HbO2 and HbR Spatiotemporal Heatmap Feature Fusion Network for ADHD Diagnosis in Children

Mengxiang Chu[1,†], Xichen Wang[1,†], Hongbo Guo[1], Xingxing Cheng[1], Xiaohan Li[1], Xiaowei He[1], Xiao Li[1], Jiaojiao Ren[2], and Jingjing Yu[3,*]

[1] School of Electronic Information (Artificial Intelligence), Northwest University, Xi'an, 710127, China.
[2] Department of Pediatric Health Care, Xi'an People's Hospital (Xi'an Fourth Hospital), Xi'an, 710004, China.
[3] School of Physics and Information Technology, Shaanxi Normal University, Xi'an, 710062, China.
[†] Co-first Author; * Corresponding Author, yujj@snnu.edu.cn.

Abstract. Functional near-infrared spectroscopy (fNIRS)-based diagnosis of childhood attention-deficit/hyperactivity disorder (ADHD) is limited by spatiotemporal dynamic decoupling and neglected physiological coupling between oxyhemoglobin (HbO2) and deoxyhemoglobin (HbR). To address these issues, we propose a Spatiotemporal Heatmap Feature Fusion Network (STHFFN) for modeling abnormal neurovascular trajectories and bilateral hemoglobin interactions in ADHD. STHFFN starts with the Spatiotemporal Heatmap Construction to convert discrete fNIRS signals into anatomically continuous cortical heatmaps to preserve brain topology and provide a physiological basis for feature learning. Based on this structured representation, the Parallel Spatiotemporal attention module uses a bi-branch structure to jointly model spatial coactivation patterns and temporal hemodynamic changes, enabling unified spatiotemporal feature learning instead of separate modeling. The Bilateral Hemoglobin Complementarity explicitly captures spatiotemporal interactions between HbO2 and HbR, improving sensitivity to pathological neurovascular decoupling. The Memory Pool mechanism further aggregates long-range temporal dynamics to characterize persistent attentional instability, and multimodal features are adaptively fused for final diagnostic prediction. Extensive experiments on 47 children with ADHD and 47 healthy controls show that STHFFN achieves 92.57% ACC and 94.25% AUC, outperforming eight representative baseline models. By integrating anatomical continuity, spatiotemporal dynamics, and physiological coupling, STHFFN provides an effect approach for fNIRS-based auxiliary diagnosis of children with ADHD. Code is at https://github.com/wangbaobao7778-cyber/STHFFN.

Keywords: Children with ADHD, fNIRS, Auxiliary Diagnosis, Spatiotemporal Attention, Hemoglobin complementarity.

1 Introduction

Attention deficit hyperactivity disorder (ADHD) is characterized by persistent inattention and impulsivity, fundamentally rooted in prefrontal hypoactivation and complex nonlinear spatiotemporal hemodynamic deviations [8, 9]. While functional near-infrared spectroscopy (fNIRS) offers superior spatiotemporal resolution over traditional subjective scales for early objective assessment [10], accurately decoding these highly irregular neural fluctuations remains a formidable computational challenge. Recent deep learning advances, such as fNIRS-Net [28], hybrid CNN-LSTMs [24], fNIRS-T [29], and T&F-DFC [7], have significantly improved multi-feature fNIRS decoding. However, these methods such as TimesFormer [3], Joint Space-Time Attention [26], Divided Space-Time Attention [11, 12, 27], Sparse Local Global Attention [6], Axial Attention [11, 12, 27], typically employ serialized modeling paradigms that treat space and time independently. This inductive bias forcibly isolates spatial topologies from temporal evolution, causing spatiotemporal dynamic decoupling and over-smoothing the transient, localized anomalies critical for ADHD identification. Most critically, sustained attentional instability, the defining core symptom of ADHD, is inherently a cumulative temporal process rather than isolated instantaneous activation [21, 19]. The sliding window (SW) strategies only capture fragmented transient brain activity, failing to track the progressive attentional decline throughout the entire task duration [22]. Static feature extraction further ignores temporal cumulative effects, making it impossible to distinguish normal transient attentional fluctuations from the persistent deficits characteristic of ADHD. This limitation severely reduces the sensitivity of existing models to ADHD's core pathophysiology.

