AEM JOURNAL.ORG

Construction and Application of Intelligent Evaluation System for Preschool Education Quality Integrating Multi-Source Big Data

Yang Yang, Jiaxin Chen, Chan Bai
Xi'an Fanyi University, Xi'an 710105, Shaanxi, China
Corresponding author: Yang Yang (e-mail: 17647362855@163.com)

ABSTRACT With the rapid advancement of intelligent sensing networks, edge computing, and Electromagnetic Waves, Antennas and Propagation technologies, efficient multimodal information fusion has become a fundamental requirement for data-driven monitoring and decision support in complex cyber–physical systems. To address the challenges of heterogeneous data inconsistency, spatiotemporal misalignment, and limited semantic interaction, this study proposes an intelligent evaluation framework integrating multi-source big data through Digital Twin–Knowledge Graph (DT-KG) fusion and a Multimodal Spatiotemporal Alignment Transformer Network (MSAT-Net). The proposed architecture combines video, audio, and textual information using heterogeneous feature encoding, cross-modal relative position encoding, adaptive gating fusion, and multi-head self-attention to construct unified semantic representations and interpretable quantitative evaluation models. Edge computing and differential privacy mechanisms are incorporated to enable real-time data acquisition while ensuring secure information processing and privacy preservation. Experimental results demonstrate that the proposed system achieves an evaluation accuracy of 92.3%, a Pearson correlation coefficient of 0.887 with ECERS-3 expert assessment, and significant improvements in robustness, inference efficiency, and long-term quality monitoring capability. Beyond preschool education, the proposed multimodal fusion framework provides an effective methodology for distributed sensing, semantic information propagation, adaptive data fusion, and communication-oriented intelligent monitoring, offering valuable engineering references for applications in Electromagnetic Waves, Antennas and Propagation.

INDEX TERMS multimodal information fusion, edge computing, distributed sensing, semantic information propagation

I. INTRODUCTION

As global digital transformation accelerates, preschool education quality evaluation is shifting from empirical observation toward data-driven evidence, mirroring the transition to high-precision fiber morphology analysis and automated yarn spatial distribution monitoring in the technical industry. This evolution enables a level of systematic rigor in educational assessment that ensures the structural integrity and quantitative robustness of longitudinal quality metrics. The quality of preschool education has recently become an essential component in education policies due to its importance as a contributing factor to the development of childrens early achievement patterns and educational fairness. A more comprehensive and scientific preschool education evaluation system provides both a required method for monitoring and enhancing quality and therefore enhances efforts to improve resource allocation in education, as well as advance sustainable development across multiple levels within society [1,2]. However, traditional methods of evaluating preschool education are still primarily based upon observing and assessing with scales. Even though the Early Childhood Learning Environment Assessment Scale-3 (ECERS-3) and the Classroom Assessment Scoring System (CLASS) are among the standards used in both a theoretical and practical manner, they encounter issues such as considerable subjectivity, limited coverage, and delays in assessment [3,4]. With these aforementioned limitations in mind, evaluating preschool education continues to draw from the use of multi-source big data as a new way to understand the quality of preschool education that breaks beyond the limitations of time and space while utilizing the integration of technology and educational science to be a major breakthrough to improve the effectiveness of evaluation of preschoolers education [5,6]. In addition to providing support during the evaluation processes for monitoring and personalising the childrens education, data-driven evaluations can also assist policy-makers in making effective decisions regarding preschool education to aid in the development of accurate and efficient systems of support for preschool children and their families by providing empirical evidence. There are still several technical constraints to building a smart evaluation system. In early education settings, different elements that make up the quality of preschool education are created and exist together dynamically over time. Such elements as teacher-child interaction, communication through language, and activities in space are temporally and spatially related to one another to the extent that traditional technology is not able to produce a unified model for multiple sources while maintaining temporal and spatial correlation between them. Video-recorded behavior trajectories, audio-recorded dialogue content, and text-recorded curriculum lesson plans demonstrate significant differences with respect to both time granularity and semantics [7,8]. The lack of a common semantic fusion mechanism limits the ability to create a unified representation of quality, and therefore, there is still room for improvement in the development of very reliable quantitative models. Finally, the laws and regulations related to privacy protection create very strict limitations for the collection and use of fine-grained data for the evaluation of early education with respect to access and level of use [9,10]. At the intersection of technology and education, there is a big problem of keeping the educational characteristics diagnostic value while maintaining privacy, and this is becoming increasingly difficult to solve where educational practices are concerned. Additionally, because of the high demand for interpretability in evaluation results in educational practice, both teachers and administrators need to have knowledge not only about "what the score is," but also "what the educational meaning of the score is." In addition to this issue, because of the "black box," or hidden nature of the algorithm itself the interpretability of evaluation results is diminished. The presence of noise interference within the real-world context of education further serves to reduce the validity and usefulness of the evaluation results when considering educational improvement [11,12]. Overall, the main technical barriers hindering the successful application of intelligent evaluation systems are the disconnect between the diversity of data formats and the "one-size-fits-all" evaluation and measurement system, the tension between the continuous evolution of the education process and the static solidification of the evaluation process, and the adaptability dilemma between the generalization ability of general models and the specificity of scenarios. There is also an absence of any common standards for either the format of standardized data interface or for frameworks of fusion algorithms, and a lack of a systematic approach for quantitative mappings of multi-dimensional data across the categories of educational theory - creating additional challenges with respect to connecting theory to technology.

Theoretical research at the academic level has created a more developed framework for evaluating quality in preschool education. According to this framework, there are two main dimensions: "process quality" and "structural quality." Elements such as teacher-child interaction, language support, and organization of activities are seen as indicators of quality. However, how to transform the above theoretical constructs into quantifiable indicators and achieve large-scale, continuous monitoring has always been a challenge in educational research. In recent years, scholars have conducted extensive research on the evaluation of educational big data. In the process of expanding the evaluation dimensions, the research hotspots have evolved from single behavior analysis to multimodal information integration, and from shallow feature fusion to deep semantic mining. Early studies used computer vision technology to identify teacher-child interaction behaviors and construct a relationship model between behavior frequency and quality indicators. Although the recognition accuracy is high, it is limited to single-modal analysis and lacks a systematic consideration of the complementarity of multi-source data [13,14]. Some scholars use natural language processing technology to analyze classroom dialogue texts, extract emotional tendencies and language interaction features, and expand the language dimension of evaluation, but ignore the potential impact of non-verbal behaviors such as facial expressions, actions, and spatial activities on educational quality [15,16]. Building upon this foundation, multimodal learning and analysis frameworks have been gradually applied to the field of educational evaluation. By integrating various data sources such as video, audio, and logs, the richness of evaluation dimensions has been significantly enhanced. However, the fusion methods remain at the feature splicing level, lacking deeper semantic interaction mechanisms [17,18]. Meanwhile, the introduction of privacy protection technologies has made compliant data collection possible, but while ensuring data security, these mechanisms have also reduced feature richness and diagnostic accuracy to some extent [19,20]. Overall, existing research continues to expand the evaluation dimensions, but significant shortcomings remain in the spatiotemporal deep alignment of heterogeneous data, semantic fusion, and result interpretability, hindering the diagnostic depth and engineering application breadth of intelligent evaluation systems in educational practice. To address the aforementioned issues, this paper proposes constructing an intelligent evaluation system for preschool education quality that integrates multi-source big data, focusing on overcoming the technical bottlenecks of difficulty in aligning spatiotemporal features of heterogeneous multi-source data and the lack of deep semantic fusion mechanisms. This research design has created a three-layered system architecture consisting of three core parts: data perception, fusion computing, and application services. In the first layer (data perception), edge computing nodes have been deployed to allow for the real-time acquisition and privacy-desensitization of multi-source data. The second layer (fusion computing) includes an algorithm engine that is based on both spatiotemporal multimodal Transformers and adaptive gating fusion mechanisms to provide enhanced semantic interaction between different types of data sources. The final layer (application services) is designed to produce visualized diagnostic reports and quantitative quality scores for education practitioners and decision makers. The design maps out how to transfer multi-source perceptions of education to quantitative quality measures, providing a feasible engineering pathway for creating an evaluation system that has both high levels of accuracy and interpretability. By establishing a theoretical and practical basis for transitioning preschool education evaluation from empirical observation to data-driven modeling, this framework mirrors the high-precision structural optimization of fiber morphology. This alignment ensures the same level of manufacturing excellence found in technical engineering, thus providing robust support for the systematic advancement of educational quality.

