Explainable ADHD Diagnostic Framework Using Weakly-Supervised Action Recognition

Ninghan Fan[1], Ming Kong[1], Jing Huang[1], Bingdi Chen[2], and Qiang Zhu[1] (✉)

[1] Zhejiang University
[2] The Institute for Biomedical Engineering & Nano Science, Tongji University
fannh02@gmail.com
{zjukongming, huangjin9, zhuq}@zju.edu.cn
inanochen@tongji.edu.cn

Abstract. The clinical diagnosis of Attention Deficit Hyperactivity Disorder (ADHD) primarily relies on scale questionnaires, clinical interviews, and executive function tests, which face challenges including limited medical resources, low diagnostic efficiency, and high dependence on clinicians' subjective experience. Existing AI-assisted diagnostic approaches based on behavioral analysis lack sufficient result interpretability, hindering their integration with conventional diagnostic workflows and practical clinical application. This paper proposes EDWAR, an Explainable ADHD Diagnostic Framework Using Weakly-Supervised Action Recognition, which establishes a collaborative diagnostic mechanism integrating behavioral analysis with traditional test records. By employing weakly-supervised action recognition methodology requiring only diagnostic labels and video-level annotations of abnormal behaviors, our framework not only achieves high diagnostic accuracy but also provides transparent interpretation through both video-level and timestep-wise anomaly action recognition. Experimental results demonstrate that EDWAR attains superior diagnostic performance while offering convincing and explainable evidence.

Keywords: ADHD Diagnosis · Weakly-Supervised Learning · Action Recognition · Explainable AI · Clinical Decision Support.

1 Introduction

Attention Deficit Hyperactivity Disorder (ADHD), a prevalent neurodevelopmental disorder [22, 25], manifests core symptoms including persistent hyperactivity, impulse dysregulation, and attentional deficits [15]. Current clinical diagnosis relies on composite evaluations combining standardized rating scales, behavioral observations, and executive function assessments [9, 13]. However, three critical limitations persist: (1) clinician-dependent subjectivity in behavioral interpretation leads to diagnostic inconsistency; (2) the absence of quantitative metrics for core hyperactive symptoms impedes objective verification; (3) disjointed analysis between qualitative observations and quantitative test results compounds diagnostic uncertainty.

The integration of artificial intelligence (AI) in ADHD assessment has yielded promising developments[20, 29]. Pioneering studies employed machine learning classifiers on structured diagnostic records[5, 23], followed by multimodal approaches incorporating neurophysiological data (EEG/MRI) and wearable sensor metrics[1, 4, 14]. Recent computer vision advancements enable video-based behavioral phenotyping through automated analysis of gaze patterns, facial microexpressions, and kinematic features [16, 21, 28]. While these vision-driven methods align well with conventional diagnostic protocols, a fundamental constraint remains unresolved: the opaque decision-making processes in prevailing black-box models hinder clinical trustworthiness.

To address this issue, we propose EDWAR, an **E**xplainable ADHD **D**iagnostic framework using **W**eakly-Supervised **A**ction **R**ecognition. The framework implements two key processes: First, it analyzes subjects' pose sequences captured during executive function tests through a weakly-supervised action recognition module, which detects activity segment proposals and quantifies activation intensities of anomaly actions [26]. Second, it synthesizes anomaly action scores with executive function test metrics to generate diagnostic conclusions. This multimodal integration enables EDWAR to not only achieve accurate ADHD diagnosis but also provide temporal-localized evidence to support its conclusions. Experimental results demonstrate EDWAR's dual capability: attaining state-of-the-art diagnostic accuracy (94.3%) while producing clinically interpretable explanations through precise anomaly action localization.

The framework advances AI-assisted ADHD diagnostics by: (1) establishing a novel paradigm that integrates executive function test metrics with explainable anomaly action recognition; (2) overcoming fine-grained anomaly action annotation bottlenecks through weakly-supervised learning adapted for clinical constraints; (3) experimentally validating the synchronization of diagnostic efficacy and explainability.

