ISSN 2522-6398 Vol. 9, Issue 6: 97-106, DOI: 10.25236/FER.2026.090613

Interpretable Multimodal LLM Evaluation Model for Comprehensive Development of Young Children Based on Multi-Source Structured Data

Xiaowen Wang[1,a,*], Wanyi Huang[1,b], Xiao Liu[1,c], Zhengfen Teng[1,d], Yaoyao Cai[1,e]

[1]*College of Education (Shanwei), South China Normal University, Shanwei, 516625, Guangdong, China*
[a]*wangxw@m.scnu.edu.cn*, [b]*20228221140@m.scnu.edu.cn*, [c]*18156733856@163.com*,
[d]*1903824420@qq.com*, [e]*2590732175@qq.com*
[*]*Corresponding Author*

Abstract: *To address the problems of traditional early childhood development assessments, such as over-reliance on teachers' subjective observations, limited data sources, and difficulty in reflecting children's continuous growth, this paper proposes an interpretable multimodal large language model (LLM) method for early childhood development assessment based on multi-source structured data. A multimodal feature fusion mechanism is designed, and an early childhood development assessment model is established through cross-modal alignment, dynamic weight allocation, and time-series information integration. Furthermore, attention mechanisms, feature contribution analysis, and inference chain interpretation techniques are introduced to enhance the transparency and credibility of the model's assessment results. Combined with dynamic modeling of developmental trajectories and growth trend analysis, the model achieves process-oriented assessment and prediction of children's comprehensive development. Experimental results show that the proposed model achieves 95.1% accuracy and 94.1% F1-score in early childhood development assessment tasks, representing improvements of 2.8% and 2.8% compared to GPT-4o, and improvements of 4.9% and 5.2% compared to the multimodal Transformer, respectively. These results demonstrate that the proposed interpretable multimodal LLM assessment model can achieve accurate assessment, reliable interpretation, and dynamic tracking of children's comprehensive development, providing a new technical path for personalized education support and scientific decision-making in early childhood.*

Keywords: *Early Childhood Development Assessment; Early Childhood Development Model; Multidimensional Indicator System; Artificial Intelligence Assessment Method; Data-Driven Analysis*

1. Introduction

Early childhood is a critical period for the formation of an individual's cognitive, language, social, and emotional abilities. A scientific and comprehensive assessment of children's developmental levels is crucial for developing personalized education plans and promoting their healthy growth. With the rapid development of digital education and artificial intelligence technologies, kindergartens have accumulated a wealth of data in their daily educational activities, including teacher observation records, behavioral videos, language and speech recordings, learning materials, and feedback from home and school collaborations. This provides a rich data foundation for the comprehensive development assessment of children.

To address the aforementioned issues, this paper proposes an interpretable multimodal large language model (LLM) method for assessing the comprehensive development of young children based on multi-source structured data. Through dynamic modeling of developmental trajectories, analysis of multi-stage growth characteristics, and representation of individual change trends, the method achieves process-oriented assessment and trend prediction of children's comprehensive development. Experimental results show that the proposed method outperforms existing comparative models in terms of assessment accuracy, interpretability, and developmental trend prediction capabilities, providing a new technical path and practical reference for leveraging artificial intelligence in early childhood education evaluation.

ISSN 2522-6398 Vol. 9, Issue 6: 97-106, DOI: 10.25236/FER.2026.090613

2. Related Work

With the rapid development of technologies such as artificial intelligence, machine learning, and large language models, intelligent education has gradually become an important direction for the digital transformation of education. Among these, utilizing artificial intelligence to analyze the learning process, conduct educational evaluation, and provide personalized support has become a crucial topic in current educational research and practice. Scholars both domestically and internationally have conducted extensive research on intelligent educational evaluation, multi-source data analysis, and personalized educational services, providing a rich theoretical foundation and technical support for empowering educational development with artificial intelligence.