II. CONSTRUCTION OF INTELLIGENT EVALUATION SYSTEM INTEGRATING MULTI-SOURCE BIG DATA

SYSTEM DESIGN CONCEPT AND OVERALL ARCHITECTURE

Preschool education quality assessments aim to reflect the true state of the educational process accurately while also offering effective guidance for improving education. This type of system design is built on the fundamental premise of "education leads to technology and technology supports education" and promotes the integration of educational theory with a data-driven approach. On the one hand, the dual-dimensional integration of process quality and structural quality—the basis of the mainstream theoretical framework of preschool education quality evaluation—is the foundation upon which the indicator systems theoretical foundation and thus the evaluation contents conformance to the consensus of education can be established. On the other hand, the use of multi-source big data utilizes the breadth of perception, persistence of time, and richness of attributes of multi-source big data to turn theoretical concepts into quantifiable, calculable and interpretable technical indicators. By integrating these elements, the evaluation system provides both educational validity and technical reliability through engineering feasibility.

The intelligent evaluation system for preschool education quality utilizes an architectural design that seeks to adhere to the principles of high cohesion and low coupling in order to develop efficient and quantifiable representations of multi-source and heterogeneous data. Figure 1 illustrates the dataflow and module interaction relationships of the system as a whole.

The kindergartens edge computing node contains a data perception layer that collects real-time and partially cleaned perception data from many sources (e.g., video streams, audio streams, text log files). The data perception layer contains a differential privacy module that applies desensitization techniques to sensitive data at its source, thus ensuring that childrens privacy is protected and the processing of their data complys with applicable laws and regulations. There is an encrypted data transmission path between the data perception layer and the data computation layer, which maintains the integrity of the information being transferred from one layer to another. Once the data has been computed by the data computation layer, it is sent to the data fusion computation layer, where it is subjected to the multimodal spatiotemporal alignment transformer algorithm engine. The algorithm engine maps heterogeneous features to a common latent semantic space by means of implementing the spatiotemporal alignment algorithm and calculating the dynamic weights of those features to create quantitative representations of quality indicators. The application service layer processes the quantitative results and creates diagnostic reports and interactive interfaces via a visualization engine, providing suggestions for quality monitoring improvement to education administrators and educators. The application service layer also encapsulates API (Application Programming Interface) interfaces, facilitating multi-terminal access and providing real-time feedback related to evaluation results. The three-layer architecture collaboratively establishes a technical closed-loop system, from physical perception through logical computation to service applications, thus supporting the continued stable operation of the evaluation system in real-world educational contexts.

This system is based on four key parts, which are: the multi-dimensional quality indicator system based on an education model; the multi-source data collection for education-related scenarios; the algorithm engine that generates quantitative representation of the quality of education; and the diagnostic application module for a diverse user group. Together, these four parts create closed loop logic. The quality indicator system drives the direction and dimensions of the data collection. The data collection provides input for the algorithm engine. The algorithm engine provides output or evaluation of the data collected. The diagnostic application module converts the evaluation result into actions that improve education. The effectiveness of these actions is then used in the next round of data collection as part of the new evaluation cycle, creating continuous evaluation, feedback, improvement, and re-evaluation.

SCENARIO-ORIENTED DATA ACQUISITION AND PRIVACY PROCESSING MECHANISM

To collect multiple sources of heterogeneous data, the data perception layer uses edge computing nodes installed in preschool locations so that data could be collected in real-time. Acquired data includes audio and video streams (high definition) as well as logs (text structured) of the preschool education process. Video data records the interaction trajectories between teachers and children, and they can gather information on where students are located during activities and how far away they are from their teacher, posture, facial expressions, etc. To capture both of these measurements, multiple cameras are placed in specific locations within the classroom where the teacher is working and where

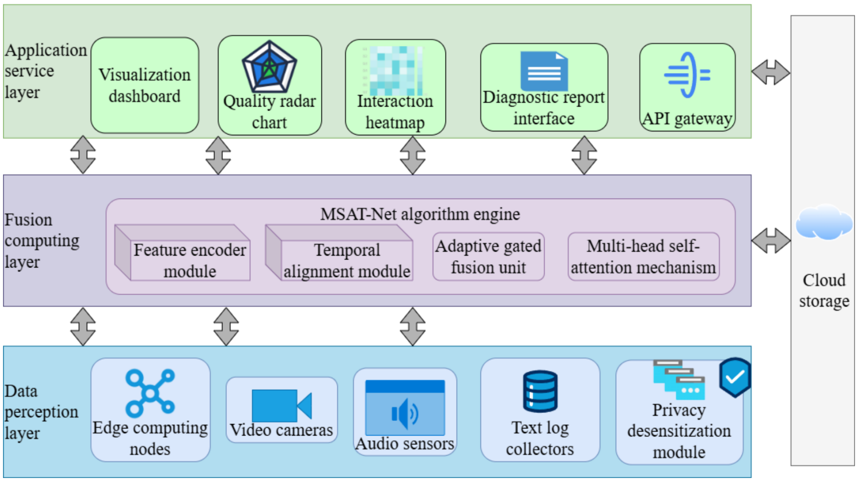

FIGURE 1. Overall architecture of the preschool education quality evaluation system

the students perform activities (main activity areas). The cameras are set to record at a rate of 25 frames per second, which allows for the detailed documentation of activity movement.

Audio data provides information about the acoustic properties of conversation in a classroom setting by capturing the teachers articulation of language, the students responses to that language, and the number of exchanges between teacher and student. Each microphone is placed throughout the classroom using an array arrangement on the ceiling with a sampling rate of 16kHz to facilitate proper sound source localization and speech separation.

Text logs are used to record structured data (e.g., lesson plans, attendance) and unstructured data (e.g., parent-teacher communications, incident reports). This data can be entered by teachers through terminal access or imported from other management systems. All text data is saved in UTF-8 format. Edge devices are equipped with accurate sensor devices that complete the transformation of raw data into a digital format and perform an initial cleaning process and until they converge upon a common format. These devices supply a uniform data input to the fusion computation process.

Protecting the privacy of children is one of the fundamental limitations that are placed on the data collection process. This system meets technical requirements and management requirements for implementing privacy protections.

At a technical level, the system employs a differential privacy approach to de-identify sensitive data. Individual characteristics are masked using differential privacy through the addition of noise into the query results [21,22]. To resolve the tension between privacy protection and the preservation of micro-expressions or subtle postures, the system employs a "feature-selective noise injection" strategy. Noise is primarily applied to global identifiers while preserving the structural relative distances of facial landmarks and joint points. This ensures that the privacy budget (ϵ) protects identity without blurring the high-frequency features necessary for emotional tone and posture analysis, thus maintaining the diagnostic accuracy of the pedagogical evaluation. The desensitized data features maintain statistical utility while eliminating the risk of individual identification. The core algorithm for adding noise is through the Laplace mechanism, which uses a probability density function to determine how much noise to add [23,24]. Below is a formula that shows how the noise addition is completed:

$$P(x) = \frac{\epsilon}{2\Delta f} \exp\left(-\frac{\epsilon|x|}{\Delta f}\right) \tag{1}$$

In Formula (1), $P(x)$ represents the probability density of the noise value; ϵ represents the privacy budget parameter; Δf represents the sensitivity of the query function; x is the added noise value. The smaller the privacy budget value, the stronger the privacy protection. The desensitized data features maintain statistical utility while eliminating the risk of individual identification. Specifically, the Laplace mechanism is calibrated to inject noise into non-essential attribute dimensions, preserving the structural integrity of spatiotemporal motion vectors and micro-expression heatmaps. This ensures that while individual identities are masked, the 3D-CNN can still extract high-fidelity behavioral gradients and interaction trajectories necessary for granular quality assessment. Furthermore, the AES-256 algorithm is strictly employed for the secure encapsulation of these preprocessed feature packets during transit to prevent manifold-level reconstruction attacks.

At the management level, the system establishes a hierarchical permission management mechanism. Different roles (teachers, principals, parents) are granted different data access permissions and can only view evaluation results and diagnostic information related to their responsibilities. The data transmission process uses the AES (Advanced Encryption Standard) encryption algorithm to ensure integrity [25,26]. The raw data is preprocessed at the edge and then enters the subsequent fusion process.