Beyond spatiotemporal tracking, robust neural decoding can be significantly enhanced by explicitly modeling the physiological interdependence of local hemodynamics. In a healthy cerebral cortex, elevated oxyhemoglobin (HbO2) indicating blood supply and depleted deoxyhemoglobin (HbR) reflecting oxygen consumption exhibit a stable opposite-phase coupling [13]. However, this delicate neurovascular synergy is frequently disrupted or decoupled by underlying metabolic disturbances in children with ADHD [25, 1]. Existing classification frameworks, including fNIRS-Net [28], fNIRS-T [29], ID-CapsuleNet [17], CT-Net [18], and IVCAN [16], primarily rely on single-modal inputs or simple channel concatenation. This naive modality isolation ignores the synergistic physiological coupling between HbO2 and HbR. Therefore, these models lack the sensitivity required to distinguish between genuine pathological decoupling and systemic physiological noise such as the respiratory or cardiac artifacts [15, 23], thus fundamentally limiting their diagnostic performance and clinical reliability.

To address these challenges, we propose a spatiotemporal heatmap feature fusion network (STHFFN) for childhood ADHD auxiliary diagnosis, modeling fNIRS nonlinear spatiotemporal anomalies and HbO2 and HbR physiological coupling. Our key contributions are threefold: (1) STHFFN converts discrete fNIRS channel signals into continuous spatiotemporal heatmaps to preserve anatomical continuity, overcoming spatiotemporal decoupling and modality isolation for integrated modeling of anatomical information, spatiotemporal dynam-

ics, and neurovascular coupling. (2) A parallel spatiotemporal (P-ST) attention module unbiasedly captures spatial co-activation and temporal evolution, while a bilateral hemoglobin complementarity (BHC) module mines HbO2 and HbR synergistic dependencies via cross-modal attention. (3) Equipped with a memory pool (MP) mechanism to aggregate long-range dynamics, STHFFN achieves 92.57% accuracy, providing a robust childhood ADHD auxiliary diagnostic tool.

2 Method

To decode the pronounced spatiotemporal heterogeneity and disrupted neurovascular coupling in childhood ADHD, we propose the STHFFN framework (Fig. 1(a)) that realizes integration of anatomical mapping, spatiotemporal modeling, cross-modal fusion, and long-term dependency aggregation. STHFFN first leverages a spatiotemporal heatmap construction (STHC) module to convert discrete fNIRS signals into continuous heatmaps as inputs [2, 14], preserving the anatomical continuity of cortical activation. These heatmaps are then concurrently processed to capture dual-hemoglobin dynamics. A P-ST attention module unbiasedly decouples and models spatial co-activation patterns and temporal evolution trajectories of HbO2 and HbR, while a BHC module explicitly mines the synergistic physiological dependencies between HbO2 and HbR via cross-modal attention. Subsequently, a MP mechanism aggregates localized features across consecutive segments to characterize long-range dynamic trajectories underlying sustained attentional instability. Finally, a linear fusion attention (LFA) module aligns and refines dual-modality features to suppress noise, feeding the integrated representations into a fully connected classifier for ADHD diagnosis.

2.1 Spatiotemporal Heatmap Construction

In the STHC module, let $X \in \mathbb{R}^{T\times C}$ denote the raw fNIRS signals, where T is the number of temporal frames and $C = 22$ represents the channels. Unlike standard grid-based image data, fNIRS channels exhibit a cross-shaped spatial distribution (Fig. 1(b)). To enable convolution operations while preserving anatomical spatial information between brain regions, the STHC module maps the channels onto a virtual 5×9 grid based on their Montreal Neurological Institute (MNI) coordinates. Then, a SW strategy with a window size and step size of L frames is employed to segment the continuous fNIRS signal. For each frame t within a SW, cubic spline interpolation is applied to generate a dense heatmap $H_t \in \mathbb{R}^{H\times W}$. This results in a spatiotemporal tensor $X_{SW} \in \mathbb{R}^{L\times H\times W}$ for each window. This STHC representation ensures that the input reliably captures the local temporal dynamics of the HbO2 and HbR responses alongside the anatomical spatial information of cortical activation.

2.2 Parallel Spatiotemporal Attention Module

Sequential modeling often introduces an inductive bias by prioritizing one dimension over another. To resolve this, the P-ST module (Fig. 1(c)) employs