Destriana et al. constructed a digital platform for monitoring early childhood development that integrates machine learning based on the Dynamic Systems Development Method (DSDM) to solve the problems of low efficiency and poor real-time performance of traditional monitoring methods. The platform integrates data entry, growth visualization, predictive analysis, and an interactive dashboard. The machine learning model has an accuracy of 89% in predicting developmental delays [1]. Chen et al. explored the application and optimization of machine learning algorithms in personalized education recommendation systems. This study pointed out that traditional education is difficult to meet the individual differences of students, while machine learning-based education systems can provide personalized learning resources and real-time feedback based on learning history, interests, and abilities, thereby improving learning outcomes [2]. Chen et al. used a systematic literature review method to analyze 134 AIGC educational application studies and explored their current development status and future trends. This study found that the United States is the most active in this field, and the research mainly focuses on topics such as technical performance evaluation, teaching application, learning outcome improvement, advantages and risks analysis, and development prospects [3]. Ouyang et al. systematically reviewed 17 empirical studies to summarize the current application status of AI-driven educational evaluation in STEM education. The study found that AI evaluation is mainly used for academic performance assessment, learning status monitoring, and teaching quality evaluation, and widely adopts algorithms such as machine learning, natural language processing, and deep learning [4]. Luo et al. used machine learning to analyze the online learning behavior of students in blended courses in Chinese universities, proposed a course classification method based on the expectation maximization algorithm, and constructed a random forest prediction model to predict learning performance. The study found that the prediction accuracy was significantly improved after classification according to learning behavior characteristics; high participation and diversified learning behaviors help to improve the performance prediction effect [5]. Wang et al. reviewed the latest application progress of large language models (LLMs) in the field of education, covering multiple aspects such as student and teacher assistance, personalized learning, adaptive teaching, and commercial tools [6].

Existing research still has some shortcomings: on the one hand, most studies mainly focus on learning behavior analysis or single educational scenarios, and there are relatively few studies on comprehensive evaluation of the entire process of early childhood development and multiple dimensions; on the other hand, the ability to integrate multi-source heterogeneous data is insufficient, the collaborative use of multimodal information such as children's behavior, language, teacher observation, and parent feedback is still insufficient, and there is a lack of a unified analytical framework that takes into account the accuracy, interpretability, and predictive ability of development trends of assessment.

3. Method

3.1 Construction of a Multi-Source Structured Data System for Early Childhood Comprehensive Development

3.1.1 Design of an Evaluation Index System for Early Childhood Comprehensive Development

To comprehensively reflect the developmental status of young children, a comprehensive evaluation index system covering five dimensions—language development, cognitive development, social development, emotional development, and motor development—is constructed based on the "Guidelines for Learning and Development of Children Aged 3 - 6." Specifically, language development focuses on language expression and comprehension abilities; cognitive development focuses on observation, memory, and problem-solving abilities; social development focuses on peer interaction and rule awareness; emotional development focuses on emotion recognition and regulation

ISSN 2522-6398 Vol. 9, Issue 6: 97-106, DOI: 10.25236/FER.2026.090613

abilities; and motor development focuses on fine motor skills and physical coordination.

The set of comprehensive development evaluation indicators for young children is defined as follows:

$$I = I_{lan}, I_{cog}, I_{soc}, I_{emo}, I_{mot} \tag{1}$$

Here, (I_{lan}), (I_{cog}), (I_{soc}), (I_{emo}), and (I_{mot}) represent indicators of language, cognitive, social, emotional, and motor development, respectively. The overall development score is expressed as:

$$S = \sum_{i=1}^{n} w_i I_i \tag{2}$$

Where (w_i) is the indicator weight, satisfying:

$$\sum_{i=1}^{n} w_i = 1 \tag{3}$$

3.1.2 Acquisition and Organization of Multi-Source Heterogeneous Data

Let the multi-source data set be:

$$D = D_t, D_v, D_a, D_w, D_f \tag{4}$$

Where (D_t) represents teacher observation text data, (D_v) represents video data, (D_a) represents audio data, (D_w) represents student work data, and (D_f) represents feedback data from parents and students.

3.1.3 Structured Representation of Multimodal Data

To address the issue of significant differences in features among different modalities, we first extract and vectorize features from text, audio, image, and video data. Then, we achieve cross-modal mapping through a unified feature space, constructing a structured data representation framework to provide input for subsequent multimodal large language model inference.

Let the feature representation of the (m)th modality be:

$$F_m = f(D_m) \tag{5}$$

Here, ($f(\cdot)$) represents the encoding function for the corresponding modality.

Then, the features of each modality are mapped to a unified representation space:

$$Z = [F_t; F_v; F_a; F_w; F_f] \tag{6}$$

Where (Z) represents the fused multimodal feature vector, and ($[\cdot]$) represents the feature concatenation operation. The final structured representation of early childhood comprehensive development is formed as follows:

$$R = \phi(Z) \tag{7}$$

Here, ($\phi(\cdot)$) represents the unified structured mapping function, and (R) serves as the input data representation for the subsequent evaluation model.