CORE ALGORITHM ENGINE DESIGN

The fusion computing layer carries a multimodal spatiotemporal alignment Transformer network to realize the accurate mapping of heterogeneous data to quality quantification representation. MSAT-Net is based on the standard Transformer encoder architecture and introduces a cross-modal relative position encoding module and an adaptive gating fusion unit for the characteristics of multimodal educational data to construct an end-to-end mapping model from multi-source perception to quality quantification. The MSAT-Net model inherits the multi-head self-attention mechanism of the Transformer architecture and makes several improvements to address the characteristics of multimodal educational data: it designs cross-modal relative position encoding to replace the original absolute position encoding; it introduces adaptive gating units to achieve dynamic calculation of modality contribution; and it constructs an educational indicator mapping layer to output interpretable evaluation results. The MSAT-Net does not operate in isolation; it is guided by a Digital Twin-Knowledge Graph (DT-KG) fusion layer. The Digital Twin provides a real-time virtual mapping of the classroom physical environment, synchronizing spatiotemporal coordinates to rectify data inconsistencies. Concurrently, the Knowledge Graph acts as a semantic constraint, mapping raw data features extracted by the Transformer to standardized pedagogical concepts. This ensures that the latent semantic space of MSAT-Net is anchored to validated educational quality indicators. The overall architecture and data flow of the MSAT-Net model are shown in Figure 2.

After the multi-source data is preprocessed by the perception layer, it enters the feature encoding stage. The original signals of each modality are mapped to a unified latent semantic space through independent encoders. The video modality uses a three-dimensional convolutional neural network (CNN) to extract spatiotemporal features [27]. The audio modality uses Mel-Frequency Cepstral Coefficients (MFCCs) combined with bidirectional LSTM (Long Short-Term Memory) to capture temporal dependencies [28]. The text modality obtains semantic embedding vectors through a BERT (Bidirectional Encoder Representations from Transformers) pre-trained model [29]. The heterogeneous feature encoding is represented as:

$$H_m = \text{Encoder}_m(X_m; \theta_m) \tag{2}$$

In Formula (2), H_m represents the hidden layer feature representation of the m-th modality; X_m is the original input data; θ_m is the learnable parameter of the encoder. The feature dimensions are uniformly projected to the d_{model} dimension space to eliminate the dimensional differences between modalities. Spatiotemporal misalignment errors are eliminated through a cross-modal relative position encoding mechanism. Traditional absolute position encoding cannot handle the temporal granularity differences of multimodal data. Relative position encoding introduces a time interval bias term to establish cross-modal temporal correlation. The position encoding is defined as:

$$PE(t_i, t_j) = \sin\left(\frac{t_i - t_j}{10000^{2k/d_{\text{model}}}}\right) \tag{3}$$

In Formula (3), t_i and t_j represent the indices of two time steps, respectively; k is the dimension index; d_{model} is the dimension of the hidden layer of the model. This encoding method enables the model to perceive the relative temporal distance between features rather than the absolute temporal position, effectively solving the alignment deviation caused by the inconsistency between video frame rate and audio sampling rate. The aligned feature sequence is input into the adaptive gating fusion unit for deep semantic interaction. The adaptive gating fusion mechanism dynamically calculates the contribution weight of each modality feature to achieve context-based importance filtering. The gating unit receives all modality encoded features and outputs a normalized weight vector. The weight calculation is expressed as:

$$G_m = \frac{\exp(W_g \cdot H_m + b_g)}{\sum_{k=1}^{M} \exp(W_g \cdot H_k + b_g)} \tag{4}$$

In Formula (4), G_m represents the gating weight of the m-th modality; W_g is the weight matrix; b_g is the bias term; M is the total number of modalities. The gating weight reflects the contribution of each modality to the current quality assessment task, and the features of high-weight modalities dominate the fusion result. The fused feature is expressed as:

$$H_{\text{fused}} = \sum_{m=1}^{M} G_m \odot H_m \tag{5}$$

In Formula (5), $\odot$ represents element-wise multiplication. This mechanism suppresses low-quality modal noise interference and enhances the expressive power of key features.

The fused features enter the multi-head self-attention layer to capture the long-range dependencies of quality indicators. The self-attention mechanism calculates the correlation strength between each position within the feature sequence, and the attention weight is expressed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

In Formula (6), Q is the query matrix; K is the key matrix; V is the value matrix; d_k is the dimension of the key vector. The scaling factor $\sqrt{d_k}$ prevents the gradient from vanishing due to excessively large dot product results. The model uses a multi-head mechanism that allows parallel computation of attention by breaking the feature space into several subspaces to help it learn how to represent various kinds of quality metrics. Attention weight data is shown and mapped to provide an understandable set of diagnostic criteria through visual representation, resulting in

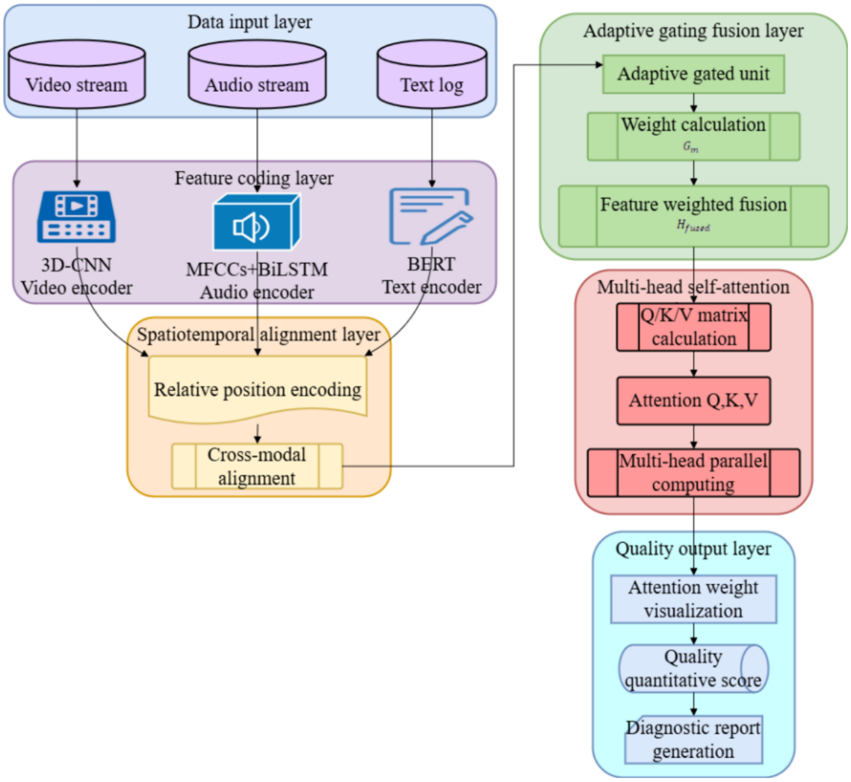

FIGURE 2. Working principle of multi-source big data fusion algorithm

how to relate the scoring logic of the attention mechanism to specific key feature areas of the image.

QUANTITATIVE CALCULATION MODULE FOR EDUCATIONAL QUALITY INDICATORS

A fully connected layer converts the fusion feature vector into the preschool education quality indicator space by mapping it to the quality indicators for preschool education divided into two key areas - process quality and structural quality. Each dimension has several secondary indicators forming an evaluation vector. The evaluation vector is expressed as:

$$I = [i_1, i_2, \dots, i_k, \dots, i_n]^T \tag{7}$$

In Formula (7), n is the total number of indicators, and i_k represents the feature value of the k-th quality indicator. The mapping process is achieved through a multilayer perceptron nonlinear transformation:

$$V = \sigma(W_1 H_{\text{fused}} + b_1) \tag{8}$$

V is the indicator feature vector; W_1 is the weight matrix; b_1 is the bias vector; σ is the activation function. The attention weight vector is normalized and used as the indicator importance coefficient. The importance coefficient vector is expressed as $A = [a_1, a_2, \dots, a_n]^T$, satisfying $\sum_{k=1}^{n} a_k = 1$. The value of a_k reflects the contribution of the corresponding indicator to the overall quality evaluation.

The quality quantification score is calculated by weighted summation. The comprehensive score is defined as:

$$S = \sum_{k=1}^{n} a_k \cdot v_k \tag{9}$$

In Formula (9), S is the final quality quantification score, and v_k is the k-th component in the indicator feature vector V. The score range is normalized to the interval 0 to 100:

$$S_{\text{norm}} = \frac{S - S_{\min}}{S_{\max} - S_{\min}} \times 100 \tag{10}$$

In Formula (10), $S_{\min}$ and $S_{\max}$ are the theoretical minimum and maximum scores, respectively. The quality level is determined based on the preset threshold interval. The level determination function is expressed as:

$$\text{Level} = \sum_{j=1}^{J} j \cdot I(T_{j-1} \le S_{\text{norm}} < T_j) \tag{11}$$

In Formula (11), Level is the output quality level; J is the total number of levels; T_j is the threshold boundary of the j-th level; $I(\cdot)$ is the indicator function. The indicator function takes the value 1 when the condition is met, and 0 otherwise.