2 Method

As shown in Fig. 1, the EDWAR framework first requires subjects to complete standardized executive function tests. Behavioral recordings are synchronously captured via cameras during testing and converted into skeletal sequences, which serve as inputs for a two-stage collaborative reasoning framework. The weakly-supervised action recognition module extracts timestep-wise anomaly activations and video-wise anomaly scores from skeletal sequences, while the ADHD diagnosis module integrates these scores with executive function test metrics for diagnostic prediction. Below, we first detail the data collection and preprocessing pipeline, followed by the framework architecture and optimization strategy.

2.1 Data Collection and Preprocess

Executive Function Test Design. Three standardized neuropsychological tests were implemented: (1) *Stroop Test* [2] to evaluate attention control and

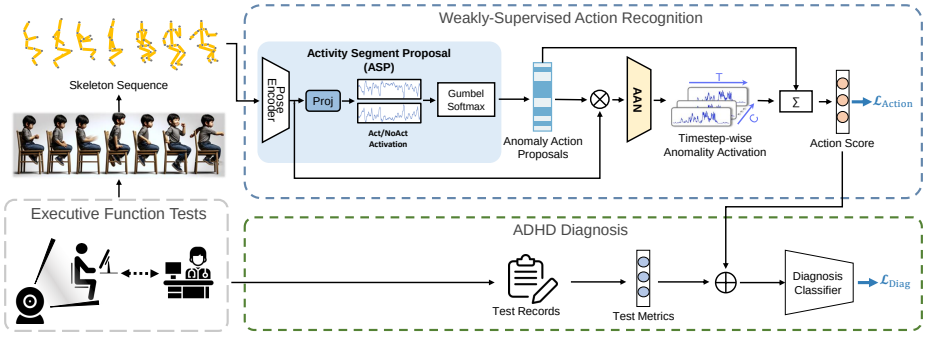

Fig. 1. The illustration of EDWAR framework.

cognitive flexibility; (2) *Wisconsin Card Sorting Test (WCST)* [19] to assess cognitive flexibility and problem-solving; (3) *Facial Emotion Recognition* [6] to measure emotional processing capacity. These tasks systematically probe distinct cognitive domains while eliciting ADHD-characteristic behavioral manifestations during focused task engagement. A certified assistant monitored the testing session in real-time to ensure protocol adherence and data acquisition reliability.

Pose Extraction. During executive function tests, participant behaviors were recorded via software-synchronized cameras to ensure temporal alignment between video segments and testing procedures. Videos were downsampled to 1 frame per second, from which 2D skeletal keypoint sequences were extracted using a pre-trained ST-GCN++ model [27]. Spatial normalization was applied to mitigate anthropometric and positional biases: joint coordinates were scaled based on limb-length ratios and aligned to a hip-centered coordinate system. The resulting skeletal sequences served for anomaly action recognition.

Training Data Annotation. We constructed video-level multi-instance anomaly action annotation for training the weakly-supervised action recognition module. Through a systematic review of the knowledge of ADHD-related motor dysfunction and clinical video analysis, we defined 6 common pathological hyperactive actions associated with ADHD: *frequently changing posture*, *wiggling body*, *shaking legs or feet*, *constantly shifting in the seat*, *looking around*, and *touching the head*. Automated annotation was performed using the Qwen2.5-VL-7B-Instruct large vision-language model (LVLM) [24]: video clips were input into this LVLM, and predefined anomaly actions were detected through multi-round question answering to generate training labels. Finally, a clinical expert team manually reviewed and refined the labels to ensure annotation accuracy.