3.2 Construction of a Multimodal LLM Early Childhood Comprehensive Development Assessment Model

3.2.1 Overall Model Architecture Design

The evaluation model designed in this study consists of a data input layer, a multimodal coding layer, a feature fusion layer, and an LLM inference layer. Its overall architecture is shown in the figure.

(1) Data Input Layer

The data input layer is responsible for receiving multi-source data such as teacher observation records, children's voice recordings, behavioral videos, artwork images, and feedback from parents and children's homes, and performing standardization processing. Let the input data set be:

$$X = X_t, X_a, X_v, X_w, X_f \tag{8}$$

Where (X_t), (X_a), (X_v), (X_w), and (X_f) represent text, speech, video, works, and feedback data, respectively.

(2) Multimodal Coding Layer

ISSN 2522-6398 Vol. 9, Issue 6: 97-106, DOI: 10.25236/FER.2026.090613

Deep semantic features are extracted using corresponding encoders for different modal data:

$$\left\{\begin{matrix} H_t=E_t(X_t) \\ H_a=E_a(X_a) \\ H_v=E_v(X_v) \\ H_w=E_w(X_w) \\ H_f=E_f(X_f) \end{matrix}\right\} \quad (9)$$

Where ($E(\cdot)$) represents the corresponding modal encoder, and (H) represents the high-dimensional feature representation.

(3) Feature Fusion Layer

The fusion layer maps the features of each modality to a unified semantic space and constructs a representation vector for the comprehensive development of young children:

$$H=H_t,H_a,H_v,H_w,H_f \quad (10)$$

$$Z=Fusion(H) \quad (11)$$

Where (Z) represents the fused global development feature vector.

(4) LLM Inference Layer

The fused feature vector is input into the large language model, and a comprehensive evaluation is completed through prompt engineering and domain knowledge constraints:

$$Y=LLM(Z,P) \quad (12)$$

Where (P) represents the evaluation prompt template, and (Y) represents the model output result.

3.2.2 Multimodal Feature Fusion Mechanism

Since different modal data have significant differences in information density, time scale, and expression form, a cross-modal feature fusion mechanism is designed to improve evaluation accuracy.

(1) Cross-modal Alignment

First, the features of each modality are mapped to a unified latent space:

$$\widehat{H}_m=W_mH_m+b_m \quad (13)$$

Where (W_m) and (b_m) represent the mapping matrix and bias parameter, respectively.

Cosine similarity is used to measure semantic consistency between different modalities:

$$Sim(i,j)=\frac{\widehat{H}_i\cdot\widehat{H}_j}{|\widehat{H}_i||\widehat{H}_j|} \quad (14)$$

Semantic alignment is achieved by maximizing cross-modal similarity.

(2) Feature-weighted fusion

Considering the varying importance of different data sources for early childhood development assessment, an attention mechanism is introduced to dynamically calculate weights:

$$\alpha_i=\sum_{k=1}^{n}\exp(e_k) \quad (15)$$

Where:

$$e_i=W^T\widehat{H}_i \quad (16)$$

It represents the importance score of the (i)th modality.

The final fusion is expressed as:

$$Z=\sum_{i=1}^{n}\alpha_i\widehat{H}_i \quad (17)$$

This mechanism can highlight key behavioral characteristics and suppress redundant information interference.

(3) Temporal Information Integration

Early childhood development is continuous and dynamic; therefore, a time window mechanism is used to model the developmental process:

ISSN 2522-6398 Vol. 9, Issue 6: 97-106, DOI: 10.25236/FER.2026.090613

$$Z_t=\sum_{k=1}^{T}\beta_k Z_k \tag{18}$$

Where:

$$\beta_k=\frac{\exp(s_k)}{\sum_{j=1}^{T}\exp(s_j)} \tag{19}$$

It represents the importance weight of different time points.

A representation of early childhood developmental trajectories obtained through temporal integration:

$$G=Z_1,Z_2,\ldots,Z_T \tag{20}$$

This provides a basis for subsequent development trend analysis.

3.2.3 Generation Mechanism of Comprehensive Development Assessment for Preschool Children

After completing the multimodal feature fusion, a comprehensive evaluation result is generated using a large language model combined with knowledge of preschool development domains.