The mapping module produces both quantification scores and level labels. While quantification scores yield continuous numerical values to represent gradations in quality, class labels provide discrete classifications, which can expedite decision making by educational administrators. Visualization of attention weight results, in conjunction with indicator scores, is considered the source data for the diagnostic report on the child. The resulting data stream enters the application service layer for rendering and display. The complete process of quantization calculations is designed in a differentiable manner to enable model training from one end of the process to the other. Backpropagation of gradients is accomplished within the quantization layer before moving to the feature extraction layer in order to optimize all parameters. The objective function used to train the model is based on mean squared error as the loss function:

$$\text{Loss} = \frac{1}{N}\sum_{i=1}^{N}\left(S^{(i)} - \hat{S}^{(i)}\right)^2 \tag{12}$$

In Formula (12), N is the sample size; $S^{(i)}$ is the model prediction score; $\hat{S}^{(i)}$ is the expert-annotated true value. The loss function minimization process drives the model to approximate the true quality distribution.

DIAGNOSTIC APPLICATION MODULE DESIGN

The evaluation system serves multiple stakeholders involved in improving the quality of preschool education, including front-line teachers, kindergarten administrators, and parents. Different user groups have varying needs for evaluation results. Therefore, this system designs differentiated application service modules to ensure that evaluation information is accurately reached and effectively used.

The application service layer receives the quality score and attention weight data output by the quantitative calculation module. The visualization engine, based on a modern Web front-end framework, constructs a dynamic interactive interface to intuitively present the evaluation results. The interface layout is shown in

The multidimensional distribution of quality is presented through a radar chart showing the differences in scores for each dimension. Hotspots in teacher-child interaction spaces are displayed through a heat map showing the frequency distribution of activity areas. The diagnostic report generation module extracts key improvement items based on attention weight ranking to form text suggestions. The diagnosis priority index can be calculated using the formula $P_k = \alpha \cdot A_k + \beta \cdot (1 - V_k)$, where P_k is the priority accorded to the kth indicator (A_k—the weight attributed by the models); V_k is the relative score of the k-th indicator; α and β are the balance coefficients. Indicators with high priority trigger the system to generate suggestions for improvement. The backend service sends data to the frontend via an interface utilizing JSON format for communication to ensure efficiency. The system also provides real-time refresh capabilities for data and historical trend backtracking in order to address the need for dynamic monitoring. Education administrators and teachers are able to view the results of evaluations via their devices in order to make informed decisions. The text for diagnostic suggestions can automatically be generated based on a combination of weighted data and preconfigured templates, so that the cost of manual interpretation of evaluation data can be minimized. When data is updated, interface components are re-rendered in order to maintain timeliness of information. The entire module translates evaluation data from numerical form into visual form to support improved educational practice.

III. EXPERIMENTAL DESIGN

CONSTRUCTION AND DESCRIPTION OF EXPERIMENTAL DATASET

The data used for the experiments has been collected from ten different preschool classrooms located within public and private schools. During this collection phase, data is collected across a full semester of school, ensuring an equal time distribution between classrooms from which data is collected in order to maintain uniformity with respect to sample size within each sample group. Multi-source heterogeneous data consists of high-definition video stream images captured at 25 frames per second per video stream, high-definition audio recordings captured at 16 kHz per audio stream, and structured text logs of the teaching and attendance records. Teaching and attendance records are captured as structured logs using UTF-8 encoded structured text format. Edge Computing Nodes convert physical image/audio signals to digital form and perform the initial signal clean-up of data collected from preschool teachers and preschool students. The primary constraint on collection is the protection of childrens privacy. To this end, the system uses a differential privacy mechanism to desensitize sensitive data, retaining their statistical utility but preventing them from being able to identify an individual. This study was approved by the Ethics Committee of Xi'an Fanyi University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

The qualified preschool education experts have completed the evaluation truth labeling process using the ECERS-3 (Early Childhood Environment Rating Scale) and CLASS assessment systems [30]. The experts independently evaluate the samples for their scoring validity and the average of their scores become the final quality label. The annotation of the dataset is based on a double-blind review process to ensure there is no bias [31,32]. The training, validation, and test sets are all divided according to the 7:2:1 ratio in order to determine if the model is able to generalize its results. Additional information about the sources of the experimental datasets and their respective sizes can be seen in Table 1.

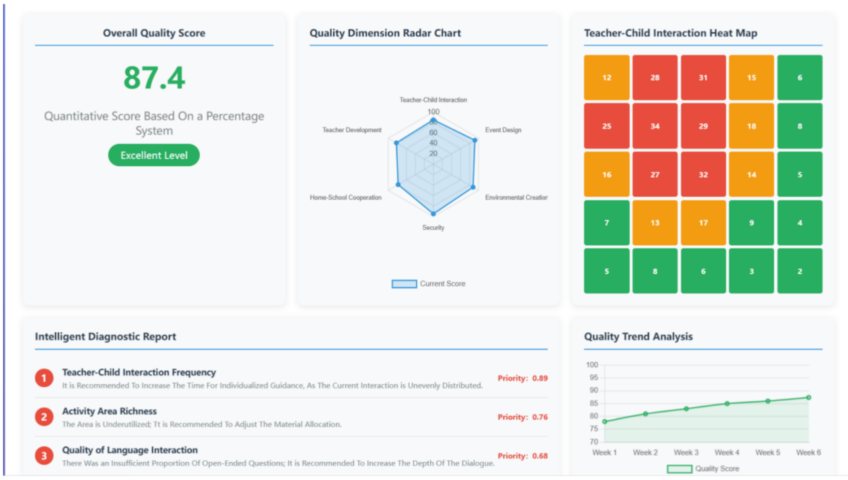

FIGURE 3. System application interface prototype

TABLE 1. Sources and sizes of experimental datasets

Data modal	Collection time/number of records	Total sample size	Labeling basis
Video modality	500 Hour	120,000 frames	ECERS-3 behavioral observation
Audio modality	500 Hour	45,000 segments	CLASS language assessment
Text modality	2,000 copies	2,000 copies	Teaching plan syllabus
Comprehensive dataset	-	5,000 sample	Expert overall score

Table 1 shows that the dataset has the characteristics of multimodal heterogeneity and sufficient scale. The duration of video and audio data covers various teaching activity time periods, ensuring sufficient learning of spatiotemporal features. The number of text logs matches the number of classes, supporting semantic association analysis. The distribution of training set and test set remains independent to avoid the risk of data leakage. The labeling confidence coefficient reaches above 0.85, ensuring the reliability of the true value. The data scale meets the training requirements of deep learning models and supports the convergence of MSAT-Net network parameters. The dataset construction process complies with privacy protection specifications and has a legal basis for engineering application.

COMPARISON OF MODEL AND PARAMETER CONFIGURATION SETTINGS

In order to verify the performance of MSAT-Net model, a representative baseline model is selected for comparative experiments. CNN-LSTM (Convolutional Neural Network - Long Short-Term Memory) model is used as the benchmark for spatiotemporal feature extraction, and the convolutional neural network and long short-term memory network are combined to process sequence data [33]. The standard Transformer encoder is used as the benchmark for attention mechanism to verify the effectiveness of multi-head self-attention structure. Early fusion multilayer perceptron is used as the benchmark for simple fusion strategy to evaluate the gain of deep semantic fusion mechanism. Support vector machines (SVMs) are used as a traditional machine learning benchmark to compare the feature representation capabilities of deep learning methods [34]. All comparison models are trained and tested using the same dataset. Hyperparameter settings follow the grid search optimization results. The optimizer uses the Adam algorithm. The loss function used is mean squared error. The number of training iterations is set to 100 rounds. An early stopping mechanism is used to prevent overfitting. The weight initialization uses the Xavier method. The hardware environment is uniformly configured to eliminate differences in computing resources. The model hyperparameter configuration and hardware environment are shown in Table 2.

TABLE 2. Model hyperparameter configuration and hardware environment

Configuration items	Symbol	MSAT-Net	CNN-LSTM	Transformer	MLP	SVM
Hidden layer dimensions	d_{model}	512	512	512	512	-
Learning rate	η	0.001	0.001	0.001	0.001	-
Batch size	batch_size	64	64	64	64	-
Number of iteration rounds	epochs	100	100	100	100	-
Optimizer	-	Adam	Adam	Adam	Adam	SMO (Sequential Minimal Optimization)
Graphics card model	-	Tesla V100	Tesla V100	Tesla V100	Tesla V100	CPU
Memory capacity	-	64 GB	64 GB	64 GB	64 GB	32 GB

The experimental environment configuration is highly consistent. Uniform learning rate and batch size parameters eliminate interference from training process variables. The increase in MSAT-Net model parameters is due to the introduction of the gating fusion module. NVIDIA Tesla V100 GPUs provide sufficient computing power to support large-scale matrix operations. Hardware configuration meets the convergence requirements of deep learning models. Experimental settings ensure the reproducibility of results and fairness of comparison.