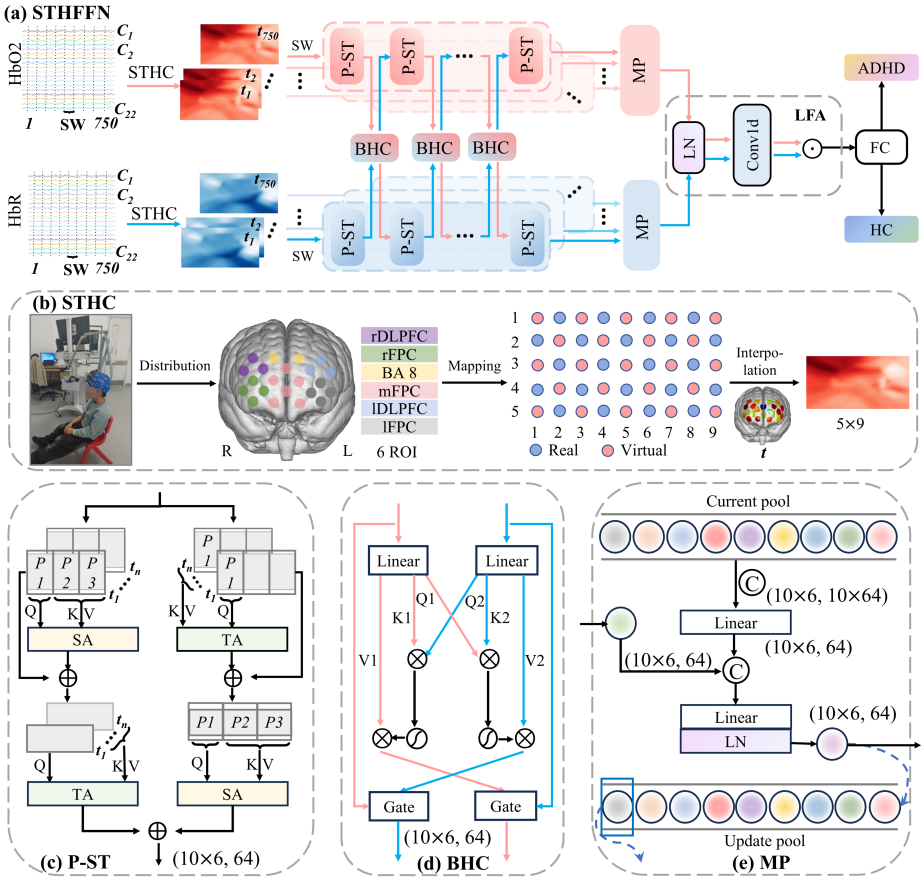

Fig. 1. (a) STHFFN framework for ADHD diagnostic pipeline; (b) Spatiotemporal heatmap construction (STHC) converts discrete fNIRS signals to continuous anatomical maps; (c)-(e) Parallel Spatiotemporal (P-ST) attention, Bilateral Hemoglobin Complementarity (BHC), Memory Pool (MP) modules for spatiotemporal modeling, HbO2 and HbR coupling, and long-range dependency aggregation.

two concurrent branches to unbiasedly decouple and capture both spatial correlations and temporal dynamics. Let $F^{(l-1)} \in \mathbb{R}^{L\times N\times D}$ be the input feature tensor, where N is the number of spatial patches and D is the feature dimension.

Spatial Co-activation Branch: To identify spatial activation patterns within each t, we compute intra-frame spatial attention (SA) for patch i across all other patches j (Pj), using a residual connection to preserve local details:

$$f_{t,i}^{\text{space}} = f_{t,i}^{(l-1)} + \text{Attention}_{j\neq i}\left(\frac{f_{t,i}^{(l-1)}(f_{t,j}^{(l-1)})^{\top}}{\sqrt{d_k}}\right) f_{t,j}^{(l-1)} \tag{1}$$

Subsequently, temporal attention (TA) is applied across time frames $\tau = 1 \dots L$ to track the evolution of these spatially-refined features:

$$f_{t,i}^{\text{left}} = \text{Attention}_{\tau \neq t} \left(\frac{f_{t,i}^{\text{space}} (f_{\tau,i}^{\text{space}})^\top}{\sqrt{d_k}} \right) f_{\tau,i}^{\text{space}} \tag{2}$$

Temporal Continuity Branch: To capture the dynamic continuity within specific regions, TA is first computed for patch i across frames τ, preserving local time-varying trends:

$$f_{t,i}^{\text{time}} = f_{t,i}^{(l-1)} + \text{Attention}_{\tau \neq t} \left(\frac{f_{t,i}^{(l-1)} (f_{\tau,i}^{(l-1)})^\top}{\sqrt{d_k}} \right) f_{\tau,i}^{(l-1)} \tag{3}$$

Then, SA is applied to these temporally-aware features to model cross-region interactions:

$$f_{t,i}^{\text{right}} = \text{Attention}_{j \neq i} \left(\frac{f_{t,i}^{\text{time}} (f_{t,j}^{\text{time}})^\top}{\sqrt{d_k}} \right) f_{t,j}^{\text{time}} \tag{4}$$

Finally, the outputs are fused ($F^{(l)} = f_{t,i}^{\text{left}} + f_{t,i}^{\text{right}}$) to produce a comprehensive representation, effectively capturing neural synchronization asymmetry in ADHD without dimensional preference.