2.2 Explainable ADHD Diagnosis Framework

Weakly-supervised Action Recognition The weakly-supervised action recognition module aims to perform feature encoding on human pose sequences while achieving anomaly action recognition and localization. Given a skeletal sequence $X \in \mathbb{R}^{T\times D}$ extracted from action videos, where T denotes the number of timesteps and D represents the feature dimension, the module addresses interference from irrelevant motions (e.g., static poses or non-ADHD-related movements) during testing. Inspired by [10], we introduce an *Activity Segment Proposal (ASP)* module to locate temporally active segments. The ASP module first extracts pose features using an encoder g_θ:

$$\mathbf{f} = g_\theta(X) \tag{1}$$

where $\mathbf{f} \in \mathbb{R}^{T\times d}$ denotes the encoded pose feature sequence. These features are then projected into a $T \times 2$ activation map:

$$\alpha^{act} = \text{Proj}(\mathbf{f}), \quad \alpha^{act} \in \mathbb{R}^{T\times 2} \tag{2}$$

where $\alpha^{act}_{i,0}$ and $\alpha^{act}_{i,1}$ represent the activation for the presence/absence of activities at the i-th timestep, respectively.

Intuitively, one could directly normalize each a_i using softmax to convert activation scores into probability distributions. However, this approach disproportionately emphasizes strongly activated segments while neglecting marginally activated timesteps, resulting in fragmented and incomplete activity proposals. To address this, we employ *Gumbel-Softmax* [12] to inject controlled stochasticity during sampling, generating complementary activity/no-activity proposals:

$$[\mathbf{P}^{\text{Act}}_i, \mathbf{P}^{\text{NoAct}}_i] = \text{Gumbel-Softmax}([a_{i,0}, a_{i,1}]), \quad \forall i \in \{1, \ldots, T\} \tag{3}$$

Crucially, Gumbel-Softmax enables differentiable gradients through soft sampling during training while switching to hard sampling for deterministic inference:

$$\mathbf{P}^{\text{Act}}_i = \text{argmax}([a_{i,0}, a_{i,1}]), \quad \forall i \in \{1, \ldots, T\} \tag{4}$$

Using $\mathbf{P}^{\text{Act}}$, we filter out timesteps containing static poses or normal motions from the pose feature sequence, enabling precise anomaly action analysis. We then introduce an *Anomaly Activation Network (AAN)* composed of multilayer perceptrons (MLPs) to predict timestep-wise anomaly activations from the masked features:

$$\alpha^{ano} = \text{AAN}(\mathbf{P}_{Act} \odot \mathbf{f}) \tag{5}$$

where $\alpha^{ano} \in \mathbb{R}^{T\times C}$ denotes the anomaly activation matrix, and C represents the number of anomaly action categories. Each element $\alpha^{ano}_{i,c}$ indicates the activation logit for anomaly action c at timestep i, providing interpretable evidence about whether anomaly c occurs at temporal segment i.

On this basis, we compute video-level anomaly scores $\mathbf{s} \in \mathbb{R}^C$ by aggregating temporal anomaly activations across activity proposals. The scoring function is

defined as:

$$s_c = \sigma\left(\frac{\sum_{i=1}^{T} P_i^{\text{Act}} \cdot \alpha_{i,c}^{ano}}{\mathcal{T}_c}\right), \quad c = 1, \ldots, C \tag{6}$$

where $\sigma(\cdot)$ denotes the sigmoid activation function, $\mathcal{T}_c$ is a learnable temperature parameter for class c. $s_c \in [0, 1]$ represents the predicted probability of anomaly class c occurring in the video, which enables the identification of present anomalies within the video segment through thresholding.

ADHD Diagnosis To enhance diagnostic accuracy, we integrate the anomaly score vector $\mathbf{s} \in \mathbb{R}^C$ with standardized executive function test metrics $\mathbf{r} \in \mathbb{R}^M$ through a lightweight multilayer perceptron (MLP) classifier. The ADHD diagnosis probability is formulated as:

$$p = \text{MLP}\left(\text{concat}(\mathbf{s}, \mathbf{r})\right) \tag{7}$$

Optimization We employ an end-to-end multi-task learning framework to jointly optimize action recognition and diagnostic tasks. The composite loss function combines two binary cross-entropy (BCE) components:

$$\mathcal{L} = \mathcal{L}_{\text{diag}} + \lambda \mathcal{L}_{\text{action}} \tag{8}$$

where:

$$\begin{aligned} \mathcal{L}_{\text{diag}} &= \text{BCE}(p, y^{diag}) \\ \mathcal{L}_{\text{action}} &= \sum_{c=1}^{C} \text{BCE}(s_c, y_c^{action}) \end{aligned} \tag{9}$$

λ is the task weighting hyperparameter, y^{diag} is the Ground-truth diagnosis label, and y_c^{action} is the binary indicator for presence of anomaly action c.