APPLICATION VALIDATION SCENARIOS AND PROCESS DESIGN

The application validation scenario is the pilot kindergarten as the data collection source. The system adopts a private deployment mode, installed on the kindergartens local server to ensure data sovereignty and security. Edge computing nodes are connected to existing monitoring equipment to complete hardware integration and network configuration. The system collects data from multiple sources and generates a report on the quality of those sources after it has started the evaluation process. The teacher receives a copy of the quality diagnostic report via an encrypted channel on the terminal of their device. The system provides specific pedagogical interventions based on the identified gaps, such as "scaffolding childrens complex language use through open-ended questioning" or "facilitating peer collaboration during block play." To prevent "gaming the system" (where scores improve via superficial behavioral changes like moving closer to a microphone), the MSAT-Net cross-validates improvements across multiple modalities—e.g., verifying that an increase in audio volume matches the semantic depth of teacher-child dialogue and the physical proximity captured by spatial trajectory data. The teacher implements interventions to help improve quality that is described in the report. After each of the 12 weeks of intervention, the system evaluates whether there has been any change in the quality indicators that are measured prior to the intervention. Longitudinal tracking of data records the evolution trajectory of quality scores, and the validation process covers the dimensions of system operational stability and the guidance of evaluation results. Data collection frequency is set to daily teaching hours. Abnormal data triggers the system alarm mechanism. The administrator logs system resource usage in the backend, and user operation logs are stored in the audit database. The entire verification process adheres to ethical guidelines for educational experiments. The implementation of intervention measures is incorporated into the teacher performance evaluation system. System version iterations are optimized based on user feedback logs. Data from the completed verification process is archived to the research database.

IV. RESULTS AND APPLICATION EFFECTS

SYSTEM MODEL PERFORMANCE COMPARISON

Model technical performance verification constitutes the core basis for the engineering feasibility of the intelligent evaluation system. The multimodal spatiotemporal aligned Transformer network must be systematically tested for classification accuracy, precision, recall, and deployment efficiency. Baseline models (SVM (Support Vector Machine), MLP (Multilayer Perceptron), CNN-LSTM, Transformer) are selected for comparison with MSAT-Net while maintaining independent and consistent distribution of test datasets. Six core performance categories (accuracy, F1-score, precision, recall, scenario robustness, and inference speed) are measured using the average of 5 independent experiments, and then the standard deviation is calculated to eliminate random variation. Technical performance comparisons of each modeling approach are presented in Figure 4.

MSAT-Net scores 0.923 in terms of accuracy and 0.915 when evaluated by the F1 metric; SVM scores 0.762 and 0.741 respectively, as shown in Figure 4(a). The difference in performance levels is due to the dynamic selection of modal feature information based on high-contribution with MSAT-Nets adaptive gating mechanism compared to the static method of SVM using manually extracted features where deep semantic relationships cannot be captured through SVM with respect to multi-source data. According to Figure 4(b), MSAT-Net has achieved a precision score of 0.918 and a recall score of 0.912. The precision score differs from the baseline transformer model by 0.053, while the recall score is 0.051 different than the baseline transformer model. The reason for this difference in scores is due to the addition of cross-modal relative position encoding, which reduces errors in time alignment that exist between the video frame rate and audio sample rate. Conversely, the absolute position encoding used in standard transformer models cannot accommodate the differences in temporal granularity between modalities; therefore, there are alignment errors caused by these differences. Robustness and inference efficiency are compared in Figure 4(c). MSAT-Net achieves cross scenario robustness of 0.894 compared to SVM, which has cross scenario robustness of 0.685, providing a 30.5% improvement to MSAT-Net over SVM. The inference efficiency normalization value is 1.0 for MSAT-Net and 0 for SVM, resulting in a significant improvement to inference efficiency with the MSAT-Net model over the SVM model. The improvement to robustness comes from the feature preservation mechanism used after desensitizing data using differential privacy by preserving statistical utility. The advantage of improved inference efficiency comes from preprocessing the data at the edge computing node and therefore decreasing the cloud computing load. As shown in Figure 4, there is a significant statistical difference between MSAT-Net and the baseline model ($p < 0.05$). MSAT-Net outperforms the baseline model in all three dimensions: classification accuracy, precision, recall, and engineering deployment efficiency, thus meeting the dual requirements of real-time and accurate operation of the intelligent evaluation system for preschool education quality.

CRITERION-RELATED VALIDITY OF EVALUATION RESULTS

Educational verification at intelligence evaluation is very important for using engineering and theorem-based engineering algorithms, so it is crucial to develop an educational quality score that is statistically correlated to preschool education quality (according to established scales) in order to validate that the scoring of educational quality is aligned with educational theory. In this experiment, ECERS-3 (Early Childhood Environment Rating Scale) and CLASS (Classroom Assessment Scoring System) are used

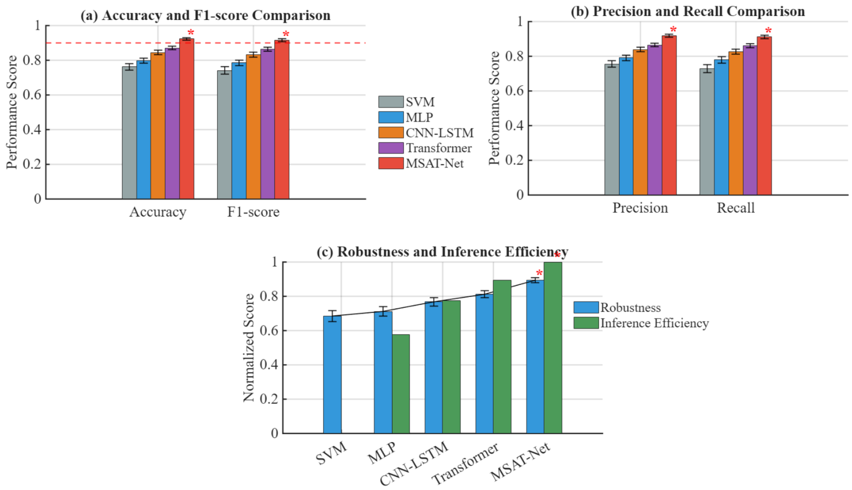

FIGURE 4. Multi-model technical performance comparison analysis: (a) Comparison of accuracy and F1 score; (b) Comparison of precision and recall; (c) Robustness and inference efficiency

as classified tools to establish these correlations. Qualified early childhood education experts independently score the same sample. Pearson correlation analysis, Bland-Altman consistency test, and Kappa coefficient calculation are performed on the system scores and expert scores. Twenty classes participate in validity verification, and a double-blind review mechanism is used to eliminate subjective bias. The criterion-related validity analysis of the evaluation results is shown in Figure 5.

Figure 5 presents the correlation characteristics between model scores and scores on the two authoritative scales. Figure 5(a) shows that the Pearson correlation coefficient r between the model scores and ECERS-3 expert scores reaches 0.887, and the coefficient of determination R^2 is 0.787. Furthermore, in Figure 5(b), the Pearson correlation coefficient r between the model scores and CLASS expert scores reaches 0.894, and the coefficient of determination R^2 reaches 0.799. This high correlation stems from the attention weight mapping mechanism of this papers evaluation system, which establishes a non-linear relationship between algorithmic features and educational indicator dimensions. The long-term dependency relationship of teacher-child interaction captured by the multi-head self-attention layer corresponds to the interaction quality construct in the ECERS-3 scale. Figure 5(c) shows that the Bland-Altman consistency analysis has an average difference of 1.24 points, with the 95% consistency threshold falling between -3.34 and 3.75 points. All twenty sample points fall within the consistency threshold. The consistency bias is attributed to the subjective judgment fluctuations in expert ratings. The system rating maintains algorithmic stability, and differential privacy desensitization maintains the statistical utility of features while eliminating individual identification risks. In Figure 5(d), the diagonal elements of the quality level confusion matrix account for 80%, and the Kappa consistency coefficient is 0.812. The consistency of the level classification stems from the alignment of the threshold division strategy with the expert level judgment criteria, and the adaptive gating fusion mechanism suppresses low-quality modal noise interference to ensure the robustness of the rating. Overall, the Bland-Altman analysis does not demonstrate any apparent systematic bias in either direction: therefore, there is no evidence of a systematic trend between the MSAT-Net output rating and the authoritative rating for preschool education quality (an example of how the evaluations met the educational and scientific validity requirements of the intelligent evaluation system).