2.3 Bilateral Hemoglobin Complementarity Module

While HbO2 typically reflects neural activation via oxygen supply, HbR reflects oxygen consumption, direct splicing of F_{HbO2} and F_{HbR} often fails to fully utilize their coupling correlation and physiological interdependence. We introduce the BHC module (Fig. 1(d)) where one modality can retrieve complementary contextual information from the other. We use F_{HbO2} as a query to guide the aggregation of global structural information from F_{HbR}:

$$M_{\text{HbO2}\leftarrow\text{HbR}} = \text{Softmax} \left(\frac{(F_{\text{HbO2}} W_Q)(F_{\text{HbR}} W_K)^\top}{\sqrt{d_k}} \right) (F_{\text{HbR}} W_V) \tag{5}$$

where W_Q, W_K, W_V are learnable projections. The HbR stream is symmetrically updated ($M_{\text{HbR}\leftarrow\text{HbO2}}$). To prevent the propagation of modality-specific noise, a Gated Fusion mechanism adaptively regulates the integration of these retrieved features: $G_{\text{HbO2}} = \sigma\left(W_g[F_{\text{HbO2}} \parallel M_{\text{HbO2}\leftarrow\text{HbR}}]\right)$,$\hat{F}_{\text{HbO2}} = G_{\text{HbO2}} \odot M_{\text{HbO2}\leftarrow\text{HbR}} + (1 - G_{\text{HbO2}}) \odot F_{\text{HbO2}}$ where $\parallel$ denotes concatenation, and $\odot$ is element-wise multiplication. Symmetrical operations yield $\hat{F}_{\text{HbR}}$. By dynamically weighting these features, the network prioritizes synergistic HbO2 and HbR patterns for robust ADHD diagnosis.

2.4 Memory Pool Updating Strategy

ADHD diagnosis requires consideration of both transient states and the overall temporal variation trend of activation. The fNIRS signal in ADHD fluctuates

irregularly, and a single SW can only capture local patterns, failing to capture cumulative changes and long-term dynamics. To capture global temporal dependencies, we implement a first-in-first-out (FIFO) MP $\mathcal{M}$ to accumulate historical activation features within continuous SWs, thereby modeling the complete brain activation trajectory. As shown in Fig. 1(e), let $\hat{F}_k$ be the feature vector of the current SW k derived from the P-ST module for HbO2 or HbR, we accumulate historical features H_{global} within SWs, thereby modeling the complete brain activation trajectory: $h_t^{\text{new}} = \text{LayerNorm}\left(W_u[H_{global} \parallel \hat{F}_k] + b_u\right)$ For the final SW, the output h_t^{new} serves as the final global feature representation.

3 Experiments and Results

This study included 47 children with ADHD (31 males, aged 8.79 ± 1.98, IQ 107.9 ± 11.2) and 47 demographically matched HCs (30 males, aged 8.62 ± 2.41, IQ 106.6 ± 11.8) recruited from Xi'an People's Hospital (Xi'an Fourth Hospital) (Approved No. KJLL-Z-H-2023005) under an approved protocol. The verbal fluency task (VFT) comprised a 30 s numerical repetition, a 60 s word generation block (3 cues, 20 s each), and a 60 s numerical repetition. fNIRS signals were acquired at 10 Hz (ETG-One, Hitachi Medical, Tokyo, Japan) and preprocessed to obtain HbO2 and HbR [8]. The 22-channel probe configuration covered 6 regions of interest (ROI) (Fig. 1(b)): left and right dorsolateral prefrontal cortex (lDLPFC, rDLPFC), medial, left and right frontopolar cortex (mFPC, lFPC, rFPC), and Brodmann area 8 (BA8) [8].

Eight state-of-the-art models were used to compare and evaluate the performance of STHFFN, including LightTS [5] by Campos et al., IVCAN [16], ID-CapsuleNet [17], CT-Net [18], iTransformer [20] by Liu et al., fNIRS-Net [28], fNIRS-T [29], and TimesNet [30] by Wu et al., using concatenated HbO2 and HbR as input. The evaluation employed a five-fold cross-validation with an 80:20 split across 60 independent trials, and the data were augmented via odd-even frame separation only after train-test splitting to eliminate data leakage and overfitting [7]. Performance was assessed using ACC, F1-score (F1), precision (PRE), recall (REC), and area under the curve (AUC). The L was set to 10, batch size of 16 and learning rate of 0.0001.

