ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

The Construction and Innovative Practice of Multimodal Data-Driven Intelligent Evaluation System of University Labor Education

Mi Ma[1,*]

[1] Department of General Courses, Xi'an Traffic Engineering University, Xi'an 710300, Shaanxi, China

SUMMARY: *In view of the demand for computable evaluation in university labor education, this paper proposes a multi-modal data-driven intelligent evaluation system for learning scenarios. The system integrates video action streams, operation logs, task submission records, collaboration trajectories and reflective texts into the feature space. The labor behavior encoder extracts posture change, tool use sequence, completion time, process consistency, and reflective semantic features. Attention fuses the quality of network evaluation participation, task completion, safety compliance, collaboration status and depth of reflection to generate evidence evaluation scores for teachers and students. A data set containing 4280 valid records was constructed from 126 undergraduates, covering tasks such as campus cleaning, green plant maintenance, equipment maintenance, community service and handcrafting. Experimental results show that the proposed model achieves 93.2% Accuracy, 0.914 F1-score and 0.176 MAE, which is 5.8, 6.4 and 3.9 percentage points higher than the CNN-LSTM, Text-BERT and Late-FusionNet baselines, respectively. It presents a stable evaluation performance.*

KEYWORDS: *Multi-modal data; Labor education evaluation; Feature fusion; Intelligent feedback*

1 Introduction

The evaluation of labor education in colleges and universities needs to face real tasks, field operations and process performance, and a single score record is difficult to express the differences in students 'labor cognition, action norms, collaborative participation and reflection expression. The platform accumulates check-in records, task submissions, image clips, tool usage traces, mutual review texts, and teacher review labels, and these data are temporal, behavioral, and semantic. If the manual scale is used to complete the summary, the results are easy to stay at the score level. Visual recognition, natural language processing, behavior sequence modeling and multi-modal feature fusion provide a computable path for the evaluation of labor education in colleges and universities, which makes the evaluation system shift from experience recording to data-driven intelligent analysis.

Yağcı studied the predictive effect of machine learning algorithms on student performance, and showed that learning behavior data can be transformed into quantifiable evaluation results through classification models [1]. Baashar et al. proposed to use artificial neural network to predict student performance, proving that the nonlinear model can deal with complex learning feature relationships [2]. Sghir et al. reviewed the development progress of predictive

*mami0613@126.com

https://doi.org/10.65102/is2026923 Published: 30 April, 2026

learning analysis and pointed out that learning platform data, behavioral data and model inference can jointly support educational decision-making [3]. Khalil et al. analyzed the application of learning theory in learning analysis, which provided a theoretical mapping basis for the evaluation model to explain students' behaviors [4]. These studies provide algorithm reference for the evaluation of labor education in colleges and universities. However, labor education has the characteristics of long operation chain and scattered evaluation evidence, so it is still necessary to establish a multimodal computing framework for labor tasks.

Data security, privacy protection, and user trust become technical constraints in system design. Starting from student data privacy in learning analytics, Prinsloo et al. proposed that technical solutions need to take into account both data governance and educational usage boundaries [5]. Mutimukwe et al. constructed a privacy concern model for students' learning analysis, emphasizing that the user's perception of the data use process will affect the system acceptance [6]. Nazaretsky et al studied the trust mechanism of teachers in artificial intelligence educational technology, and pointed out that the model output needs to be understandable and professional reviewable [7]. Therefore, the intelligent evaluation of labor education not only pursues classification accuracy, but also provides traceable evidence, visual feedback and teacher review entry, so that the algorithm conclusion can enter the curriculum management.

The development of multimodal learning analysis provides a technical foundation for labor education evaluation. Ouhaichi et al. systematically mapped the research trend of multi-modal learning analysis and showed that the fusion of video, audio, text, sensing and interaction log could enhance the ability to identify the learning process [8]. Giannakos and Cukurova discussed the role of learning theory in multimodal learning analysis, indicating that multi-source data needs to be organized by educational semantics to form stable evaluation explanations [9]. Prinsloo et al. analyzed the privacy boundary in multimodal learning analysis and suggested that the use of visual and behavioral data must include anonymization, authorization and minimization [10]. Labor education in colleges and universities involves tasks such as cleaning and finishing, planting maintenance, equipment maintenance, community service and manual production. The evaluation model needs to deal with information such as action duration, operation sequence, collaborative interaction, outcome quality and text reflection.

Based on the above research basis, this paper constructs a multimodal data-driven intelligent evaluation system for university labor education, which maps image sequence, task log, process time, mutual evaluation record and reflection text into a unified feature vector. The system uses a multimodal coding module to extract labor action, task progress, collaboration status and semantic expression features. The participation intensity, standardization degree, continuous engagement and reflection depth are identified by behavior state modeling, and the attention fusion network is used to generate evaluation levels, confidence levels and evidence weights. The platform is set up with student feedback interface, teacher review interface, data management module and visual report module, so that the evaluation results reflect the individual labor process and service curriculum quality analysis. The objectives of this paper include: verifying the applicability of multi-modal feature encoding; The intelligent evaluation model and system process are constructed. Performance and stability were evaluated by model comparison, result analysis and module ablation test.

2 Related work

Learning analytics dashboards have moved from statistical interfaces to learning process explanation systems. Paulsen and Lindsay studied the evolution of learning analysis dashboard and pointed out that the dashboard design was shifting from data presentation to learning behavior support, which provided interface basis for labor education evaluation feedback [11]. Palanci et al. studied the application of learning analysis in distance education, sorted out the relationship between data tracking, behavior prediction and learning support, and showed that online logs and task process data could enter the analysis process [12]. Sakr and Abdullah studied the impact of virtual reality, augmented reality and learning analytics on learners and educators, and showed that interaction records, action traces and feedback data generated by immersive environments can be used as process evaluation evidence [13]. These results illustrate that labor tasks, field operations, and reflective texts in college labor education can be transformed into model inputs through platform logs, image sequences, and semantic features.