TYPICAL KINDERGARTEN APPLICATION CASE TRACKING

Determining if the educational use of the intelligent evaluation system is effective establishes the principal basis to determine the value of the project. Therefore, the diagnostic results obtained as a quality measure must be transformed into practical recommendations for teaching improvements and subsequently validated with follow-up evaluations in real educational contexts. The experiment selects three typical kindergartens— a demonstration kindergarten, a regular kindergarten, and a rural kindergarten—as pilot subjects, denoted as Kindergarten A, B, and C, respectively. The MSAT-Net evaluation system is deployed, and continuous monitoring is conducted for twelve weeks. Weekly quality quantification scores and diagnostic reports are generated, and teachers adjust their teaching strategies based on the reports recommendations. The trajectory of quality score changes records the guiding effectiveness of the system evaluation results on teaching practice. The follow-up results of the typical kindergarten application case are shown in Figure 6.

Figure 6 presents the evolution characteristics of quality scores of the three types of kindergartens over twelve weeks and the average intervention effect. Figure 6(a) shows that the quality score of Kindergarten A increases from 79.2 points before intervention to 91.8 points, an increase of 12.6 points; Kindergarten B increases from 72.1 points before intervention to 88.4 points, an increase of 16.3 points; and

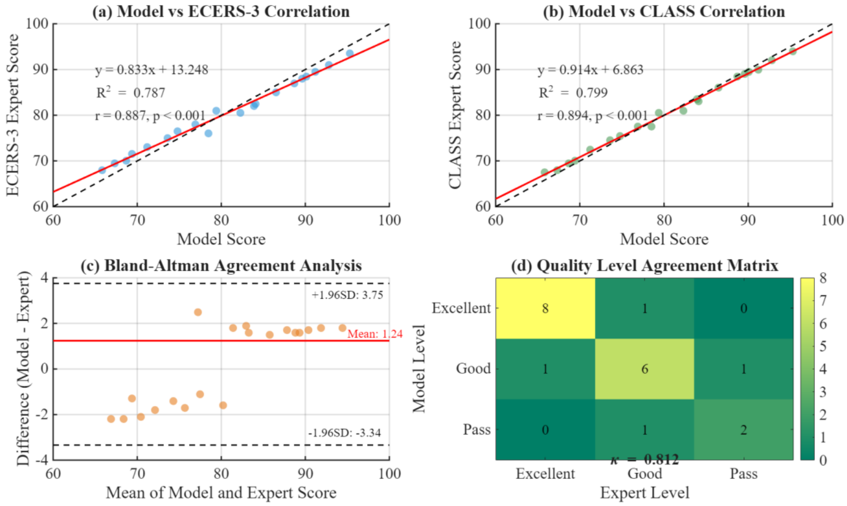

FIGURE 5. Criterion-related validity of the evaluation results: (a) Correlation between the model and ECERS-3; (b) Correlation between the model and CLASS; (c) Bland-Altman consistency analysis; (d) Quality rating matrix

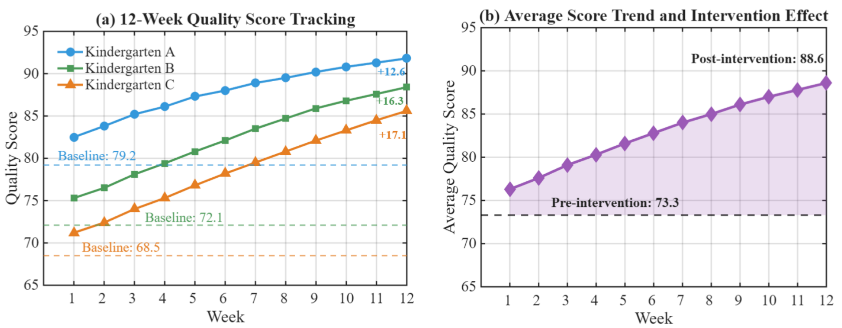

FIGURE 6. Follow-up results of typical kindergarten application cases: (a) Longitudinal tracking of quality scores of the three kindergartens over 12 weeks; (b) Trend of average scores and intervention effect of the three kindergartens

Kindergarten C increases from 68.5 points before intervention to 85.6 points, an increase of 17.1 points. The difference in the improvement rate stems from the differentiated improvement space of kindergartens with different baseline levels. Rural kindergartens have lower initial quality, but the marginal effect of intervention is more significant. Figure 6(b) shows that the average score of the three kindergartens increases from 73.3 points before intervention to 88.6 points, an overall increase of 15.3 points, representing an increase of 20.9%. The sustained growth trend is attributed to the closed-loop feedback mechanism formed by the weekly diagnostic reports generated by the system. Teachers can identify blind spots in teacher-child interaction and implement targeted adjustments based on the visualized results of attention weights. The slopes of the three tracking curves are decreasing. The average number of points added per week from week 1 to week 4 equals 1.33 and from week 9 to week 12 equals 0.83. The reason for the decline in the growth rate is due to decreased marginal benefit of quality improvement. The first step forward generally focuses on improving easy-to-reach points, while additional improvements tend to focus on the education structure. Each of these curves is steadily increasing and does not exhibit fluctuations or declines, and the average score of all three curves is above the baseline threshold following intervention. MSAT-Net has provided an ongoing quality improvement result in actual education and meets the performance criteria of an intelligent evaluation system for quality preschool education.

EFFECTIVENESS OF QUALITY IMPROVEMENT BASED ON EVALUATION FEEDBACK

The degree to which the diagnostic suggestions of the intelligent evaluation system are adopted determines the actual effectiveness of quality improvement. The experiment collects score data of key quality indicators from three kindergartens before and after intervention, calculated Cohens d effect size to quantify the intervention intensity, and statistically analyzed the teachers adoption rate of the systems diagnostic suggestions to verify the credibility of the diagnostic results. The evaluation index system comprises process indicators such as teacher-child interaction frequency, activity area richness, language interaction quality, safety guarantee index, and home-school cooperation level, along with the comprehensive score. Figure 7 shows the results of the quality improvement effectiveness analysis based on evaluation feedback.

Figure 7 presents the comparison of key indicators before and after intervention, improvement magnitude, effect size

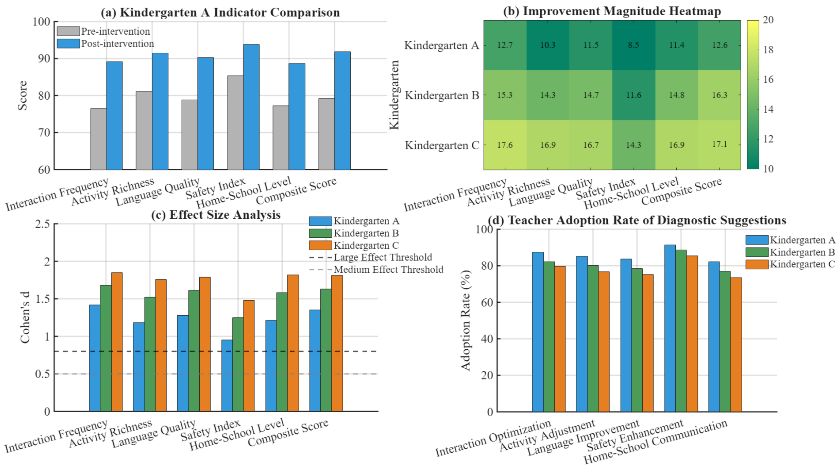

FIGURE 7. Quality improvement effects based on assessment feedback: (a) Comparison of key indicators before and after intervention in kindergarten A; (b) Comparison of improvement magnitude; (c) Effect size analysis; (d) Teacher adoption rate of diagnostic suggestions