3.1 Results

Temporal Variation of HbO2 and HbR: The HC exhibited strong, rhythmic high-amplitude fluctuations with distinct phase-oppositional coupling during VFT (Fig. 2), indicating intact neurovascular regulation and synchronization. In contrast, the ADHD group displayed significant temporal turbulence and signal attenuation. The erratic waveforms and divergent channel trajectories reflected an inability to sustain stable cortical activation and increased spatiotemporal heterogeneity, consistent with Bian's study [4]. These complex nonlinear spatiotemporal deviations justified the necessity of STHFFN for accurate diagnosis.

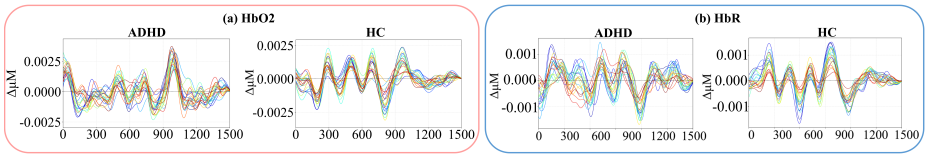

Fig. 2. Temporal Variation of HbO2 and HbR in ADHD and HC group.

Table 1. Performance comparison with baselines in ADHD diagnosis.

Methods	ACC (%)	F1 (%)	PRE (%)	REC (%)	AUC (%)
LightTS [5]	86.50 ± 3.95	86.30 ± 5.01	85.77 ± 6.04	87.86 ± 3.94	87.03 ± 5.31
IVCAN [16]	89.45 ± 4.20	89.30 ± 4.80	88.60 ± 5.50	90.20 ± 5.10	90.50 ± 4.00
ID-CapsuleNet [17]	86.95 ± 5.20	87.10 ± 5.80	86.50 ± 6.90	87.75 ± 6.50	87.80 ± 4.40
CT-Net [18]	88.42 ± 3.65	88.65 ± 3.95	88.10 ± 5.15	89.25 ± 4.30	89.80 ± 3.45
iTransformer [20]	81.45 ± 4.50	81.79 ± 4.12	81.50 ± 4.30	82.10 ± 5.15	81.85 ± 4.88
fNIRS-Net [28]	83.90 ± 6.80	83.95 ± 8.20	83.10 ± 7.10	84.85 ± 8.60	84.45 ± 5.25
fNIRS-T [29]	85.20 ± 6.30	85.42 ± 7.50	83.80 ± 8.90	87.15 ± 8.50	86.10 ± 3.85
TimesNet [30]	84.15 ± 7.12	83.84 ± 8.80	83.30 ± 7.50	85.15 ± 8.95	84.95 ± 5.05
STHFFN	$\mathbf{92.57 \pm 2.53}$	$\mathbf{92.35 \pm 3.70}$	$\mathbf{89.52 \pm 8.09}$	$\mathbf{96.18 \pm 4.69}$	$\mathbf{94.25 \pm 3.00}$

Comparison with baselines: Table 1 presents the performance of STHFFN and the compared models in ADHD diagnosis. STHFFN achieved the best overall diagnostic performance among all compared methods, with an accuracy of 92.57%±2.53%, F1 of 92.35%±3.70%, and AUC of 94.25%±3.00%. Compared with the strongest IVCAN, STHFFN improved ACC by approximately 3.1% and AUC by 3.7%, indicating that the proposed bilateral hemoglobin complementary modeling and memory-enhanced fusion strategy provided more discriminative representations for the assisted diagnosis of children with ADHD. Furthermore, STHFFN achieved a PRE of 89.52% and REC of 96.18%, exhibiting high sensitivity while maintaining stable specificity.

Ablation experiments: The ablation results in Table 2 demonstrate that each proposed component contributed meaningfully to performance. The most significant performance degradation occurred without the MP, yielding an ACC of 87.31% and an AUC of 86.24%, confirming that capturing the complete brain activity trajectory was crucial for detecting the cumulative sustained attention deficits inherent to ADHD. Removing the BHC and LFA resulted in an approximate 4.2% decrease in ACC and a 3.6%–4.6% decrease in F1, suggesting that modeling the synergistic relationship between HbO2 and HbR provided crucial supplementary information missed by single-modality or simple fusion methods. Removing the P-ST module reduced the bilateral spatiotemporal modeling ability of brain regions, resulting in an ACC of 89.94% and an F1 of 89.85%.

