Proceedings of the 7th International Conference on Education Innovation and Philosophical Insights
DOI: 10.54254/2753-7048/2026.38323

Research on Identifying and Avoiding Overdiagnosis of ADHD in Students

Taoqi Wang

Beijing RCF Dongba School, Beijing, China
18601057933@163.com

Abstract. Attention Deficit Hyperactivity Disorder (ADHD) is a common neurodevelopmental disorder that affects the physical and mental development. The traditional recognition method relying on scale evaluation is easily influenced by subjective factors, and there is a risk of misjudgment and overdiagnosis. This paper proposes a student ADHD assisted recognition method based on the fusion of eye movement and facial expression temporal features. This article uses the eye tracking MPIIFaceGaze dataset and facial CK+dataset to construct an EFF-ARM recognition model using Transformer network. The experiment adopts a cross student testing method and compares it with CNN and RNN models for analysis. The results showed that the accuracy, precision, and recall of the EFF-ARM model reached 86.42%, 85.76%, and 85.91%, respectively, and the overall recognition performance was better than the comparison model. Finally, this article implemented a system platform for identifying and avoiding overdiagnosis of ADHD in students, which achieved feature analysis and risk result display, providing an auxiliary analysis method based on objective behavioral data for early screening of ADHD in students.

Keywords: ADHD recognition, Transformer, Multimodal fusion, Overdiagnosis

1. Introduction

The main manifestations of ADHD include difficulty maintaining attention and impulsive behavior. ADHD has a high incidence rate in children and adolescents and has gradually become a new problem in education and medicine. The prevalence of ADHD in children and adolescents in China is as high as 6.3%, and about two-thirds of children still have ADHD symptoms in adulthood [1]. ADHD may show significant differences when assessed using different self-assessment scales, making it necessary to combine standardized assessments for evaluation [2]. Identifying ADHD also requires attention to differences in performance among different groups to reduce misdiagnosis caused by evaluation bias. However, the recognition process of ADHD still faces significant challenges, and the issue of overdiagnosis is gradually attracting researchers' attention. This article provides a new approach for ADHD recognition using AI algorithms, with eye tracking, facial features, behavioral task performance, and other criteria gradually being applied for auxiliary analysis [3]. Therefore, this article attempts to explore more effective ADHD recognition pathways through multimodal fusion, hoping to provide practical reference for ADHD.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.38323

2. Research methods

2.1. Multimodal data representation model

This paper constructs a multimodal fusion model based on eye movement and facial expression time-series data for student ADHD recognition tasks. The overall structure of the model includes a data acquisition layer, a feature extraction layer, a feature fusion layer, and a recognition output layer [4].

Eye tracking data is represented as:

$$X_e = \{x_1^e, x_2^e, \ldots, x_T^e\} \tag{1}$$

Among them, $x_t^e \in R^{d_e}$, which represents the eye movement feature vector at the t-th time point, including fixation position, fixation time, eye movement distance, eye movement speed, etc.

Facial expression data is represented as:

$$X_f = \{x_1^f, x_2^f, \ldots, x_T^f\} \tag{2}$$

Among them $x_t^f \in R^{d_f}$, which representing facial behavioral features corresponding to time nodes, including facial keypoints, facial expression changes, head posture, behavioral states, etc.

2.2. Transformer time series feature extraction algorithm

2.2.1. Input embedding

Firstly, map the input data to a unified feature space. Assuming the input behavior sequence is: $X = \{x_1, x_2, \ldots, x_n\}$. Among, n represents the length of the time series, and x_i represents the behavioral features extracted. After linear mapping, the input feature representation is obtained:

$$E = XW_e + b_e \tag{3}$$

Among them, W_e and b_e represent mapping weights and bias parameters, respectively.