To present the correspondence between the existing research and the task of this paper, Table 1 summarizes the common data sources, calculation methods, and transfer points in learning analytics and intelligent evaluation. The comparison shows that performance prediction, real classroom analysis, dashboard feedback and generative modeling can all provide method references for the intelligent evaluation of labor education, and labor education also needs to add scene characteristics such as action specification, task continuity, collaboration status and outcome quality.

Table 1: Correspondence between learning analytics and intelligence evaluation research

Research Direction	Main Data Sources	Computational Methods	Implications for Labor Education Evaluation
Student performance prediction	Grades, participation records, assignment data	Classification algorithms, deep models	Supports grade-level prediction
Real classroom analysis	Videos, interaction logs, classroom events	Behavior recognition, large-scale analysis	Supports state recognition
Learning analytics dashboards	Platform logs, feedback records	Visualization modeling, user analysis	Supports result interpretation
Multi-scenario learning analytics	VR/AR interactions, online behaviors	Multimodal fusion, temporal modeling	Supports process tracking

Khairy et al. studied the application of data mining classification algorithms in student test performance prediction, and showed that methods such as decision tree, random forest and support vector machine can complete classification in structured educational data [14]. Fazil et al. proposed a deep learning model for student performance prediction by using participation data, emphasizing the combined expression ability of participation data [15]. Patidar et al. proposed Edulyze real classroom learning analysis system, which transformed classroom behavior data into scalable analysis results and provided systematic ideas for batch evaluation [16]. Haas et al. studied the Bayesian generative modeling method of student outcomes in course networks, and showed that probabilistic inference can express performance differences [17]. Schneider and Sung studied the influence of teacher's face and gaze on learning effect in online videos, and proved that visual cues can participate in behavior interpretation [18]. These studies provide algorithm support for the recognition of task completion, behavior persistence and interaction quality in labor education evaluation.

In the application level of evaluation system, Kotorov et al. studied a multi-case scheme to support teaching and learning innovation decision-making, indicating that learning analysis results need to be transformed into executable management and feedback information [19]. Freitas et al. proposed MMALA learning analytics maturity model to evaluate the ability level of institutions to adopt learning analytics, which provides reference for data governance and system deployment of labor education platforms in colleges and universities [20]. De Vreugd et al. studied the design and evaluation of learning analysis dashboard to support students' self-regulation, emphasizing that the feedback interface should present understandable behavioral evidence and action suggestions [21]. In summary, the existing research has formed the technical foundation of data prediction, process identification and feedback visualization. For labor education in colleges and universities, the evaluation model still needs to connect multi-modal coding, behavior state modeling, fusion reasoning and visual feedback into a complete system, so that actions, texts, logs and results form an intelligent evaluation link that can be calculated, reviewed and updated sustainably.

3 Multimodal data-driven intelligent evaluation model and system design of university labor education

3.1 Multi-modal feature coding method for Labor education evaluation

The evaluation of labor education in colleges and universities involves many links such as classroom organization, field operation, results submission and reflection expression, and the original data is obviously heterogeneous. The image sequence can record the student's action posture, tool use and achievement status, the operation log can reflect the task progress sequence, the text material can show the labor understanding and reflection depth, and the mutual evaluation record can supplement the peer feedback in the collaboration process. The core task of multi-modal feature coding is to transform these data with different sources, structures and time granularities into a unified vector, so that the system can complete labor behavior recognition, process state judgment and evaluation results generation in the same computing space.

In order to make labor process evidence from different sources enter the unified computing space, a sample container with time index should be established first, as shown in the following equation:

$$X_i = \{x_{i,t}^{v}, x_{i,t}^{o}, x_{i,t}^{r}, x_{i,t}^{s}\}_{t=1}^{T_i} \tag{1}$$

where X_i represents the set of multimodal samples of the i student in one labor task; $x_{i,t}^{v}$ represent image or video frame features; $x_{i,t}^{o}$ denote the operation log characteristics; $x_{i,t}^{r}$ represent achievement records and mutual evaluation information; $x_{i,t}^{s}$ represent the semantic features of the reflective text; T_i represents the number of task duration slices. This formula is used to fix the basic organization form of the sample, so that visual, text, log and evaluation records no longer exist as independent files, but enter a unified sample unit with time order and task semantics.

In order to clearly present the processing process before the multimodal features enter the model, Fig. 1 uses a hierarchical process structure to show the coding link. The left side of the figure is the original data entry in the labor education scene, including the scene image frame, operation log, task achievement record, student reflection text and peer assessment information. The middle layer was the data processing layer, which completed identity desensitization, time alignment, segment segmentation, modal coding and quality verification in turn. The right side is the unified feature output layer, where the system converts different modalities into labor evaluation vectors of the same dimension and writes them into the feature cache synchronously.