analysis, and statistical data on the adoption rate of diagnostic suggestions for the three kindergartens. The data from the four subplots shows that the teacher-child interaction frequency in Kindergarten A increases from 76.5 points to 89.2 points, an increase of 12.7 points, with an effect size of 1.42; the effect size in Kindergarten B reaches 1.68; the effect size in Kindergarten C reaches 1.85. The difference in effect size gradient stems from the greater behavioral changes of teachers in rural kindergartens after accepting the systems diagnostic suggestions, while the existing teaching behaviors of teachers in demonstration kindergartens are already close to optimal, resulting in limited room for improvement. The average effect size for activity area richness is 1.49; the average effect size for language interaction quality is 1.56; the average effect size for safety assurance index is 1.23; the average effect size for home-school collaboration level is 1.54. The safety assurance index has a low effect size compared to its high baseline value, which is 85.3 points prior to intervention at kindergarten A. On the other hand, the average effect size of the indicators is greater than or equal to large effect size of 0.8 due to the strong impact of adaptive gating fusion mechanisms through suppression of low quality modal noise interference thus providing robustness of the scores, differential privacy desensitizating individual identification risks and preserving the statistical utility of the systems and the resulting trust in the system by teachers that encourage higher rate of intervention implementation. Statistics on the adoption rate of diagnostic recommendations show that the average adoption rate of recommendations for improving teacher-child interaction in the three kindergartens reaches 83.1%; the average adoption rate of recommendations for improving the safety environment is 88.5%; the average adoption rate of recommendations for home-school communication is 77.5%. The differences in adoption rates correspond to the operability and effectiveness of the safety-related suggestions being rapid as compared with home/school communication suggestions. Home-school communication also requires outside coordination for its implementation, which creates difficulties for its implementation. In Figure 7, bar charts before and after the intervention reveal positive growth for each of the indicators, and heat map color depth analysis indicates greater color depth from Kindergarten A to Kindergarten C. All bars for effect sizes exceed the large effect threshold. The MSAT-Net provides a high rate of adopting diagnostic recommendations and a high effect size for all diagnostic recommendations met by the MSAT-Net diagnostic evaluations, meeting the standards set forth by an intelligent systems evaluation of preschool education quality.

IMPACT OF THE EVALUATION SYSTEM ON KINDERGARTEN MANAGEMENT AND TEACHING RESEARCH PRACTICE

This research further examines how well the evaluation system works within a real-life learning context by conducting in-depth interviews with three Kindergarten managers after the pilot program is completed. Systematic process materials including records of teaching research activity, resource allocation plans, and teacher training files are collected. The research assesses the evaluation results from five areas to explore how the results support kindergarten management and how those results support teaching research practices: 1) teaching research management; 2) resource allocation; 3) teacher development; 4) management effectiveness; 5) home-school collaboration. A comparison of the effects of the kindergarten pilot program and the effects of a similar type of kindergarten that does not implement the system is presented in Table 3.

Table 3 shows that the pilot kindergartens have improved their management and teaching research in a systematic way. Currently, the average number of adjustments made to teaching research plans is 3.2 times per semester, which is an increase of 113.3% over the control kindergartens 1.5 times. The systems diagnostic reporting identifies common weaknesses in the teaching process and serves as a valid

TABLE 3. Statistics on system operation efficiency and stability indicators

Indicator category	Specific indicator	Pilot kindergartens (n=3)	Control kindergartens (n=2)	Difference analysis
Teaching research management	Frequency of semester teaching research plan adjustments (times/semester)	3.2	1.5	+113.3%
	Proportion of data-driven teaching research activities (%)	68.4%	22.5%	+45.9 percentage points
	Teacher participation rate in teaching research (%)	94.2%	86.7%	+7.5 percentage points
Resource allocation	Frequency of resource adjustments based on evaluation results (times/semester)	5.7	2.0	+185.0%
	Improvement rate of area utilization after resource adjustment (%)	18.6%	6.3%	+12.3 percentage points
Teacher development	Average frequency of teacher participation in kindergarten-based training (times/semester)	4.3	2.8	+53.6%
	Teacher satisfaction with training relevance (5-point scale)	4.41	3.86	+0.55
Management Efficiency	Quality monitoring report generation cycle (days)	7 (weekly)	30 (monthly)	Shortened by 76.7%
	Time spent by managers on data interpretation (hours/week)	1.8	4.5	-60.0%
Home-Kindergarten Collaboration	Frequency of referencing evaluation data in home-kindergarten communication (times/month)	12.4	3.2	+287.5%
	Parent satisfaction with kindergarten services (5-point scale)	4.52	4.18	+0.34

basis for the selection of teaching research topics. As a result, the percentage of data-driven teaching research activities has increased from 22.5% to 68.4%. The provision of visualisation reports allows administrators to include data evidence into their decision making processes for teaching research, thus expanding the traditional limitations on relying solely on classroom observations. In addition, the pilot kindergartens use their evaluation results to change resource allocation 5.7 times per semester or nearly three times as often (2.0 times) as the control kindergartens. Furthermore, administrators can use the data provided by the system to accurately determine where the differences in inter-class quality exist, allowing them to more accurately identify the weaknesses in their use of funds for resource allocation. The frequency at which teachers participate in training has increased from 4.3 times per semester — 53.6% increase from control kindergartens. The diagnostic reports generated on each teacher contain profiles of their teaching behaviours, which allows for kindergartens to create different training levels (i.e., novice/career). The average satisfaction rating of teachers for the relevance of training content is 4.41, which is 0.55 above the control kindergartens. The diagnostic data generated from the pretraining system provides accurate input for the design of training. In terms of management efficiency, the cycle time for generating reports has decreased to 7 days and established a reliable and consistent method of producing weekly reports. The amount of time that administrators spend interpreting the data has been reduced by 60.0%, to 1.8 hours per week compared with the control kindergartens. The average number of times that evaluation data is used in home-school communication is 12.4 times per month, which is approximately four times more than the control kindergartens. Data on childrens growth trajectories provides concrete and tangible evidence to support communication between teachers and parents. According to these data, the evaluation system has changed from a technical tool to an organizational mechanism to assist kindergarten personnel in management decision-making and in conducting research in order to generate systematic effects on: 1) research focus; 2) resource allocation; 3) teacher development; 4) management efficiency; and 5) homeschool collaboration.

V. CONCLUSION

Drawing upon the principles of high-precision fiber morphology and manufacturing excellence, this paper develops an intelligent evaluation system for preschool education quality that integrates multi-source big data. This framework ensures the structural integrity and spatial distribution of evaluation metrics to rectify the fragmented nature of educational data, mirroring the technical precision required in high-performance engineering. The use of a multimodal spatiotemporal alignment Transformer network enables semantic fusion and quality quantification of heterogeneous data. The experimental verification of the system produces an F1 score of 0.915 as well as a Pearson correlation coefficient of 0.887 between the evaluation results produced by using the system and the ECERS-3 rating scale. In addition, there is a 15.3-point increase in average quality scores of pilot kindergartens. These results indicate that the algorithm employed in this study, which is a combination of a multimodal spatiotemporal alignment transformer network and the adaptive gating fusion mechanism for dynamic filtering of modal features, is effective for resolving the challenges associated with spatiotemporal alignment of multi-source data. The system also has demonstrated strong robustness and interpretability. The application of cross-modal relative position encoding has significantly reduced the potential for temporal alignment errors. The system supports engineering and technical assistance in the provision of preschool education quality management through a change from experience-based to data-driven evaluation models that are both stable and reliable and meet the requirements of large-scale implementations. The processing of data on edge-computer nodes decreases the amount of processing required by the cloud computing environment, and implementing differential privacy techniques to the pre-processed data can leave the statistical utility of the features intact but can remove individual identification risk of each unique data point. By establishing a direct and accurate mapping between multi-source perception and educational quality quantification, this work mirrors the high-precision alignment of fiber morphology and yarn spatial distribution required for manufacturing excellence. This integrated approach provides educational authorities with data-driven engineering solutions that ensure the same level of structural robustness and interpretability found in technical optimization, thereby facilitating informed decisions for the advancement of educational practices.

AUTHOR CONTRIBUTIONS

Conceptualization – Yang Yang, Jiaxin Chen and Chan Bai; methodology – Yang Yang, Jiaxin Chen and Chan Bai; investigation – Yang Yang, Jiaxin Chen and Chan Bai; writing-original draft preparation – Yang Yang, Jiaxin Chen and Chan Bai. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Xi'an Fanyi University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

REFERENCES

[1] P. Niu, “An artificial intelligence method for comprehensive evaluation of preschool education quality,” Frontiers in psychology, vol. 13, no. 1, pp. 955870-955876, 2022, doi: 10.3389/fpsyg.2022.955870.

[2] X. Wang, J. Zhao, Y. Lu, et al., “Spatial pattern, quality evaluation, and implications of preschool education facilities in new urban areas using multi-source data: A case study from Lingui New District in West China,” Buildings, vol. 14, no. 6, pp. 1718-1722, 2024, doi: 10.3390/buildings14061718.

[3] S. Wolf, H. Jukes M C, H. Yoshikawa, et al., “Examining the validity of an observational tool of classroom support for childrens engagement in learning,” Early Childhood Education Journal, vol. 53, no. 4, pp. 1325-1339, 2025, doi: 10.1007/s10643-024-01731-8.