2.2.2. Self attention calculation

Due to the fact that Transformer itself cannot obtain the sequential relationship between different positions in a sequence, position encoding information needs to be added:

$$Z = E + PE \tag{4}$$

Among them, PE represents the position encoding matrix.The Transformer is the multi head self attention mechanism, which first obtains the Query, Key, and Value through linear transformation:

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V \tag{5}$$

Then use scaled dot product attention to calculate the degree of correlation between different features:

DOI: 10.54254/2753-7048/2026.38323

$$\mathrm{Attention}\left(\mathrm{Q},\mathrm{K},\mathrm{V}\right)=\mathrm{Softmax}\left({}^{\mathrm{QK}^{\mathrm{T}}}\!/_{\sqrt{\mathrm{d_k}}}\right)\mathrm{V} \tag{6}$$

Among them, d_k represents the dimension of the key vector.

2.2.3. Multi head attention mechanism

In order to enhance the learning ability of the model for different feature space information, Transformer adopts a multi head attention mechanism. The calculation formula is as follows:

$$\mathrm{MultiHead}\left(\mathrm{Q},\mathrm{K},\mathrm{V}\right)=\mathrm{Concat}\left(\mathrm{head}_1,\ldots,\mathrm{head}_\mathrm{h}\right)\mathrm{W}^{\mathrm{O}} \tag{7}$$

Among them,

$$\mathrm{head_i}=\mathrm{Attention}\left(\mathrm{QW}_\mathrm{i}^{\mathrm{Q}},\mathrm{KW}_\mathrm{i}^{\mathrm{K}},\mathrm{VW}_\mathrm{i}^{\mathrm{V}}\right) \tag{8}$$

The output result is processed through residual connection and normalization:

$$\mathrm{Z}^{'}=\mathrm{Layer\ Norm}\left(\mathrm{Z}+\mathrm{MultiHead}\left(\mathrm{Q},\mathrm{K},\mathrm{V}\right)\right) \tag{9}$$

Then input the feedforward neural network for further nonlinear feature transformation:

$$\mathrm{FFN}(\mathrm{x})=\sigma(\mathrm{xW}_1+\mathrm{b}_1)\mathrm{W}_2+\mathrm{b}_2 \tag{10}$$

The eye and facial features are as follows:

$$H_e=Transformer(X_e),H_f=Transformer(X_f) \tag{11}$$

2.2.4. Multimodal attention fusion algorithm

Firstly, concatenate the two modal features,Then calculate the modal weights:

$$\mathrm{H_c}=\left[\mathrm{H_e};\mathrm{H_f}\right],\qquad \alpha_\mathrm{i}=\frac{\exp(\mathrm{W_i};\mathrm{H_i})}{\sum_{\mathrm{j}=1}^{\mathrm{M}}\exp(\mathrm{W_j};\mathrm{H_j})} \tag{12}$$

Among them, M=2 represents two modes; α_e is the eye movement weight; α_f is the facial weight.

Calculate category probability using Softmax function:

$$P(y=i|x)=\frac{e^{z_i}}{\sum_{j=1}^{C}e^{z_j}},Y=argmax(P(y)) \tag{13}$$

Among them, C=2 represents two types of results; Y=0: Normal student; Y=1: ADHD risk students.

2.2.5. Final mathematical model summary

Therefore, this article proposes a student ADHD recognition model, which can be expressed as:

DOI: 10.54254/2753-7048/2026.38323

$$Y = \mathrm{Softmax}(W_c(\alpha_e \mathrm{Transformer}(X_e) + \alpha_f \mathrm{Transformer}(X_f) + b_c)) \quad (14)$$

Among them, $\mathrm{Transformer}(X_e)$ is extracting temporal features of eye; $\mathrm{transformer}(X_f)$ is extracting temporal features of facial; α_e, α_f are multimodal attention weights; W_c, b_c are classification parameters.

2.3. Evaluation indicators

This article uses accuracy, precision, and recall as performance evaluation metrics for the model.

$$A=\frac{TP+TN}{TP+TN+FP+FN}, \quad P=\frac{TP}{TP+FP}, \quad R=\frac{TP}{TP+FN} \quad (15)$$

3. Research results

3.1. ADHD identification process

In the process of identifying ADHD in students, a single behavioral indicator often fails to accurately reflect an individual's attention status. This article utilizes eye movement information and facial behavior information from students completing attention tasks to establish a multimodal data-driven ADHD assisted recognition process [5]. The overall process mainly includes five steps: data collection, data processing, feature extraction, feature fusion, and classification judgment, as shown in Figure 1.