The joint training of diagnosis and action recognition facilitates mutual knowledge transfer: the weakly-supervised module detects disease-specific anomalies to supply diagnostic evidence, while diagnostic gradients guide the action network to focus on clinically salient patterns. This synergy enhances feature discriminability and interpretability while reducing overfitting via shared feature extraction. Experiments show this synergy improves accuracy and provides clinicians with interpretable decisions supported by anomaly-action correlations and clinical metrics.

3 Experiments

3.1 Dataset

We collected test records from 441 subjects in real-world clinical settings, comprising 324 ADHD children and 117 typically developing controls. Participants

Table 1. Comparison of ADHD diagnosis performance between different methods. T and A represent using executive function test and action information, respectively.

Method	Type	Accuracy	Recall	Precision	F1-score
SVM [3]	T	0.684	0.504	0.508	0.506
Decision Tree [5]	T	0.722	0.643	0.646	0.645
Random Forest [23]	T	0.692	0.518	0.535	0.526
Logistic Regression [23]	T	0.714	0.568	0.606	0.586
MLP [8]	T	0.741	0.922	0.766	0.837
bi-LSTM [28]	A	0.849	0.929	0.868	0.897
Bert [28]	A	0.853	0.910	0.887	0.898
ADR [17]	A	0.805	0.894	0.840	0.866
ADR-T [17]	T+A	0.797	0.904	0.825	0.863
AVEN [21]	A	0.867	0.917	0.899	0.908
AVEN-T [21]	T+A	0.871	0.925	0.898	0.911
Bert* [28]	T+A	0.916	0.905	0.885	0.895
EDWAR (Ours)	T+A	**0.943**	**0.950**	**0.915**	**0.932**

aged 6–12 years with a male-to-female ratio of 3:1 completed assessments lasting 19.5 ± 4.5 minutes. The videos were segmented into 1-minute clips, yielding 8,608 valid clips (ADHD: 6,514, control: 2,094). All clips were preprocessed as described in Section 2.1 to generate normalized skeletal sequences and anomaly action annotations.

To ensure rigorous evaluation, we implemented 5-fold stratified cross-validation, preserving original class distribution ratios in each fold through stratified random sampling.

3.2 Implementation Details

The experiments were conducted on an NVIDIA A100 80GB PCIe GPU. Input skeleton sequences were padded to 60 frames using the COCO skeleton format (17 keypoints with 2D coordinates) [18]. The Adam optimizer was employed with a learning rate of $1e$-4, and Binary Cross-Entropy (BCE) loss was used for training. Evaluation metrics included accuracy, recall, precision, and F1-score for both diagnosis and action detection results.

3.3 Results

ADHD Diagnosis. To systematically validate the clinical efficacy of the EDWAR framework, we conducted comparative analyses with state-of-the-art AI-assisted diagnostic methods that integrate clinical test metrics and hyperactivity behavior analysis. This includes approaches employing traditional machine learning models (SVM, Decision Tree, Random Forest, Logistic Regression, and MLP) for clinical metric evaluation [3, 5, 23], as well as temporal pattern recognition methods using bi-LSTM and BERT architectures to capture sequential behavioral characteristics in patient actions [7, 11, 28]. Additionally, we included two

Table 2. Ablation study results of EDWAR framework components.