[4] V. Huynh H, E. Puffer, J. Ostermann, et al., “A comparison of assessment tools for childcare centers in high vs,” low resource settings. Frontiers in Public Health, vol. 12, no. 1, pp. 1331423-1331429, 2024, doi: 10.3389/fpubh.2024.1331423.

[5] A. Samsul S, N. Yahaya, and H. Abuhassna, “Education big data and learning analytics: A bibliometric analysis,” Humanities and Social Sciences Communications, vol. 10, no. 1, pp. 709-715, 2023, doi: 10.1057/s41599-023-02176-x.

[6] Sun Lihua, “Evaluation literacy of preschool teachers under the digital perspective: development judgment and optimization path,” Educational Science, vol. 40, no. 3, pp. 90-95, 2024.

[7] Y. Zhou, S. Zou, M. Liwang, et al., “A teaching quality evaluation framework for blended classroom modes with multidomain heterogeneous data integration,” Expert Systems with Applications, vol. 289, no. 1, pp. 127884-127892, 2025, doi: 10.1016/j.eswa.2025.127884.

[8] C. Li, C. Liu, W. Ju, et al., “Prediction of teaching quality in the context of smart education: application of multimodal data fusion and complex network topology structure,” Discover Artificial Intelligence, vol. 5, no. 1, pp. 19-27, 2025, doi: 10.1007/s44163-025-00240-w.

[9] S. Prabowo, G. Putrada A, D. Oktaviani I, et al., “Privacy-preserving tools and technologies: Government adoption and challenges,” Ieee Access, vol. 13, no. 1, pp. 33904-33934, 2025, doi: 10.1109/ACCESS.2025.3540878.

[10] Y. Song and N. Wang, “Application of big data analytics in education management: Enhancing teaching quality and resource allocation efficiency,” International Journal of High Speed Electronics and Systems, vol. 35, no. 4, pp. 2540637-2540648, 2025, doi: 10.1142/S0129156425406370.

[11] T. Jiao, C. Guo, X. Feng, et al., “A comprehensive survey on deep learning multi-modal fusion: Methods, technologies and applications,” Computers, Materials & Continua, vol. 80, no. 1, pp. 1-9, 2024, doi: 10.32604/cmc.2024.053204.

[12] E. Dritsas and M. Trigka, “Big data analytics in e-learning: ethical challenges and opportunities for engineering education,” AI and Ethics, vol. 6, no. 1, pp. 4-8, 2026, doi: 10.1007/s43681-025-00866-7.

[13] D. Hu, J. Liu, and W. Hu, “A classroom interaction behavior analysis method based on image processing and artificial intelligence,” Traitement du Signal, vol. 41, no. 6, pp. 3173-3182, 2024, doi: 10.18280/ts.410633.

[14] B. Elbaum, K. Perry L, and S. Messinger D, “Investigating childrens interactions in preschool classrooms: An overview of research using automated sensing technologies,” Early childhood research quarterly, vol. 66, no. 1, pp. 147-156, 2024, doi: 10.1016/j.ecresq.2023.10.005.

[15] T. Shaik, X. Tao, Y. Li, et al., “A review of the trends and challenges in adopting natural language processing methods for education feedback analysis,” Ieee Access, vol. 10, no. 1, pp. 56720-56739, 2022, doi: 10.1109/ACCESS.2022.3177752.

[16] G. BREVIÁRIO A, “Natural language processing and computational tools in portuguese language teaching: exploring text mining and sentiment analysis in the development of communicative skills,” Current Scientific Journal, vol. 5, no. 1, pp. 370-390, 2025.

[17] W. Chango, A. Lara J, R. Cerezo, et al., “A review on data fusion in multimodal learning analytics and educational data mining,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 12, no. 4, Art. no. e1458-e1463, 2022, doi: 10.1002/widm.1458.

[18] J. Moon, S. Yeo, K. Banihashem S, et al., “Using multimodal learning analytics as a formative assessment tool: Exploring collaborative dynamics in mathematics teacher education,” Journal of Computer Assisted Learning, vol. 40, no. 6, pp. 2753-2771, 2024, doi: 10.1111/jcal.13028.

[19] R. Marshall, A. Pardo, D. Smith, et al., “Implementing next generation privacy and ethics research in education technology,” British Journal of Educational Technology, vol. 53, no. 4, pp. 737-755, 2022, doi: 10.1111/bjet.13224.

[20] V. Balachandar and K. Venkatesh, “Privacy-enhanced secure framework for educational data protection and analysis,” International Journal of Information Technology, vol. 17, no. 5, pp. 2887-2904, 2025, doi: 10.1007/s41870-025-02458-4.

[21] R. Subramanian, “Have the cake and eat it too: Differential Privacy enables privacy and precise analytics,” Journal of Big Data, vol. 10, no. 1, pp. 117-125, 2023, doi: 10.1186/s40537-023-00712-9.

[22] Y. Zhao and J. Chen, “A survey on differential privacy for unstructured data content,” ACM Computing Surveys (CSUR), vol. 54, no. 10s, pp. 1-28, 2022, doi: 10.1145/3490237.

[23] S. Kumar G, K. Premalatha, U. Maheshwari G, et al., “Differential privacy scheme using Laplace mechanism and statis tical method computation in deep neural network for privacy preservation,” Engineering Applications of Artificial Intelligence, vol. 128, no. 1, pp. 107399-107405, 2024, doi: 10.1016/j.engappai.2023.107399.

[24] G. Muthukrishnan and S. Kalyani, “Grafting Laplace and Gaussian distributions: A new noise mechanism for differential privacy,” IEEE Transactions on Information Forensics and Security, vol. 18, no. 1, pp. 5359-5374, 2023, doi: 10.1109/TIFS.2023.3306159.

[25] K. Assa-Agyei, F. Olajide, and T. Alade, “Optimizing the performance of the advanced encryption standard techniques for secured data transmis-

sion," International Journal of Computer Applications, vol. 185, no. 21, pp. 31-36, 2023, doi: 10.5120/ ijca2023922941.

[26] B. Sarkar, A. Saha, D. Dutta, et al., "A survey on the advanced encryption standard (AES): a pillar of modern cryptography," International Journal of Computer Science and Mobile Computing, vol. 13, no. 4, pp. 68-87, 2024, doi: 10.47760/ijcsmc.2024.v13i04.008.

[27] H. Wu, X. Ma, and Y. Li, "Spatiotemporal multimodal learning with 3D CNNs for video action recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 3, pp. 1250-1261, 2021, doi: 10.1109/TCSVT.2021.3077512.

[28] A. Sajitha, K. Rao R A V, K. Chandana, et al., "Speech emotion recognition for adaptive learning experiences using LSTM & MFCC," Applications of AI in Smart Technologies and Manufacturing, vol. 2, no. 1, pp. 70-78, 2025, doi: 10.1201/9781003682325-9.

[29] S. Shreyashree, P. Sunagar, S. Rajarajeswari, et al., "A literature review on bidirectional encoder representations from transformers," Inventive Computation and Information Technologies: Proceedings of ICICIT 2021, vol. (1), pp. 305-320, 2022, doi: 10.1007/978-981-16-6723-7_23.

[30] P. Nasiopoulou, E. Mellgren, S. Sheridan, et al., "Conditions for childrens language and literacy learning in Swedish preschools: Exploring quality variations with ECERS-3," Early childhood education journal, vol. 51, no. 7, pp. 1305-1316, 2023, doi: 10.1007/s10643-022-01377-4.

[31] E. Shmidt and C. Jacobson B, "Double-blind reviews: a step toward eliminating unconscious bias," Clinical and translational gastroenterology, vol. 13, no. 1, Art. no. e00443-e00452, 2022, doi: 10.14309/ctg.0000000000000443.

[32] M. Sun, J. Barry Danfa, and M. Teplitskiy, "Does double-blind peer review reduce bias? Evidence from a top computer science conference," Journal of the Association for Information Science and Technology, vol. 73, no. 6, pp. 811-819, 2022, doi: 10.1002/asi.24582.

[33] O. Adefemi K and B. Mutanga M, "A robust hybrid CNN-LSTM model for predicting student academic performance," Digital, vol. 5, no. 2, pp. 16-18, 2025, doi: 10.3390/digital5020016.

[34] D. Valkenborg, J. Rousseau A, M. Geubbelmans, et al., "Support vector machines," American journal of orthodontics and dentofacial orthopedics, vol. 164, no. 5, pp. 754-757, 2023, doi: 10.1016/j.ajodo.2023.08.003.