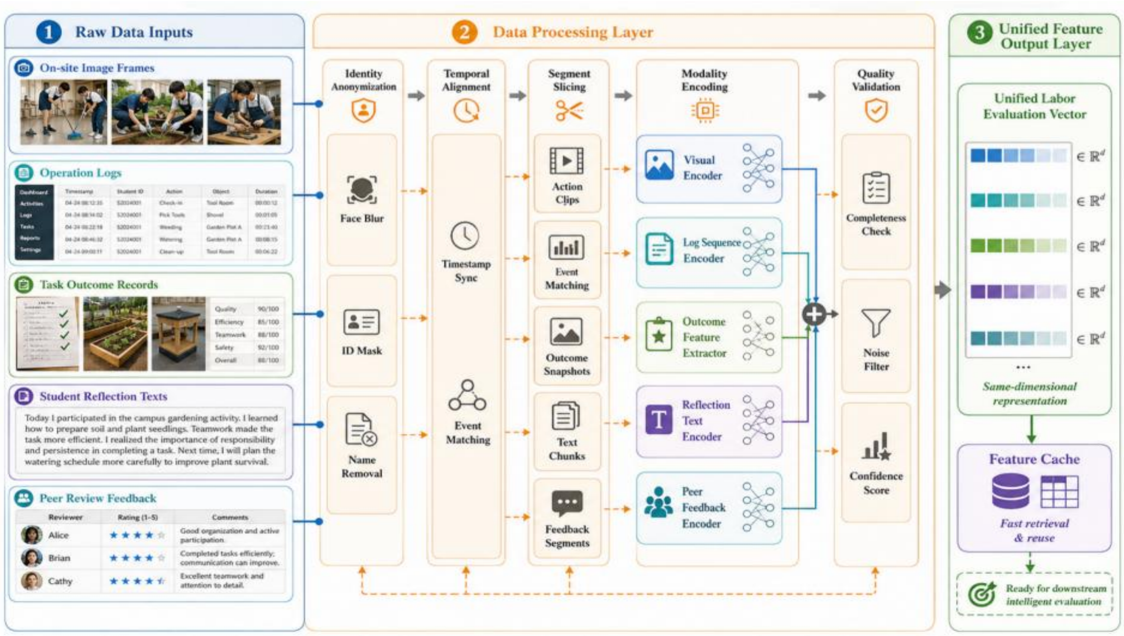

Figure 1: Multimodal feature coding process for labor education evaluation

Due to the different generation frequencies of image, log, and text data, a temporal kernel function is needed to complete the cross-modal alignment process, as shown in the following equation:

$$\hat{x}^{m}_{i,t} = \sum_{\tau=1}^{T_i} \kappa_m(t,\tau)\, x^{m}_{i,\tau} \tag{2}$$

Here, $\hat{x}^{m}_{i,t}$ denote the alignment result of mode m on a uniform time slice t. $\kappa_m(t,\tau)$ is the mode-dependent temporal kernel weight. $x^{m}_{i,\tau}$ denote the observations at the original time point τ. This equation can alleviate the time difference between the sparse log, the text lag and the dense image, so that the data of students in receiving tools, completing steps, submitting results and filling in reflections can fall into the same evaluation window.

On the basis of temporal alignment, each modality still needs to retain its own expression characteristics. In this paper, a special encoding function is set to complete the representation learning, as shown in the following equation:

$$e_i^m = G_m(\hat{x}^{m}_{i,1:T_i}; \theta_m) \tag{3}$$

Here, e_i^m denotes the encoding vector of mode m; G_m represents the feature encoder of the corresponding modality. Let θ_m denote the trainable parameters. The visual branch extracts action pose, tool area and result appearance, the log branch extracts operation sequence, dwell time and submission node, and the text branch extracts topic semantics, reflection level and task matching degree. This formula undertakes the transformation task from original evidence to deep features, which is the basis for the intelligent evaluation model to form effective input.

There will be occlusions, omissions, repeated submissions and differences in text length in the labor education scene, so the quality gating mechanism needs to be introduced after coding, as shown in the following equation:

$$\alpha_i^m = \frac{\exp(w_m^{\top}[e_i^m; q_i^m; c_i])}{\sum_{m' \in M} \exp(w_{m'}^{\top}[e_i^{m'}; q_i^{m'}; c_i])} \tag{4}$$

Here, α_i^m denotes the mass weight of modality m. q_i^m represents the data integrity, clarity and time coverage of the mode. c_i stands for labor task scene encoding; w_m represents the gating parameter; Let M denote the set of modes. This formula does not process all data equally, but dynamically assigns weights according to the modal quality to reduce the interference of low-definition images, abnormal logs and invalid texts on the evaluation results.

The final output of the multi-modal coding needs to enter the unified evaluation space and keep the numerical scale stable. The fusion expression is shown in the following equation:

$$z_i = \mathrm{LayerNorm}\left(\sum_{m \in M} \alpha_i^m W_m e_i^m + b_c\right) \tag{5}$$

Here, z_i represents the unified labor evaluation feature vector of the i student; W_m represents modal mapping matrix; b_c represents task scene bias; $\mathrm{LayerNorm}$ represents the layer normalization operation. This formula compreses the multi-source labor evidence into a unified vector, so that the data of different students, different tasks and different modalities can participate in the calculation in the same model.

This coding method makes the evaluation of labor education in colleges and universities from a single record to the collaborative expression of multi-source evidence. The system saves the desensitized vector, time slice index, task label and quality weight, and does not directly rely on manual summary scores. The coding results can not only support the automatic calculation of the model, but also retain the evidence sources required by the teacher's review. For labor tasks such as cleaning and finishing, planting maintenance, equipment maintenance, community service and manual production, the proposed method can integrate action specification, process continuity, result quality and reflection expression into the same evaluation entry, ensuring that the subsequent intelligent evaluation has a stable data basis and interpretable calculation basis.

3.2 Labor education learning behavior recognition and process state modeling

The learning behavior of labor education is continuous and scene, and a single action or a single log cannot fully express the student's labor performance. Students may experience states such as task understanding, tool preparation, action execution, peer collaboration, examination of outcomes, and reflective revision in the same task. The goal of learning action recognition is to convert the encoded multi-modal vectors into specific action categories. The goal of process state modeling is to judge whether student participation is stable, whether operation conforms to task flow, whether collaboration actually occurs, and whether results are consistent with process evidence in time series. This module undertakes the intermediate computational tasks from "feature input" to "evaluation evidence".

In order to establish the continuous connection between labor behavior and historical state, the system is represented by the gated timing unit update fragment, as shown in the following equation:

$$h_{i,t} = \mathrm{GRU}([z_{i,t}; a_{i,t}; g_{i,t}], h_{i,t-1}) \tag{6}$$

where $h_{i,t}$ denotes the hidden state of the i student at time slice t; $z_{i,t}$ denote the multimodal evaluation vector; $a_{i,t}$ denotes the action embedding; $g_{i,t}$ denote the task phase embedding; $h_{i,t-1}$ denotes the state memory of the previous time slice; GRU stands for Gated recurrent unit. This formula can preserve the relationship in the labor process, and make the model distinguish different states such as normal waiting, process stagnation, repeated correction and collaborative handover.