Parameter sensitivity experiment: We further investigated the model's sensitivity to key parameters (Table 3). First, a 2-layer P-ST and BHC depth yielded optimal performance, while deeper networks exhibited performance degradation, likely due to overfitting caused by the limited sample size. Second, a MP size of 10 resulted in the best performance; smaller pools size of 5 with ACC

Table 2. Ablation study of STHFFN components.

Methods	ACC (%)	F1 (%)	PRE (%)	REC (%)	AUC (%)
STHFFN	**92.57 ± 2.53**	**92.35 ± 3.70**	**89.52 ± 8.09**	96.18 ± 4.69	**94.25 ± 3.00**
w/o P-ST	89.94 ± 4.15	89.85 ± 5.74	84.48 ± 9.21	**97.00 ± 6.00**	92.06 ± 5.94
w/o BHC	88.30 ± 3.91	88.72 ± 3.99	83.19 ± 6.24	96.00 ± 8.00	92.50 ± 1.95
w/o MP	87.31 ± 5.29	86.74 ± 8.41	82.12 ± 10.05	92.67 ± 9.04	86.24 ± 5.52
w/o LFA	88.36 ± 6.08	87.82 ± 7.49	85.35 ± 9.45	92.18 ± 11.64	91.60 ± 6.19

of 88.30% failed to capture sufficient historical blood oxygenation dynamics, while larger pools size of 15 with ACC of 90.58% introduced irrelevant outdated information. Third, patch partitioning based on six biological ROIs achieved the highest accuracy of 92.57%, indicating that biologically defined partitioning better reflected the brain's intrinsic functional organization than arbitrary grid partitioning. Finally, cubic spline interpolation significantly outperformed zero-padding, which yielded an ACC of 66.38%, highlighting the importance of maintaining the anatomical continuity of HbO2 and HbR dynamic signals.

Table 3. The results of parameter sensitivity analysis.

Parameter	Setting	ACC (%)	F1 (%)	PRE (%)	REC (%)	AUC (%)
(P-ST +BHC) Layers	1	91.58 ± 5.37	90.06 ± 8.76	89.66 ± 6.85	90.27 ± 8.43	91.00 ± 6.09
	2	**92.57 ± 2.53**	**92.35 ± 3.70**	**89.52 ± 8.09**	**96.18 ± 4.69**	**94.25 ± 3.00**
	3	89.36 ± 3.34	87.82 ± 7.23	89.71 ± 5.44	88.05 ± 6.90	88.77 ± 4.10
	4	88.36 ± 3.84	87.01 ± 7.17	89.02 ± 5.41	87.25 ± 6.84	87.75 ± 4.43
Memory Pool Size	5	88.30 ± 2.06	88.67 ± 4.40	81.54 ± 7.90	98.00 ± 4.00	93.73 ± 4.08
	10	**92.57 ± 2.53**	**92.35 ± 3.70**	**89.52 ± 8.09**	**96.18 ± 4.69**	**94.25 ± 3.00**
	15	90.58 ± 6.32	88.73 ± 9.42	90.49 ± 8.57	90.14 ± 5.03	90.67 ± 9.04
Patch Configuration	**6 (ROIs)**	**92.57 ± 2.53**	**92.35 ± 3.70**	**89.52 ± 8.09**	**96.18 ± 4.69**	**94.25 ± 3.00**
	1 (5×9)	88.30 ± 3.92	86.80 ± 7.22	88.87 ± 5.48	87.05 ± 6.90	87.72 ± 4.46
	2 (5×4+5×5)	88.36 ± 6.08	86.88 ± 8.52	89.00 ± 6.62	87.16 ± 8.19	87.77 ± 6.46
	2 (5×5+5×4)	87.31 ± 6.25	85.10 ± 11.09	87.71 ± 8.26	85.59 ± 10.34	86.53 ± 7.20
	3 (5×3×3)	89.41 ± 4.64	88.78 ± 5.14	90.37 ± 4.17	88.92 ± 5.03	88.97 ± 4.75
Interpolation	**Cubic spline**	**92.57 ± 2.53**	**92.35 ± 3.70**	**89.52 ± 8.09**	**96.18 ± 4.69**	**94.25 ± 3.00**
	Zero-padding	66.38 ± 7.97	61.63 ± 14.46	65.84 ± 16.54	65.11 ± 11.24	64.83 ± 7.96