(1) Development Level Identification

The model first predicts the preschooler's level score $S_i=f_i(Z)$ on each developmental dimension. Where (S_i) represents the score of the (i)th developmental dimension. The comprehensive development score is expressed as $S=\sum_{i=1}^{n} w_i S_i$.

Development levels are divided according to the scoring interval:

$$L=\begin{cases}1,S<60\\2,60\leq S<75\\3,75\leq S<90\\4,S\geq 90\end{cases} \tag{21}$$

Where (L) represents the developmental level.

(2) Comprehensive Ability Evaluation

An evaluation prompt, $Prompt=Z,K,R$ is constructed based on the fusion of feature vectors and domain knowledge base. Where (K) represents the early childhood development knowledge base, and (R) represents the evaluation rule set. Subsequently, the comprehensive evaluation result $E=LLM(Prompt)$ is generated by the large language model. The output includes strengths, weaknesses, and growth suggestions.

(3) Developmental Trend Representation

To reflect the early childhood development process, the developmental trend is described using the time series rate of change:

$$T_i=\frac{S_i^{(t)}-S_i^{(t-1)}}{S_i^{(t-1)}} \tag{22}$$

Where (T_i) represents the rate of change in the (i)th dimension.

The comprehensive trend index is defined as:

$$TI=\frac{1}{n}\sum_{i=1}^{n} T_i \tag{23}$$

When TI > 0, it indicates an overall upward trend in development; when TI < 0, it indicates a risk of stagnation or decline in development. Based on the trend index and historical assessment records, the large language model further generates a report analyzing the children's growth trajectory, enabling process assessment and development prediction.

3.3 Evaluation Model Interpretability Mechanism Design

3.3.1 Attention-Based Explanation Method

Multimodal LLM assigns different levels of attention to different modalities and behavioral features when generating evaluation results. By analyzing the model's attention weights, key behaviors, and important features that influence the evaluation results can be identified.

ISSN 2522-6398 Vol. 9, Issue 6: 97-106, DOI: 10.25236/FER.2026.090613

Let the attention weight of the (i)th feature be:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_j)} \tag{24}$$

Here, (e_i) represents the feature importance score.

When the weight of a certain behavioral feature is significantly higher than that of other features, it can be considered an important basis for evaluating early childhood development. Visualizing attention distribution can intuitively display the key behaviors, language expressions, and interactive activities that the model focuses on, providing teachers with a reference for understanding the evaluation results.

3.3.2 Interpretation Method Based on Feature Contribution

To further quantify the impact of each evaluation indicator on the final result, this study uses characteristic contribution analysis to assess the importance of different development dimensions.

Let the comprehensive evaluation result be (S), and the contribution of the (i)th indicator be defined as:

$$C_i = \frac{|w_i S_i|}{\sum_{j=1}^{n} |w_j S_j|} \tag{25}$$

Where (w_i) represents the indicator weight, and (S_i) represents the score of the corresponding dimension.

3.3.3 Interpretation of Evaluation Results Based on the Chain of Thought

During the generation of the evaluation report, the large language model completes reasoning step by step based on the input features and evaluation rules. To enhance the interpretability of the results, this study introduces a chain of thought mechanism to record the main decision-making process of the model from evidence extraction and feature judgment to result generation.

Let the model reasoning process be represented as:

$$R = r_1, r_2, \ldots, r_n \tag{26}$$

Where (r_i) represents the reasoning result of step (i).

3.4 A Process-Based Assessment Mechanism for Early Childhood Comprehensive Development

3.4.1 Dynamic Modeling of Developmental Trajectories

Early childhood development is a continuous evolutionary process with significant temporal correlations between different developmental stages. To characterize the trajectory of early childhood development, the multimodal fusion features at each time point are represented as $Z_t = z_1, z_2, \cdots, z_t$. Here, (z_t) represents the comprehensive developmental feature vector at time (t).

A developmental trajectory model is constructed based on the time series concept:

$$G = Z_1, Z_2, \cdots, Z_T \tag{27}$$

Where (G) represents the complete developmental trajectory of the child, and (T) represents the length of the observation period.

To reflect the relationship between the current developmental state and historical performance, a time decay weight $\omega_t = \frac{exp(\lambda t)}{\sum_{k=1}^{T} exp(\lambda k)}$ is introduced. Here, (λ) is the time decay coefficient. The final dynamic developmental representation is $D = \sum_{t=1}^{T} \omega_t Z_t$.