To illustrate the relationship between learning behavior recognition and state modeling, Fig. 2 uses a timeline structure to show the whole calculation process. In the figure, the unified labor evaluation feature vector is input to the left, and the labor process segments are segmented according to a fixed window in the middle, and each segment is sent to the time series state unit. The upper branch outputs behavioral categories such as tool use, outcome collation, collaborative communication, waiting for pauses, and reflective recording. The lower branch holds the state probability curve, which is used to observe the variation of student participation in different stages. The right side outputs the process state label, confidence, and reviewable evidence fragments, so that the model results can correspond to specific labor scenarios.

Figure 2: Labor education learning behavior recognition and process state modeling process

After obtaining the hidden state, the system needs to map each time slice into a labor learning behavior category, and the classification calculation is as follows:

$$P(y_{i,t}=c)=\frac{\exp(\phi_c^{\mathrm{T}}h_{i,t}+\mu_c)}{\sum_{j=1}^{C}\exp\left(\phi_j^{\mathrm{T}}h_{i,t}+\mu_j\right)} \tag{7}$$

Here, $P(y_{i,t}=c)$ represents the probability that time slice t belongs to the c labor behavior. ϕ_c and μ_c denote the class parameters; C represents the total number of behavioral categories. The output of this formula is not a simple label, but a recognition result with probability distribution, which can reflect the confidence degree of the model on the judgment of tool use, collaborative communication, result collation and reflection record.

In order to avoid state jump caused by short-term occlusion or accidental actions, adjacent segment constraints are added to the process state modeling, as shown in the following equation:

$$\mathcal{L}_s=\sum_{t=2}^{T_i}\gamma_{i,t}\left\|P_{i,t}-A_{g_{i,t}}P_{i,t-1}\right\|_2^2 \tag{8}$$

Here $\mathcal{L}_s$ represents the state smoothing loss; $P_{i,t}$ denotes the action probability vector of the current time slice; $P_{i,t-1}$ denotes the behavior probability vector of the previous time slice; $A_{g_{i,t}}$ denotes the state transition matrix associated with the task phase; Let $\gamma_{i,t}$ denote the fragment credibility weight. This formula can restrain the unreasonable state jump, make the normal transition in the labor process more stable, and reduce the influence of low quality images and missing logs on the state judgment.

To further illustrate how the state results enter the evaluation feedback, Fig. 3 shows the state mapping and evidence traceback structure. The left part of the figure is the behavior probability curve arranged by time, and the middle part is the key evidence fragments screened by the model, including high confidence fragments, low confidence fragments and teacher review fragments. On the right side is the output area of evaluation results, including process scores, state proportion, risk tips and visual feedback reports. This structure enables teachers to see the behavioral sources behind the scores, and also enables students to understand the specific labor process corresponding to the evaluation results.

Figure 3: State mapping and evaluation feedback structure of labor education process

The final evaluation needs to integrate the behavior probability, state duration, task phase weight and achievement verification information, and the process score is calculated as follows:

$$S_i = \sum\nolimits_{c=1}^{C} \omega_c \frac{\sum_{t=1}^{T_i} P\,(y_{i,t} = c)\Delta t_{i,t}}{\sum_{t=1}^{T_i} \Delta\, t_{i,t}} + \lambda Q_i \tag{9}$$

where S_i represents the labor process evaluation score of the ith student; ω_c denotes the action class weight; $\Delta t_{i,t}$ denotes the time slice length; Q_i refers to the result quality and reflection consistency verification results; Let λ denote the coefficient of the check term. This equation puts process behavior, ongoing engagement, and outcome evidence into the same scoring framework, avoiding evaluation relying only on outcome submission or teacher impression.

In this modeling method, the behavior recognition, status tracking and evaluation generation in labor education are connected as a continuous computing link. The model records not only whether the student completes the task, but also whether the action is standardized, whether the collaboration occurs, whether the time investment is stable, and whether the reflection content corresponds to the labor process. The state curve, fragment evidence and evaluation score of the system output together constitute the reviewable result. For the evaluation of labor education in colleges and universities, this method can transform the implicit performance in the labor process into structured data, which not only retains the real scene of practical activities, but also meets the technical requirements of stability, interpretability and scalability of the intelligent evaluation system.

3.3 Multi-source feature fusion and intelligent evaluation algorithm design

The task of multi-source feature fusion is to establish computable associations between labor behavior states, operation logs, outcome quality, reflective semantics, and peer reviews. The evaluation of labor education can not only read the strong signal of a single mode, and the image fragment can reflect the action posture, but it is difficult to explain the labor understanding. Textual reflection can show cognitive processing, but it may lag behind field operation. Logging can mark task nodes, but it cannot determine the quality of actions alone. Therefore, the intelligent evaluation algorithm needs to establish dynamic weights between multi-source evidence, retain high confidence evidence, weaken noise input, and map the fusion result into an interpretable evaluation level.

In order to characterize the relative contribution of visual, log, text and mutual evaluation evidence in the same labor task, and make the fusion weight change with the task scene and sample quality, this paper constructs a cross-modal attention allocation function, as shown in the following equation:

$$\pi_i^m = \frac{\exp((U_m z_i^m)^\top V c_i + \delta_m q_i^m)}{\sum_{r\in M} \exp\,((U_r z_i^r)^\top V c_i + \delta_r q_i^r)} \tag{10}$$

Here, π_i^m represents the fusion weight of student i on modality m, z_i^m represents the encoding vector of this modality, c_i represents the labor task scenario vector, q_i^m represents the modality quality factor, and U_m, V, and δ_m are trainable parameters. The formula combines scene semantics and data quality into the weight calculation, which is suitable for dealing with unbalanced evidence sources in tasks such as planting, cleaning, and equipment maintenance.