AR mode	WSAR	TM	$\mathcal{L}_{Action}$	Accuracy	Recall	Precision	F1-score
Co-Learning				0.921	0.905	0.897	0.901
		✓		0.916	0.905	0.885	0.895
			✓	0.918	0.916	0.889	0.902
		✓	✓	0.941	0.931	0.916	0.923
	✓			0.938	0.933	0.920	0.926
	✓	✓		0.943	0.926	**0.923**	0.924
	✓		✓	0.925	0.909	0.905	0.907
	✓	✓	✓	**0.943**	**0.950**	0.915	**0.932**
Pre-trained AR			✓	0.767	0.500	0.383	0.434
		✓	✓	0.406	0.580	0.665	0.620
	✓		✓	0.767	0.500	0.383	0.434
	✓	✓	✓	0.842	0.684	0.849	0.758

further comparative baselines: ADR group uses skeleton extraction and abnormal motion detection techniques from video data, similar to our approach[17]; AVEN group refers to a recent study that presents an effective ADHD diagnostic method with promising results[21]. Furthermore, we developed a BERT-based hybrid-modal baseline that synergistically combines sequential action recognition models with MLP-based clinical metric processing, establishing an integrated framework for cross-modal diagnostic validation.

The results demonstrate that the multimodal integration of test metrics and action methods yields significantly better performance than single-modality approaches. Notably, EDWAR outperforms the mixed-modality baseline BERT* across all metrics, confirming that its improvement stems from enhanced ADHD-specific action pattern recognition and mutual knowledge transfer between modalities rather than mere information augmentation. Moreover, the results of additional comparative experiments, including the ADR and AVEN methods, consistently demonstrate lower performance compared to EDWAR, further validating the superior efficacy of our model.

Ablation Study. As illustrated in Table 2, we conducted ablation studies to evaluate the contributions of three key components in the EDWAR framework: the Weakly-Supervised Action Recognition module, execution functional test metrics, and the Anomaly Action Recognition Loss $\mathcal{L}_{Action}$. Specifically, for the *WSAR* module, we replaced it with a baseline BERT model lacking components of Activity Segment Proposal (ASP) and Anomaly Activity Network (AAN); For test metrics (*TM*), we directly used raw predicted action scores for diagnostic prediction; For $\mathcal{L}_{Action}$, we retained only the diagnostic loss $\mathcal{L}_{Diag}$ during training. Additionally, we investigated whether using pre-trained action recognition models (rather than collaborative learning) could achieve effective ADHD diagnosis, comparing it against our joint action recognition and diagnostic prediction framework (*Co-Learning* vs *Pre-trained*).

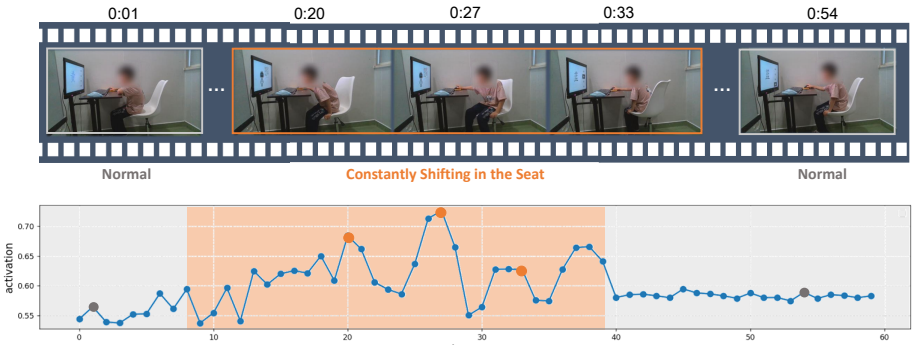

Fig. 2. Explainability Illustration example of *constantly shifting in the seat* action.

The experimental results reveal three critical insights: First, the proposed weakly-supervised action recognition framework consistently enhances diagnostic performance. This demonstrates that leveraging ASP and AAN for timestep-wise activation learning effectively mitigates interference from non-activity intervals and cross-category anomalies, thereby improving ADHD diagnosis accuracy. Second, while neither test metrics nor action recognition constraints alone improved performance, their combined integration significantly boosted model effectiveness. Notably, directly applying pre-trained anomaly action detection models yielded poor diagnostic results. This starkly highlights the necessity of our collaborative learning framework. When trained in isolation, action recognition models may overfit to ADHD-irrelevant action activities, failing to capture disorder-specific activation patterns critical for diagnosis.

