.3934/math.2024517.

[9]   S. Agarwal, A. Raj, A. Chowdhury, G. Aich, R. Chatterjee, and K. Ghosh, "Investigating the impact of standard brain atlases and connectivity measures on the accuracy of ADHD detection from fMRI data using deep learning," *Multimedia Tools and Applications*, vol. 83, no. 25, pp. 67023–67057, 2024, doi: 10.1007/s11042-023-17962-7.

[10] P. G. Gülhan and G. Özmen, "The use of FMRI regional analysis to automatically detect ADHD through a 3D CNN-based approach," *Journal of Imaging Informatics in Medicine*, 2024, doi: 10.1007/s10278-024-01189-5.

[11] S. Saurabh and P. K. Gupta, "Deep learning-based modified bidirectional LSTM network for classification of ADHD disorder," *Arabian Journal for Science and Engineering*, vol. 49, no. 3, pp. 3009–3026, 2024, doi: 10.1007/s13369-023-07786-w.

[12] E. Salah, M. Shokair, F. E. A. El-Samie, and W. A. Shalaby, "Utilization of fMRI with optical amplification to diagnose attention deficit hyperactivity disorder," *Journal of Optics (India)*, 2024, doi: 10.1007/s12596-023-01485-3.

[13] P. Amado-Caballero, P. Casaseca-de-la-Higuera, S. Alberola-López, J. M. Andrés-de-Llano, J. A. López-Villalobos, and C. Alberola-López, "Insight into ADHD diagnosis with deep learning on actimetry: Quantitative interpretation of occlusion maps in age and gender subgroups," *Artificial Intelligence in Medicine*, vol. 143, p. 102630, Sep. 2023, doi: 10.1016/j.artmed.2023.102630.

[14] Y. Qin, Y. Lou, Y. Huang, R. Chen, and W. Yue, "An ensemble deep learning approach combining phenotypic data and fMRI for ADHD diagnosis," *Journal of Signal Processing Systems*, vol. 94, no. 11, pp. 1269–1281, 2022, doi: 10.1007/s11265-022-01812-0.

[15] Z. Mao *et al.*, "Spatio-temporal deep learning method for ADHD fMRI classification," *Information Sciences*, vol. 499, pp. 1–11, 2019, doi: 10.1016/j.ins.2019.05.043.

[16] "The ADHD-200 global competition." http://fcon_1000.projects.nitrc.org/indi/adhd200/results.html (accessed Oct. 01, 2017).

[17] S. Liu *et al.*, "Deep spatio-temporal representation and ensemble classification for attention deficit/hyperactivity disorder," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1–10, 2021.

[18] D. Kuang, X. Guo, X. An, Y. Zhao, and L. He, "Discrimination of ADHD based on fMRI data with deep belief network," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8590 LNBI, pp. 225–232, 2014, doi: 10.1007/978-3-319-09330-7_27.

[19] L. Zou, J. Zheng, C. Miao, M. J. McKeown, and Z. J. Wang, "3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI," *IEEE Access*, vol. 5, pp. 23626–23636, 2017, doi: 10.1109/ACCESS.2017.2762703.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[21] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277–281, 2020, doi: 10.1109/LGRS.2019.2918719.

[22] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Deep generative endmember modeling: An application to unsupervised spectral unmixing," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 374–384, 2019, doi: 10.1109/tci.2019.2948726.

[23] S. H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, and Y. D. Zhang, "Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network," *Information Fusion*, vol. 67, pp. 208–229, 2021, doi: 10.1016/j.inffus.2020.10.004.

[24] Q. Cui, S. Wu, Q. Liu, W. Zhong, and L. Wang, "MV-RNN: A multi-view recurrent neural network for sequential recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 2, pp. 317–331, 2020, doi: 10.1109/TKDE.2018.2881260.

[25] B. Kong *et al.*, "Learning tree-structured representation for 3D coronary artery segmentation," *Computerized Medical Imaging and Graphics*, vol. 80, 2020, doi: 10.1016/j.compmedimag.2019.101688.

[26] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Journal of Machine Learning Research*, vol. 9, pp. 249–256, 2010.

[27] D. Dai, J. Wang, J. Hua, and H. He, "Classification of ADHD children through multimodal magnetic resonance imaging," *Frontiers in Systems Neuroscience*, no. SEPTEMBER, pp. 1–8, 2012, doi: 10.3389/fnsys.2012.00063.

[28] S. Dey, R. Rao, and M. Shah, "Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects," *Frontiers in Neural Circuits*, vol. 8, no. JUNE, 2014, doi: 10.3389/fncir.2014.00064.

BIOGRAPHIES OF AUTHORS

Zarina Begum graduated from Jawaharlal Nehru Technological University in Kakinada with a bachelor's and master's degree in engineering. She is presently enrolled at VIT-AP University, which is close to Vijayawada, Andhra Pradesh, to pursue a Ph. D. in computer science. Machine learning, deep learning, optimization techniques, the internet of things (IoT), data mining, and big data are among her current research interests. Her email address is zarinabegum.714@gmail.com.

Kareemulla Shaik is a senior assistant professor at VIT-AP University's School of Computer Science & Engineering, which is located close to Vijayawada, Andhra Pradesh. He has been a teacher for more than fifteen years. He received a Ph.D. from KLEF, a B. Tech. from JNTU-Hyderabad, and an M. Tech. from JNTU-Kakinada. In addition to guiding numerous undergraduate and graduate projects, he has presented and published his work in a number of reputable journals and conferences. Additionally, he published patents from India. His research focuses on machine learning, deep learning, and computer networks and security. Additionally, he arranged conferences and gave numerous talks. He emails address is kareemulla.shaik@vitap.ac.in.