In order to avoid the loss of the original process information after multiple linear mappings of the fusion vector, the residual memory item is introduced into the algorithm, so that the high-level evaluation features can still retain the key traces in the behavior sequence, as shown in the following equation:

$$r_i = \mathrm{GELU}\left(W_f \sum_{m \in M} \pi_i^m z_i^m + W_h \bar{h}_i + b_f\right) + \xi z_i^{base} \tag{11}$$

Here, r_i represents the evaluation representation after fusion, $\bar{h}_i$ represents the time-averaged vector of student labor process states, z_i^{base} represents the basic task features, and W_f, W_h, b_f, and ξ are the model parameters. This formula combines multi-modal evidence, process state and task basic information to reduce the excessive influence of a single segment on the overall evaluation.

The evaluation grade output needs to reflect labor participation, norm execution, collaboration performance and reflection quality simultaneously. The model uses a hierarchical classification head to generate grade probabilities, as shown in the following equation:

$$\hat{p}_{i,k} = \frac{\exp(a_k^T \mathrm{Dropout}(r_i) + s_k)}{\sum_{j=1}^{K} \exp\left(a_j^T \mathrm{Dropout}(r_i) + s_j\right)} \tag{12}$$

Here, $\hat{p}_{i,k}$ represents the probability that student i belongs to the k evaluation level,K represents the number of levels, and a_k and s_k represent the classification parameters. The formula converts the fusion representation into a probability distribution, which facilitates the system to output the grades of excellent, good, qualified and need to improve, and retains the confidence information required by the teacher's review.

In order to constrain the consistency between classification results, rank order and confidence calibration, the joint loss function is used for model learning in the training phase, as shown in the following equation:

$$\mathcal{L} = \mathcal{L}_{ce} + \lambda_1 \sum_i \left| \sum_{k=1}^{K} k\, \hat{p}_{i,k} - y_i^{ord} \right| + \lambda_2 \sum_i \left(\max_k \hat{p}_{i,k} - d_i \right)^2 + \lambda_3 \|\Theta\|_2^2 \tag{13}$$

Here, $\mathcal{L}_{ce}$ represents the cross-entropy loss, y_i^{ord} represents the ordered rank labels formed by manual review, d_i represents the sample evidence agreement, Θ represents all training parameters. The formula takes into account classification accuracy, rank distance and confidence calibration, so that the model not only pursues label matching, but also pays attention to the continuous meaning of evaluation grades.

In order to enhance the interpretability of the evaluation results, the system calculates the evidence contribution of each mode to the final grade, which is convenient for teachers to view the score sources, as shown in the following equation:

$$\chi_i^m = \frac{\pi_i^m \|W_m z_i^m\|_1}{\sum_{r \in M} \pi_i^r \left\|W_r z_i^r\right\|_1 + \epsilon} \tag{14}$$

Here, χ_i^m represents the contribution proportion of mode m and ϵ is used to avoid the denominator being zero. This formula converts the abstract fusion results into the proportion of evidence that can be viewed, so that teachers can see whether an evaluation mainly comes from action clips, task logs, achievement records or reflective texts.

The multi-source fusion algorithm integrates visual, log, text and mutual evaluation evidence into a unified evaluation representation, and outputs the results through grade probability, confidence and modal contribution. This process not only retains the action process in the labor task, but also retains the reflective expression and collaboration evidence, so that the evaluation results no longer rely on a single data source. The corrected labels after teacher review can continue to be written into the sample pool, providing a stable basis for model iteration.

3.4 Architecture and functional modules of the Intelligent Evaluation System of University Labor Education

The intelligent evaluation system of university labor education adopts a hierarchical service architecture, which deploys data access, model reasoning, evaluation management and feedback separately. The student end is responsible for task viewing, process submission and feedback reading, the teacher end is responsible for project configuration, result review and class analysis, and the management end is responsible for data permissions, model versions and running audits. The architecture encapsulates multimodal input, algorithmic reasoning and visualization results as independent services, which facilitates the expansion of new labor items and evaluation metrics.

To illustrate the organizational relationship of the internal modules of the system, Fig. 4 adopts a four-layer structure of "user layer -- business layer -- algorithm layer -- data layer". The upper part of the figure is the entrance of the student end, teacher end and management end, the middle part is the task management, evaluation and review, feedback generation and report export module, the lower part is the multimodal coding, behavior recognition, fusion evaluation and model monitoring service, and the bottom part is the desensitization sample library, vector cache, evaluation result library and log audit library. The figure presents the complete path of the system from user action to model computation to result writeback.

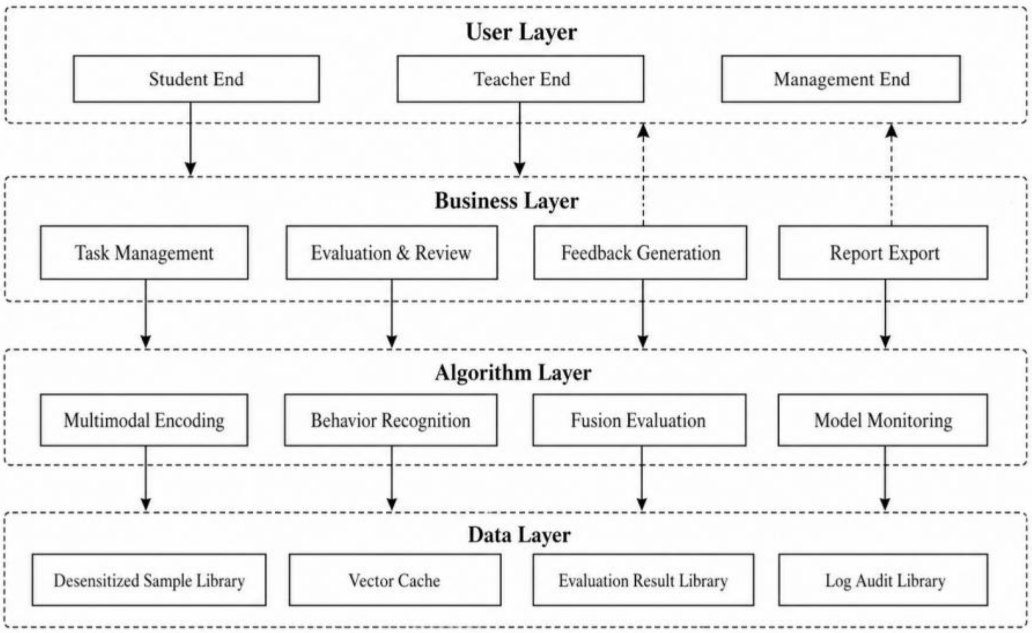

Figure 4: The overall architecture and functional modules of the intelligent evaluation system for labor education in colleges and universities