3.4.2 Analysis of Multi-Stage Developmental Characteristics

Early childhood development exhibits distinct stage-specific characteristics, with different developmental focuses at each stage. To analyze the developmental characteristics of each stage, the observation period is divided into (K) developmental stages:

$$P = P_1, P_2, \cdots, P_K \tag{28}$$

Where (P_k) represents the data set for stage (k).

For each development dimension, the average development level of the stage is calculated:

ISSN 2522-6398 Vol. 9, Issue 6: 97-106, DOI: 10.25236/FER.2026.090613

$$M_k=\frac{1}{N_k}\sum_{i=1}^{N_k} S_i \tag{29}$$

Where (S_i) represents the individual development score within the stage, and (N_k) represents the sample size.

Further calculation is the stage growth rate $R_k=\frac{M_k\text{-}M_{k\text{-}1}}{M_{k\text{-}1}}$. When ($R_k$>0), it indicates that the ability in that stage is continuously improving; when (R_k<0), it indicates that the development speed is slowing down or fluctuating.

3.4.3 Characterization of Individual Development and Change Trends

To further reveal the future developmental direction of young children, this study constructs a developmental trend representation mechanism based on historical assessment results. Let the score sequence of a child on the (i)th developmental dimension be $S_i=s_i^1,s_i^2,\cdots,s_i^T$. A linear trend model is used to describe the process of ability change: where (a) represents the developmental trend coefficient, and (b) represents the initial developmental level.

When a > 0, it indicates that the ability is continuously increasing; when a = 0, it indicates relatively stable development; when a < 0, it indicates a risk of developmental stagnation or decline. Simultaneously, a comprehensive developmental trend index is defined:

$$TI=\frac{1}{n}\sum_{i=1}^{n}\frac{s_i^T\text{-}s_i^1}{s_i^1} \tag{30}$$

Where (n) represents the number of development dimensions.

4. Results and Discussion

4.1 Data Sources and Sample Composition

The data for this study primarily come from long-term developmental records of an early childhood education institution, including multi-source information such as observational data on children's daily behaviors, teacher evaluation data, and periodic developmental assessment data. During data collection, the focus is on core dimensions such as cognitive development, language ability, social development, and motor coordination to ensure the comprehensiveness and representativeness of the data.

The sample includes continuous developmental records of several individual children, spanning multiple developmental stages to reflect the dynamic changes in children's growth. All data has been anonymized to ensure compliance and privacy in data use.

4.2 Data Preprocessing and Labeling

Before using the data, the raw data are first cleaned, including missing value completion, outlier detection, and noise removal, to improve data quality. Subsequently, data from different sources underwent standardized processing to ensure consistency in units and format, facilitating subsequent model input.

During the annotation phase, based on the early childhood development evaluation index system, educational experts and frontline teachers jointly participated in grading and annotating the developmental levels of each dimension. The annotation results are cross-validated for consistency verification to improve the reliability and objectivity of the annotation results.

4.3 Evaluation Indicator Setting

To comprehensively evaluate model performance, this paper constructs an evaluation index system from two aspects: prediction accuracy and stability. Regarding accuracy, the main indicators used are accuracy and mean squared error (MSE) to measure the deviation between the model's predictions and the actual labeled data.

Regarding stability, a mechanism of repeated experimental validation is introduced to analyze the model's fluctuations under different data partitioning conditions, thereby verifying the model's robustness. Furthermore, error analysis of sub-indicators for each development dimension is

ISSN 2522-6398 Vol. 9, Issue 6: 97-106, DOI: 10.25236/FER.2026.090613

incorporated to more meticulously assess the model's performance differences across different capability dimensions.