4 Conclusion

The proposed STHFFN is designed for fNIRS-based childhood ADHD diagnosis. It jointly models three core pathological characteristics of ADHD: abnormal prefrontal spatiotemporal activation, impaired neurovascular coupling, and sustained attentional instability. The STHC module converts discrete channel signals into continuous spatiotemporal heatmaps to preserve anatomical continuity. P-ST attention captures spatial and temporal dynamics. The BHC module models neurovascular coupling. The MP module aggregates long-range dependencies associated with sustained attentional deficits. Together, these components establish an integrated anatomical, spatiotemporal, and physiological modeling framework. Experiments on a clinical dataset demonstrate that STHFFN achieves

92.57% ACC and 94.25% AUC, outperforming eight representative baselines, with ablation analyses confirming the contribution of each core module.

Four key considerations regarding our methodology and findings are addressed as follows. First, given the hemodynamic response time constant of 2–6 s (0.1-0.5 Hz), the 10 Hz fNIRS system satisfies the Nyquist theorem and is standard practice in fNIRS research. Higher sampling rates increase data volume but introduce excessive high-frequency noise. Second, cubic spline interpolation outperforms zero-padding. It preserves anatomical continuity of brain activation, maintains MNI-aligned cortical topology, and avoids biologically invalid spatial artifacts. Third, we have completed quantitative analysis of brain activation across 6 ROIs, and identified rFPC, mFPC, and BA8 as core pathological hubs exhibiting abnormal activation and neurovascular uncoupling, which aligns with established findings in the ADHD literature [8]. Fourth, while multi-center validation is critical for generalizability, public fNIRS-ADHD datasets remain scarce, and recent high-quality studies [25, 9] also used single-center cohorts of comparable size. We will collaborate with other hospitals to expand data collection and extend validation to related pediatric neurological disorders in future.

In conclusion, STHFFN provides an effective framework for ADHD children diagnosis and establishes a foundation for future multicenter validation.

Acknowledgments. This study was funded by: National Natural Science Foundation of China (12571457, 62506298); General Project of the Key Research and Development Plan of Shaanxi Province (2025SF-YBXM-240); Research incubation project of Xi'an People's Hospital (Xi'an Fourth Hospital) (ZD-20).

Disclosure of Interests. The authors have no conflicts to disclose.

References

1. Ahirwar, L.K., Blackburn, S.L., McBride, D.W., T, P.K.: Reviewing vascular influences on neuronal migration, cortical development, and neurodevelopmental disorders: focus on autism, adhd and schizophrenia. Molecular Psychiatry **30**(12), 5953–5966 (2025)
2. Bashivan, P., Rish, I., Yeasin, M., Codella, N.: Learning representations from eeg with deep recurrent-convolutional neural networks. arXiv preprint arXiv:1511.06448 (2015)
3. Bertasius, G., Wang, H., Torresani, L.: Is space-time attention all you need for video understanding? In: Icml. vol. 2, p. 4 (2021)
4. Bian, Z., Ma, X., Wang, Y., Yu, X.: Verbal fluency tasks and attention problems in children with adhd: evidence from fnirs. Frontiers in Neurology **16**, 1541589 (2025)
5. Campos, D., Zhang, M., Yang, B., Kieu, T., Guo, C., Jensen, C.S.: Lightts: Lightweight time series classification with adaptive ensemble distillation. Proceedings of the ACM on Management of Data **1**(2), 1–27 (2023)
6. Child, R., Gray, S., Radford, A., Sutskever, I.: Generating long sequences with sparse transformers. arXiv preprint arXiv:1904.10509 (2019)
7. Chu, M., Ma, Y., He, X., Li, X., Ren, J., Zhong, Z., Yu, J., Guo, H.: T&f-dfc fusionnet: Time&frequency-dynamic functional connectivity fusion network for adhd