In order to ensure that the functions and data ranges accessed by different roles meet the system permission requirements, the platform uses role permission matrix to describe the access boundaries, as shown in the following equation:

$$\Omega_{u,j} = \mathbb{I}(\mathrm{role}_u \in R_j) \cdot \mathbb{I}(\mathrm{scope}_u \cap \mathrm{scope}_j \neq \emptyset) \cdot \zeta_{u,j} \quad (15)$$

where $\Omega_{u,j}$ represents the access permission of user u to function or data object j, role_u represents the user role, R_j represents the set of roles that the object is allowed to access, scope_u and sscope_j represent the scope of the organization and the scope of the course, $\zeta_{u,j}$ represents the quadratic verification factor. In this formula, students can only read personal tasks and feedback, teachers can only access authorized class data, and management end operates system resources according to audit rules.

In the system operation, the algorithm service needs to keep responding when multiple students submit tasks concurrently, so the service scheduling function is introduced to calculate the request priority of the model, as shown in the following equation:

$$\Psi_b = \rho_1 N_b + \rho_2 \bar{L}_b + \rho_3 C_b - \rho_4 H_b \quad (16)$$

Here, Ψ_b represents the scheduling priority of batch b, N_b represents the number of samples to be processed, $\bar{L}_b$ represents the average waiting time, C_b represents the urgency of the course task, H_b represents the cache hit rate, and ρ_1 to ρ_4 represent the scheduling coefficient. This formula is used to control the request order of multi-class tasks running at the same time, so that real-time feedback, batch grading and teacher review can share computing resources.

Interfaced data packet transmission is used between the front-end business module and the algorithm module, and the system writes the task number, sample index and model version into the reasoning request, as shown in the following equation:

$$\mathcal{B}_i = \{\mathrm{id}_i, \mathrm{task}_i, \mathrm{ver}_i, z_i, t_i, \mathrm{hash}_i\} \quad (17)$$

Here, $\mathcal{B}_i$ represents the i inference request packet, id_i represents the desensitization student identification, $,\mathrm{task}_i$ represents the labor task number, ver_i represents the model version, z_i represents the evaluation vector, t_i represents the submission timestamp, and hash_i represents the integrity check code. This formula makes the data exchange between each module have a traceable structure, and facilitates the back-end to locate the model version, sample source and result writing status.

To monitor whether the system meets the real-time feedback requirements, the platform records the end-to-end service time and splits it into four parts: encoding, reasoning, library writing and rendering, as shown in the following equation:

$$T_{sys} = T_{enc} + T_{inf} + T_{db} + T_{ui} + \varepsilon_{net} \quad (18)$$

Here, T_{sys} represents the total system time consumption of an evaluation request, T_{enc} represents the feature encoding time consumption, T_{inf} represents the model inference time consumption, T_{db} represents the database reading and writing time consumption, T_{ui}represents the interface rendering time consumption, and ε_{net} represents the network fluctuation term. This formula was used to locate the source of delay. If the inference time increased, the system could switch the batch queue. If the writing time is abnormal, the system calls the cache writeback mechanism.

The division of function modules makes the evaluation system of labor education in colleges and universities have engineering boundary. The student end pays attention to the task process and feedback explanation, the teacher end pays attention to the class evaluation distribution, low confidence samples and review records, and the management end pays attention to data permissions, model versions and operation logs. The algorithm service does not directly expose the original data, only receives the desensitization vector and the task index. The database holds the resulting scores, evidence fragments, model versions, and manual correction records. Such a system structure supports the operation of different labor courses, and makes the evaluation process consistent between curriculum implementation, model calculation, manual review and data governance.

3.5 Evaluation result generation, visual feedback and data management mechanism

The generation stage of evaluation results is responsible for transforming the model output into structured information that students can understand, teachers can review, and the system can save. The evaluation results of university labor education not only include a total score, but also should present dimensions such as task participation, step specification, collaborative contribution, quality of results and reflective matching. The system synchronously retains evidence fragments, modal weights, confidence and version identification when generating results, so that each evaluation can be traced back to the data source. The visual feedback module converts complex model inference results into charts, labels, and short suggestions, so that the evaluation information does not stay in the background.

To calculate a comprehensive evaluation score that can be used in the presentation of the report, the system incorporates the dimension score, modal contribution, and rank probability into the summary function, as shown in the following equation:

$$E_i = \sum_{d=1}^{D} \theta_d v_{i,d} + \sum_{m \in M} \varphi_m \chi_i^m + \sum_{k=1}^{K} \psi_k \hat{p}_{i,k} \tag{19}$$

Here, E_i represents the comprehensive evaluation score of student i, $v_{i,d}$ represents the score of the d evaluation dimension, χ_i^m represents the modal contribution proportion, $\hat{p}_{i,k}$ represents the grade probability, and θ_d, φ_m, and ψ_k represent the weight parameters. This formula unifies the behavior process, evidence contribution and grade prediction into the report output, so that the evaluation results are no longer determined by a single dimension.

In order to reduce the influence of model overconfidence on feedback judgment, the system performs temperature calibration on the output probability and generates a credible interval as shown in the following equation:

$$\tilde{p}_{i,k} = \frac{\exp(\log \hat{p}_{i,k} / \tau)}{\sum_{j=1}^{K} \exp(\log \hat{p}_{i,j} / \tau)}, \quad CI_i = [E_i - \omega\sigma_i, E_i + \omega\sigma_i] \tag{20}$$

Here, $\tilde{p}_{i,k}$ represents the calibrated grade probability, τ represents the temperature coefficient, CI_i represents the evaluation confidence interval, σ_i represents the sample uncertainty, and ω represents the interval adjustment coefficient. This formula enables the system to distinguish between high confidence evaluation and evaluation that needs to be reviewed, and avoids low quality samples from directly entering the final conclusion.