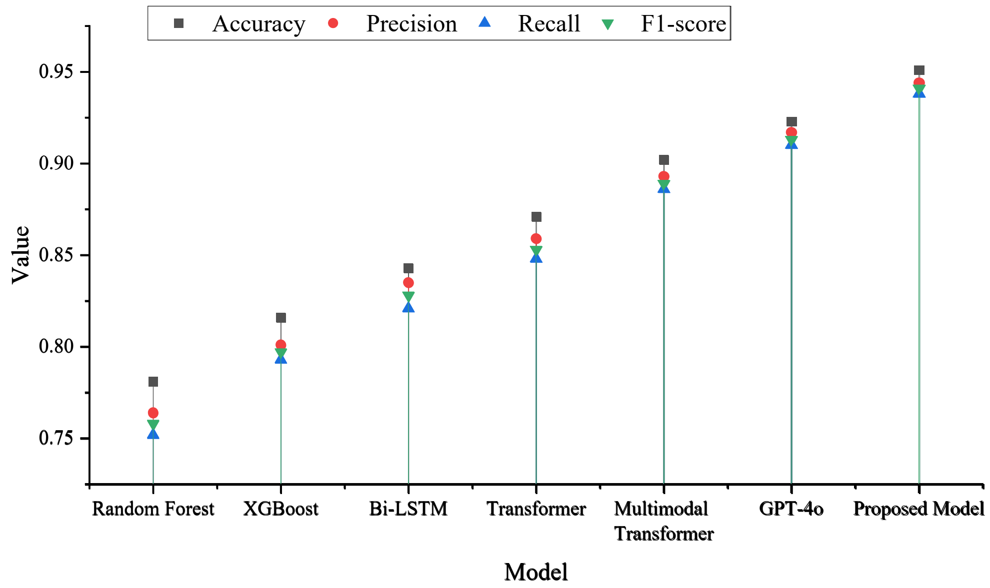

Figure 1. Comparison of the comprehensive development evaluation performance of different models

As shown in Figure 1, the multimodal Transformer demonstrates superior performance compared to the single-structure model, with its accuracy improved to 0.902, indicating that multi-source information fusion effectively enhances the ability to characterize children's developmental status. Meanwhile, GPT-4o, as a large language model, further improved to 0.923 in the comprehensive assessment task, demonstrating the advantages of large models in semantic understanding and cross-modal knowledge integration. However, the model proposed in this paper achieves the best overall performance, reaching 0.951, 0.944, 0.938, and 0.941 in Accuracy, Precision, Recall, and F1-score, respectively, all outperforming all compared methods. This result demonstrates that the proposed method possesses stronger expressive and generalization capabilities in multi-source information fusion and feature modeling, enabling it to more accurately characterize children's developmental levels and maintain stable assessment performance in complex educational scenarios.

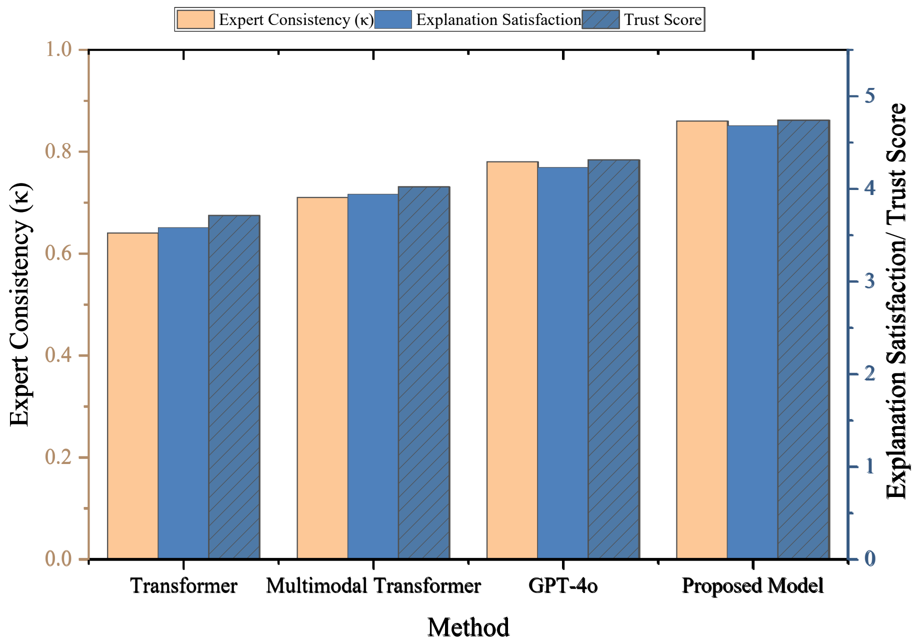

Figure 2. Results of interpretability evaluation

Note: Satisfaction and credibility are assessed using a 5-point Likert scale.

Early childhood education experts are invited to conduct the evaluation manually.

ISSN 2522-6398 Vol. 9, Issue 6: 97-106, DOI: 10.25236/FER.2026.090613

Using:

$$\kappa=\frac{P_o-P_e}{1-P_e} \tag{31}$$

That is, Cohen's Kappa coefficient.