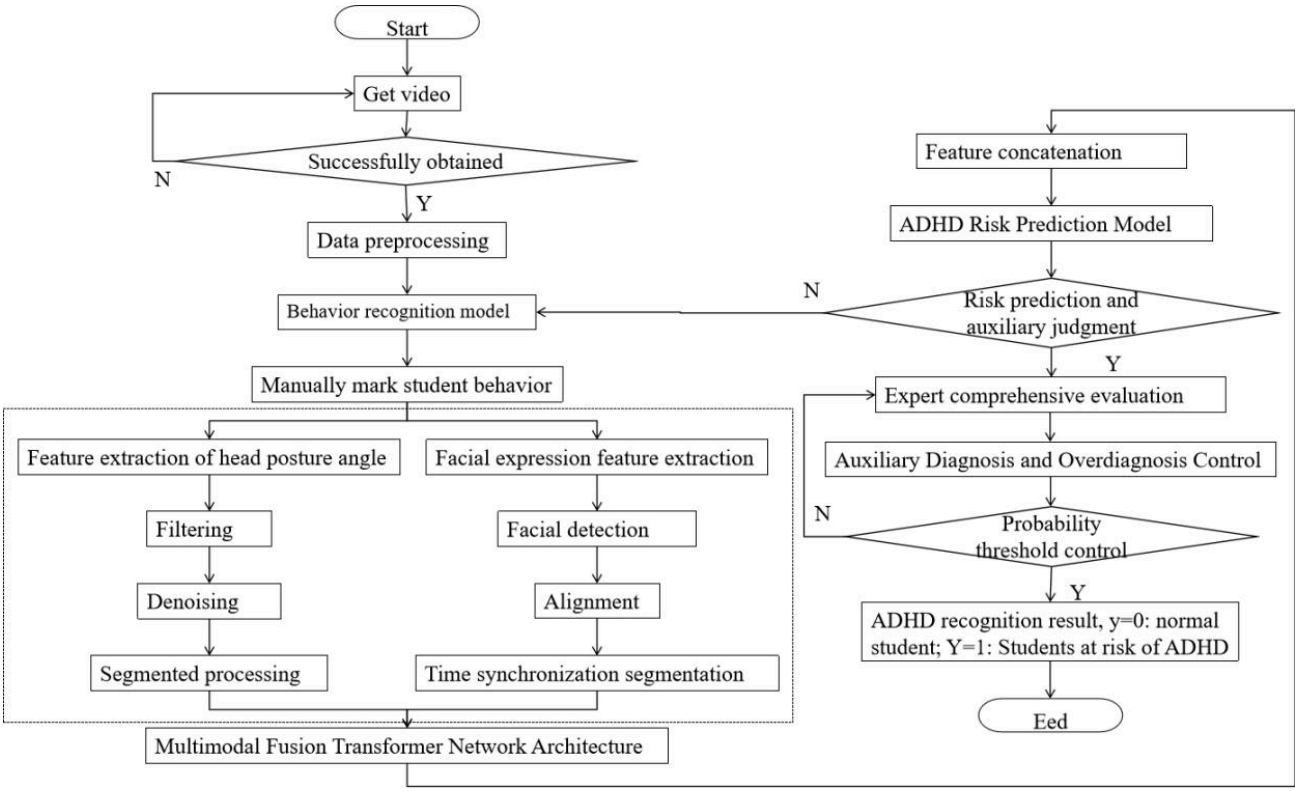


Figure 1. ADHD identification and avoidance of overdiagnosis process

DOI: 10.54254/2753-7048/2026.38323

3.2. ADHD training test set experimental results

During the experiment, CNN, RNN, and the EFF-ARM model proposed in this paper were selected for comparison. The experimental evaluation adopts three indicators: accuracy, precision, and recall. The experimental results are shown in Table 1.