Visual feedback needs to maintain comparability between different dimensions. The system maps the raw dimension scores into a standardized presentation matrix, as shown in the following equation:

$$\mathrm{V}_{\mathrm{i,d}}^{*} = \frac{\mathrm{V}_{\mathrm{i,d}} - \min(\mathrm{V}_{:,\mathrm{d}})}{\max(\mathrm{V}_{:,\mathrm{d}}) - \min(\mathrm{V}_{:,\mathrm{d}}) + \epsilon} \tag{21}$$

where $\mathrm{V}_{\mathrm{i,d}}^{*}$ is the normalized display value, $\mathrm{V}_{\mathrm{i,d}}$ is the raw dimension score, $\max(\mathrm{V}_{:,\mathrm{d}})$ $\min(\mathrm{V}_{:,\mathrm{d}})$ are the maximum and minimum values of that dimension in the class sample, and ϵ is used to prevent the denominator from being zero. This formula ensures that task engagement, collaborative performance, operational norms, and reflective quality can be presented in the same diagram.

To generate personalized feedback, the system calculates feedback weights based on dimension gap, confidence level and teacher review priority, as shown in the following equation:

$$\mathrm{F}_{\mathrm{i,d}} = \eta_1(1 - \mathrm{V}_{\mathrm{i,d}}^{*}) + \eta_2(1 - \max_{\mathrm{k}} \tilde{\mathrm{p}}_{\mathrm{i,k}}) + \eta_3 \mathrm{R}_{\mathrm{i,d}} \tag{22}$$

Here, $\mathrm{F}_{\mathrm{i,d}}$ represent the feedback priority of the d dimension, $\mathrm{R}_{\mathrm{i,d}}$ represent the teacher review or task rule trigger mark, and η_1, η_2, and η_3 represent the weight coefficients. This formula is used to decide what content in the feedback report should be presented first, such as missing operational steps, insufficient evidence of collaboration, or reflection text inconsistent with the actual process.

The data management mechanism needs to support the long-term preservation of evaluation results, version tracking and sample updating. The system uses the life cycle index to describe the data status, as shown in the following equation:

$$\mathrm{D}_{\mathrm{i}}^{(\mathrm{t}+1)} = \mathcal{U}(\mathrm{D}_{\mathrm{i}}^{(\mathrm{t})}, \mathrm{E}_{\mathrm{i}}, \mathrm{CI}_{\mathrm{i}}, \mathrm{A}_{\mathrm{i}}, \mathrm{ver}_{\mathrm{t}}) \tag{23}$$

where $\mathrm{D}_{\mathrm{i}}^{(\mathrm{t})}$ represents the student labor evaluation file at time t, A_{i} represents the teacher review action, $\mathrm{ver}_{\mathrm{t}}$ represents the model version, and $\mathcal{U}$ represents the file update function. The formula writes the model results, confidence information, manual correction and version information into the same data chain, so that the subsequent query can restore the evaluation basis at that time.

The evaluation result generation and data management mechanism uniformly stored the model scores, evidence fragments, confidence intervals and manual review records. The student end gets understandable personal feedback, the teacher end gets traceable class analysis results, and the management end can view model versions and data flow records. The mechanism ensures that the intelligent evaluation results of labor education have the ability of readability, rechecking and continuous updating.

4 Results and discussion

4.1 Experimental data collection and sample construction

The experiment was carried out relying on the labor education course of a university, and 126 undergraduates participated in the experiment. The task scenarios included campus cleaning, green plant maintenance, equipment maintenance, community service and manual production. Each student completed check-in, tool collection, on-site operation, result submission, peer collaboration and reflective filling in according to a unified process, and the system synchronously recorded image sequences, operation logs, submission nodes, collaboration trajectories and text materials. The image data was collected by fixed camera equipment and mobile terminal, and segmented into action segments according to the 5s window. Log data records events such as confirmation, upload, modification and feedback viewing. The text data were segmented, denoised and semantic vectorized. Teachers reviewed and labeled according to task specification, outcome quality, collaboration performance and reflection content to form supervision labels required for model training. A total of 4280 valid samples were obtained in the experiment. After deleting missing time, serious screen occlusion, log breakage and invalid text records, the training set, validation set and test set were divided according to 7:2:1, and the proportions of the five types of tasks were kept consistent. In the training phase, brightness perturbation, random clipping and Angle fine-tuning are added to image segments, slight time shift is added to log sequences, and synonymous substitution is used to extend semantic coverage of text materials. All samples were processed by identity desensitization, time alignment, missing modal completion and normalization before entering the model. Finally, a unified sample tensor was generated for multi-modal coding, behavior state recognition, model comparison and ablation experiments. The experimental organization enables different modalities to participate in the calculation under the same labor task conditions, which ensures the comparability of the evaluation results and facilitates the reproduction and verification of the results.

4.2 Comparative analysis of intelligent evaluation models

The comparison experiment of intelligent evaluation model is used to test the adaptation ability of the model in the multimodal evaluation task of university labor education. The comparison objects include CNN-LSTM, Text-BERT, Late-FusionNet and the model of this paper. CNN-LSTM mainly processes image action sequences, Text-BERT is used for reflective Text semantic analysis, Late-FusionNet fuses multi-source prediction results at the decision level, and the proposed model introduces vision, log, Text and mutual evaluation information in the feature encoding stage, and combines process state modeling and confidence calibration to complete the evaluation output. All models use the same training set, validation set and test set, the training rounds, batch size and optimizer parameters are consistent, and the evaluation metrics include Accuracy, F1-score, MAE and average inference delay.