diagnosis in children based on fnirs. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 627–637. Springer (2025)
8. Chu, M., Ren, J., Yang, L., Ma, Y., Zhong, Z., He, X., Guo, H., Li, X., Yu, J., Wang, Z.: Exploring the neural mechanisms of adhd in children: a multifeature cross-task fnirs analysis. Cerebral Cortex **35**(6) (2025)
9. Gao, Z., Duberg, K., Warren, S.L., Zheng, L., Hinshaw, S.P., Menon, V., Cai, W.: Reduced temporal and spatial stability of neural activity patterns predict cognitive control deficits in children with adhd. Nature Communications **16**(1), 2346 (2025)
10. Gu, Y., Miao, S., Yang, J., Li, X.: Adhd children identification with multiview feature fusion of fnirs signals. IEEE Sensors Journal **22**(13), 13536–13543 (2022)
11. Ho, J., Kalchbrenner, N., Weissenborn, D., Salimans, T.: Axial attention in multidimensional transformers. arXiv preprint arXiv:1912.12180 (2019)
12. Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Liu, W.: Ccnet: Criss-cross attention for semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 603–612 (2019)
13. Ji, H., Xu, Z., Wang, M., Zou, H., Chen, Y., Ai, J.: A flexible optoelectronic device for continuous cerebral blood flow monitoring. Biosensors **12**(11), 944 (2022)
14. Jiao, M., Yang, S., Xian, X., Fotedar, N., Liu, F.: Multi-modal electrophysiological source imaging with attention neural networks based on deep fusion of eeg and meg. IEEE Transactions on Neural Systems and Rehabilitation Engineering **32**, 2492–2502 (2024)
15. Lamata, P., Cookson, A., Smith, N.: Clinical diagnostic biomarkers from the personalization of computational models of cardiac physiology. Annals of biomedical engineering **44**(1), 46–57 (2016)
16. Li, Y., Li, S., Yuan, Z., Zhao, S., Wan, F., Xu, T., Zhang, H., Wang, H.: Ivcan: An improved visual curve attention network for fnirs-based motor imagery/execution classification. Biomedical Signal Processing and Control **104**, 107679 (2025)
17. Li, Y., Xu, T., Li, J., Wan, F., Wang, H.: Improved dilation capsulenet for motor imagery and mental arithmetic classification based on fnirs. Brain-Apparatus Communication: A Journal of Bacomics **3**(1), 2335886 (2024)
18. Liao, L., Lu, J., Wang, L., Zhang, Y., Gao, D., Wang, M.: Ct-net: an interpretable cnn-transformer fusion network for fnirs classification. Medical & Biological Engineering & Computing **62**(10), 3233–3247 (2024)
19. Lin, C., Lu, H., Pan, C., Ma, S., Zhang, Z., Tian, R.: Mbrstcformer: a knowledge embedded local–global spatiotemporal transformer for emotion recognition. Cognitive Neurodynamics **19**(1), 95 (2025)
20. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M.: itransformer: Inverted transformers are effective for time series forecasting. arXiv preprint arXiv:2310.06625 (2023)
21. Qin, H., Zhou, D., Xu, T., Bian, Z., Li, J.: Factorization vision transformer: Modeling long-range dependency with local window cost. IEEE Transactions on Neural Networks and Learning Systems **36**(2), 3151–3164 (2023)
22. Rahimi, N., Kumar, C., McLinden, J., Hosni, S.I., Borgheai, S.B., Shahriari, Y., Shao, M.: Topology-aware multimodal fusion for neural dynamics representation learning and classification. IEEE Sensors Journal **24**(13), 21062–21073 (2024)
23. Scarciglia, A., Catrambone, V., Bonanno, C., Valenza, G.: Physiological noise: Definition, estimation, and characterization in complex biomedical signals. IEEE transactions on biomedical engineering **71**(1), 45–55 (2023)
24. Shin, J., Konnai, S., Maniruzzaman, M., Hasan, M.A.M., Hirooka, K., Megumi, A., Yasumura, A.: Identifying adhd for children with coexisting asd from fnirs signals using deep learning approach. IEEE Access **11**, 82794–82801 (2023)

25. Su, S., Zhao, J., Dai, Y., Lin, L., Zhou, Q., Yan, Z., Qian, L., Cui, W., Liu, M., Zhang, H., et al.: Altered neurovascular coupling in the children with attention-deficit/hyperactivity disorder: a comprehensive fmri analysis. European Child & Adolescent Psychiatry **33**(4), 1081–1091 (2024)
26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
27. Wang, H., Zhu, Y., Green, B., Adam, H., Yuille, A., Chen, L.C.: Axial-deeplab: Stand-alone axial-attention for panoptic segmentation. In: European conference on computer vision. pp. 108–126. Springer (2020)
28. Wang, Z., Fang, J., Zhang, J.: Rethinking delayed hemodynamic responses for fnirs classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering **31**, 4528–4538 (2023)
29. Wang, Z., Zhang, J., Zhang, X., Chen, P., Wang, B.: Transformer model for functional near-infrared spectroscopy classification. IEEE Journal of Biomedical and Health Informatics **26**(6), 2559–2569 (2022)
30. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186 (2022)