Table 1. Recognition results of different models on cross student test sets

Models	Accuracy/%	Precision/%	Rrecall/%
CNN	78.36	77.54	76.82
RNN	81.25	80.63	80.47
EFF-ARM	86.42	85.76	85.91

Table 3.1 shows that the EFF-ARM model achieves the highest results in three indicators, because eye movement data can reflect changes in visual attention, and facial data can supplement the status in task execution, which can reduce the problem of insufficient information in a single mode.

At the same time, this article used a five fold cross validation method for testing, with each experiment changing the test student sample and the remaining data used for model training [6]. A total of five sets of experimental results were obtained, as shown in Table 2.

Table 2. Five fold cross validation results of EFF-ARM model

Experimental group	Accuracy/%	Precision/%	Rrecall/%
Fold1	85.73	84.96	85.21
Fold2	86.58	85.92	86.14
Fold3	87.06	86.37	86.72
Fold4	85.94	85.31	85.67
Fold5	86.79	86.02	86.41
Average value	86.42	85.72	86.03

From Table 2, it can be seen that the accuracy of EFF-ARM remained above 85% in all five experiments, indicating that the model still has a relatively stable performance. Overall, the EFF-ARM model performs better than CNN and RNN models in cross student recognition experiments.

3.3. System platform implementation

This article develops an auxiliary screening system platform based on the recognition needs of students with ADHD. The overall structure of the system is shown in Figure 2.

The system mainly consists of five parts: data collection, data processing, feature analysis, model recognition, and result display. The modules are connected according to the data flow direction, completing the entire process from obtaining raw behavioral information to outputting recognition results.

DOI: 10.54254/2753-7048/2026.38323



Figure 2. ADHD recognition and overdiagnosis avoidance system platform

4. Conclusion

This article focuses on the subjective evaluation bias and overdiagnosis issues in students' ADHD awareness, proposes an ADHD recognition model based on EFF-ARM, and designs a corresponding platform to form the following conclusions.

(1) This article constructs an EFF-ARM model, extracts eye movement and facial temporal feature data through Transformer, and uses attention fusion algorithm to achieve multimodal fusion.

(2) TThe experimental results show that the accuracy, precision, and recall of the EFF-ARM model reach 86.42%, 85.76%, and 85.91%, respectively, with better performance than CNN and RNN models.

(3) This article implements a system platform for identifying and avoiding overdiagnosis of ADHD in students. By collecting information on students' eye movements and facial behavior, it assists in analyzing their attention status and outputs risk warning results.

References

[1] Brandt, L., & Fischer, G. (2019). Adult ADHD is associated with gambling severity and psychiatric comorbidity among treatment-seeking problem gamblers. *Journal of Attention Disorders, 23*(12), 1383-1395. https://doi.org/10.1177/1087054717690232.

[2] Lee, N. Y. W., & Zhang, M. W. B. (2025). Systematic review on prevalence of ADHD, possible ADHD or ADHD symptoms in medical students. *Frontiers in Psychiatry, 16*, 1684727. https://doi.org/10.3389/fpsyt.2025.1684727.

[3] Brunetti, C., Clair, C., Schlueter, V., Eggimann, A., & Candaux, F. (2025). ADHD in adults and gender differences: Identification by general practitioners. *Revue Médicale Suisse, 21*(938), 1970-1975. https://doi.org/10.53738/REVMED.2025.21.938.47825.

[4] Gelashvili, A., Resheff, Y. S., & Blumrosen, G. (2026). ADHD identification from real-life recorded video. *Scientific Reports, 16*(1), 21482. https://doi.org/10.1038/s41598-026-52068-8.

[5] Yu, R., Zhang, X., & Chen, Y. (2025). Graph neural networks for ADHD identification in children. *Studies in Health Technology and Informatics, 329*, 1774-1775. https://doi.org/10.3233/SHTI251208.

[6] Bogdańska-Chomczyk, E., Majewski, M. K., & Kozłowska, A. (2025). ADHD in adulthood: Clinical presentation, comorbidities, and treatment perspectives. *International Journal of Molecular Sciences, 26*(22), 11020. https://doi.org/10.3390/ijms262211020.