As shown in Figure 2, the interpretability evaluation results reveal significant differences among different models in terms of expert consistency and user subjective evaluation. The Transformer model has a Cohen's Kappa coefficient of 0.64, indicating a moderate degree of consistency deviation between it and expert evaluation. Furthermore, its explanation satisfaction and credibility scores are relatively low, at 3.58 and 3.71 respectively, suggesting limited interpretability in complex educational decision-making scenarios. With an enhanced model structure, the multimodal Transformer shows improvement across all metrics, with its Kappa value increasing to 0.71. This indicates that multi-source information fusion helps enhance the consistency between model judgments and expert cognition, while explanation satisfaction and credibility also steadily improve. GPT-4o further improves to 0.78, demonstrating strong performance in both expert consistency and subjective evaluation, indicating that large language models have certain advantages in semantic explanation and decision explanation.

Table 1. Predicted development trends in preschool children

Model	MAE	RMSE	Trend Accuracy
Bi-LSTM	4.28	5.41	0.791
Transformer	3.81	4.92	0.824
Multimodal Transformer	3.22	4.13	0.861
GPT-4o	2.89	3.76	0.887
Proposed Model	2.31	3.08	0.926

Where:

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}|y_i-y_i| \tag{32}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-y_i)^2} \tag{33}$$

As can be seen from the prediction results of early childhood development trends in Table 1, there are significant differences in the performance of different models in the time series prediction task. The multimodal Transformer, by fusing multi-source data, provides a more comprehensive characterization of the developmental process, with its MAE and RMSE decreasing to 3.22 and 4.13 respectively, and its Trend Accuracy increasing to 0.861, indicating that multimodal information can effectively enhance trend recognition capabilities. GPT-4o further improves prediction performance on this basis, reducing MAE to 2.89, RMSE to 3.76, and achieving a trend accuracy of 0.887, demonstrating the advantages of large models in modeling complex semantics and temporal correlations.

5. Conclusion

This paper addresses the issue of multidimensional assessment and trend prediction in early childhood development. It constructs an intelligent assessment model that integrates heterogeneous data from multiple sources. By integrating behavioral videos, audio information, and educational text records, it achieves comprehensive modeling of children's cognitive, linguistic, and social development. Experimental results show that the proposed method outperforms the comparative model in comprehensive development assessment, interpretability analysis, and trend prediction tasks, validating the effectiveness and superiority of the multimodal fusion and unified modeling framework in early childhood development assessment. It should be noted that this study still has certain limitations. For example, the data sources are still mainly from specific educational scenarios, and the sample diversity and cross-regional generalization ability need further verification. Future research will further expand the data coverage, introduce larger-scale real-world educational scenario data, and combine more advanced self-supervised learning and causal inference methods to improve the model's generalization and causal explanation capabilities, thereby better supporting intelligent early childhood education assessment and decision-making applications.

ISSN 2522-6398 Vol. 9, Issue 6: 97-106, DOI: 10.25236/FER.2026.090613

Acknowledgement

This work was supported by Shanwei Philosophy and Social Sciences Planning Project Interim Achievements, Project Number: SWSKYF—202619.

References

[1] Destriana R, Aksani M L, Priyanggodo D Y, et al. Design of a Digital Platform for PAUD Child Development Monitoring Using the Dynamic Systems Development Method and Machine Learning[J]. Jurnal Teknik Informatika (Jutif), 2025, 6(5): 3587-3601.

[2] Chen W, Shen Z, Pan Y, et al. Applying machine learning algorithm to optimize personalized education recommendation system[J]. Journal of Theory and Practice of Engineering Science, 2024, 4(01): 101-108.

[3] Chen X, Hu Z, Wang C. Empowering education development through AIGC: A systematic literature review[J]. Education and Information Technologies, 2024, 29(13): 17485-17537.

[4] Ouyang F, Dinh T A, Xu W. A systematic review of AI-driven educational assessment in STEM education[J]. Journal for STEM Education Research, 2023, 6(3): 408-426.

[5] Luo Y, Han X, Zhang C. Prediction of learning outcomes with a machine learning algorithm based on online learning behavior data in blended courses[J]. Asia Pacific Education Review, 2024, 25(2): 267-285.

[6] Wang S, Xu T, Li H, et al. Large language models for education: A survey and outlook[J]. IEEE Signal Processing Magazine, 2026, 42(6): 51-63.