3.4 Explainability

As demonstrated in Fig. 2, we illustrate how EDWAR leverages weakly-supervised action detection results to provide interpretable evidence supporting diagnostic conclusions. The upper panel displays a video segment of a child diagnosed with ADHD during executive function testing, capturing their postural dynamics. The lower panel visualizes EDWAR's timestamp-wise activation predictions for the *"constantly shifting in the seat"* anomaly action category. The background-highlighted segments indicate periods of anomalous behaviors, with color-coded markers aligned to corresponding video frames.

Notably, the elevated anomaly activation scores between 8–39s are closely matched clinician-annotated intervals of abnormal movements. High-activation phases correlate with predefined anomaly criteria (e.g., frequent seat shifting), while low-activation periods reflect normative behaviors.

In clinical practice, EDWAR enhances diagnostic transparency by delivering timestamp-wise activation maps and aggregated anomaly scores, enabling clinicians to trace evidence directly to specific behavioral episodes. This capability reduces manual review time by rapidly localizing high-probability anomalies

and standardizes interpretation by aligning automated predictions with expert judgments.

4 Conclusion

This study proposes EDWAR, an explainable diagnostic framework for ADHD that innovatively integrates weakly-supervised action recognition with clinical test analysis through collaborative learning. The framework demonstrates superior diagnostic accuracy (94.3%) while providing multi-granularity clinical evidence from video-level anomaly scoring to temporally localized action activations effectively bridging the gap between algorithmic decisions and clinical reasoning.

Future research will extend this paradigm to other neurodevelopmental disorders (e.g., ASD or depression) through multimodal expansion incorporating eye-tracking and facial expression analysis, as well as cross-disorder knowledge transfer mechanisms for generalized behavioral understanding. These developments aim to establish a new standard for transparent, evidence-based AI diagnostics in developmental psychiatry.

Acknowledgments. This work was supported in part by National Key R & D Program of China (2022YFF1202400) and the Earth System Big Data Platform of the School of Earth Sciences, Zhejiang University.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Amado-Caballero, P., Casaseca-de-la Higuera, P., Alberola-Lopez, S., Andres-de Llano, J.M., Villalobos, J.A.L., Garmendia-Leiza, J.R., Alberola-Lopez, C.: Objective adhd diagnosis using convolutional neural networks over daily-life activity records. IEEE journal of biomedical and health informatics **24**(9), 2690–2700 (2020)
2. Bench, C., Frith, C., Grasby, P., Friston, K., Paulesu, E., Frackowiak, R., Dolan, R.J.: Investigations of the functional anatomy of attention using the stroop test. Neuropsychologia **31**(9), 907–922 (1993)
3. Bledsoe, J.C., Xiao, C., Chaovalitwongse, A., Mehta, S., Grabowski, T.J., Semrud-Clikeman, M., Pliszka, S., Breiger, D.: Diagnostic classification of adhd versus control: support vector machine classification using brief neuropsychological assessment. Journal of attention disorders **24**(11), 1547–1556 (2020)
4. Brown, M.R., Sidhu, G.S., Greiner, R., Asgarian, N., Bastani, M., Silverstone, P.H., Greenshaw, A.J., Dursun, S.M.: Adhd-200 global competition: diagnosing adhd using personal characteristic data can outperform resting state fmri measurements. Frontiers in systems neuroscience **6**, 69 (2012)
5. Chen, T., Antoniou, G., Adamou, M., Tachmazidis, I., Su, P.: Automatic diagnosis of attention deficit hyperactivity disorder using machine learning. Applied Artificial Intelligence **35**(9), 657–669 (2021)
6. Dalgleish, T., Power, M.: Handbook of cognition and emotion. John Wiley & Sons (2000)

7. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). pp. 4171–4186 (2019)
8. Duda, M., Haber, N., Daniels, J., Wall, D.: Crowdsourced validation of a machine-learning classification system for autism and adhd. Translational psychiatry **7**(5), e1133–e1133 (2017)
9. Edition, F., et al.: Diagnostic and statistical manual of mental disorders. Am Psychiatric Assoc **21**(21), 591–643 (2013)
10. Huang, J., Kong, M., Chen, L., Liang, T., Zhu, Q.: Temporal rpn learning for weakly-supervised temporal action localization. In: Asian Conference on Machine Learning. pp. 470–485. PMLR (2024)
11. Huang, Z., Xu, W., Yu, K.: Bidirectional lstm-crf models for sequence tagging. arXiv preprint arXiv:1508.01991 (2015)
12. Jang, E., Gu, S., Poole, B.: Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144 (2016)
13. Kao, G.S., Thomas, H.M.: Test review: C. keith conners conners 3rd edition toronto, ontario, canada: Multi-health systems, 2008. Journal of Psychoeducational Assessment **28**(6), 598–602 (2010)
14. Kim, W.P., Kim, H.J., Pack, S.P., Lim, J.H., Cho, C.H., Lee, H.J.: Machine learning–based prediction of attention-deficit/hyperactivity disorder and sleep problems with wearable data in children. JAMA network open **6**(3), e233502–e233502 (2023)
15. Lange, K.W., Reichl, S., Lange, K.M., Tucha, L., Tucha, O.: The history of attention deficit hyperactivity disorder. ADHD Attention Deficit and Hyperactivity Disorders **2**, 241–255 (2010)
16. Leo, M., Carcagnì, P., Distante, C., Spagnolo, P., Mazzeo, P.L., Rosato, A.C., Petrocchi, S., Pellegrino, C., Levante, A., De Lumè, F., et al.: Computational assessment of facial expression production in asd children. Sensors **18**(11), 3993 (2018)
17. Li, Y., Nair, R., Naqvi, S.M.: Video-based skeleton data analysis for ADHD detection. SSCI (2023)
18. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. ECCV (2014)
19. Monchi, O., Petrides, M., Petre, V., Worsley, K., Dagher, A.: Wisconsin card sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. Journal of Neuroscience **21**(19), 7733–7741 (2001)
20. Nash, C., Nair, R., Naqvi, S.M.: Machine learning in adhd and depression mental health diagnosis: a survey. IEEE Access **11**, 86297–86317 (2023)
21. Nash, C., Nair, R., Naqvi, S.M.: Insights into detecting adult adhd symptoms through advanced dual-stream machine learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering (2024)
22. Song, P., Zha, M., Yang, Q., Zhang, Y., Li, X., Rudan, I.: The prevalence of adult attention-deficit hyperactivity disorder: A global systematic review and meta-analysis. Journal of global health **11**, 04009 (2021)
23. Tachmazidis, I., Chen, T., Adamou, M., Antoniou, G.: A hybrid ai approach for supporting clinical diagnosis of attention deficit hyperactivity disorder (adhd) in adults. Health Information Science and Systems **9**(1), 1 (2020)

24. Team, Q.: Qwen2.5-vl (January 2025), https://qwenlm.github.io/blog/qwen2.5-vl/
25. Thomas, R., Sanders, S., Doust, J., Beller, E., Glasziou, P.: Prevalence of attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. Pediatrics **135**(4), e994–e1001
26. Wang, L., Xiong, Y., Lin, D., Van Gool, L.: Untrimmednets for weakly supervised action recognition and detection. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 4325–4334 (2017)
27. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
28. Zhang, Y., Kong, M., Zhao, T., Hong, W., Xie, D., Wang, C., Yang, R., Li, R., Zhu, Q.: Auxiliary diagnostic system for adhd in children based on ai technology. Frontiers of Information Technology & Electronic Engineering **22**(3), 400–414 (2021)
29. Zhang, Y., Kong, M., Zhao, T., Hong, W., Zhu, Q., Wu, F.: Adhd intelligent auxiliary diagnosis system based on multimodal information fusion. In: Proceedings of the 28th ACM International Conference on Multimedia. pp. 4494–4496 (2020)