In order to present the differences in the comprehensive performance of different models, Fig. 5 uses radar charts to show five indicators: Accuracy, F1-score, MAE reverse score, stability and response efficiency. The radar chart can reflect the classification ability, error control and inference efficiency at the same time, which is suitable for observing the balance degree of different models in multi-index evaluation tasks. CNN-LSTM has a stable performance in the dimension of action recognition, but lacks text and collaboration evidence support. Text-BERT can identify the semantic differences in reflective texts, but it is difficult to explain the on-site labor process. Late-FusionNet has the ability of multi-source input, but the fusion takes place at the decision end, and the process state constraint is insufficient. The model in this paper has a more balanced distribution on the five indicators, the Accuracy reaches 93.2%, the F1-score is 0.914, the MAE is 0.176, and the average inference delay is 41ms.

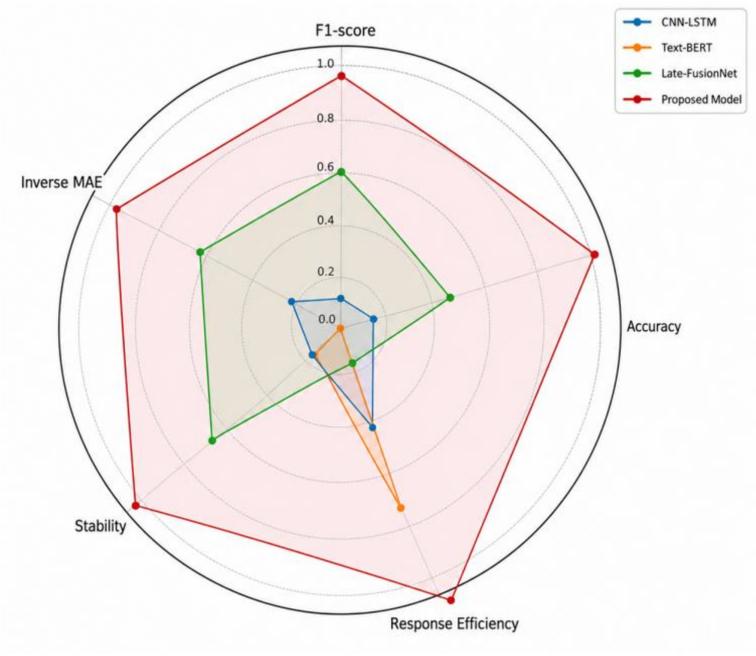

Figure 5: Radar chart of comprehensive performance of intelligent evaluation model

To further observe the dispersion degree of model errors, box plots are used in Fig. 6 to show the evaluation error distribution of the four types of models on the test set. The box plot can intuitively present the median error, the upper and lower interquartile range, and the number of outlier samples, which is suitable for analyzing the stability of the intelligent evaluation model in different labor tasks. The median error of CNN-LSTM is 0.241, Text-BERT is 0.263, Late-FusionNet is 0.205, and the proposed model is reduced to 0.157, and the number of outliers is relatively small. This result shows that feature-level fusion and state modeling can reduce the fluctuation caused by single modal distortion, and make the labor process score maintain more stable consistency with the teacher review label.

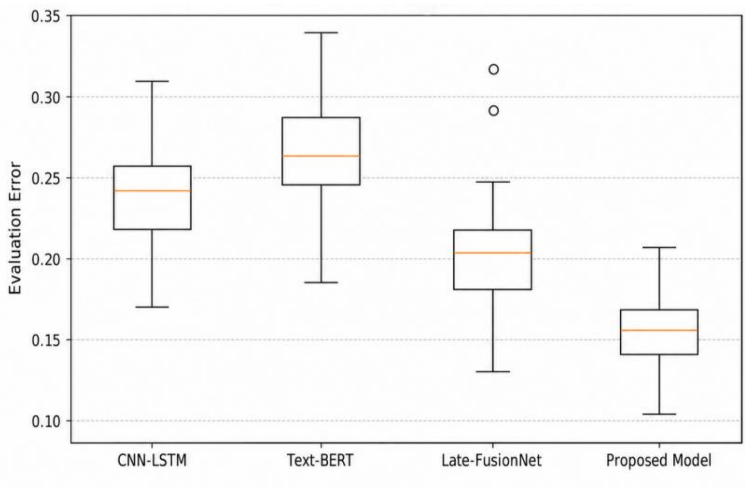

Figure 6: Box plots of error distribution for different intelligent evaluation models

Comprehensive comparison shows that compared with CNN-LSTM, Text-BERT and Late-FusionNet, the proposed model increases the Accuracy by 5.8, 6.4 and 3.9 percentage points respectively. The multi-modal coding retains the complementary information between actions, logs, achievements and texts, and the fusion evaluation algorithm further compresses the noise evidence, so that the system can maintain high recognition accuracy in multiple types of labor tasks. This comparison shows that the intelligent evaluation of labor education cannot only rely on a single modal feature, and the evaluation results are closer to the real labor performance after the process state, task log and text reflection are jointly involved in the modeling.

4.3 Analysis of multi-modal labor education evaluation results

The analysis of multimodal labor education evaluation results was carried out around five types of labor tasks, focusing on the differences in the output of the model in different task types and evaluation dimensions. The test samples cover campus cleaning, green plant maintenance, equipment maintenance, community service and handmaking, and each type of task contains image sequences, operation logs, achievement records, collaboration trajectories and reflective texts. The system outputs the scores of five dimensions of task participation, operation specification, collaboration contribution, achievement quality and reflective matching, and synchronously saves the proportion of modal contribution, grade probability and confidence interval. This analysis is able to test whether the model is able to generate interpretable evaluation results based on labor task differences.

To present the correspondence between different tasks and evaluation dimensions, Fig. 7 uses the heat matrix to show the normalized scores of five types of labor tasks on five dimensions. The rows in the heat matrix correspond to the labor task type, the columns correspond to the evaluation dimension, and the cell values represent the standardized scores after the model output. Campus cleaning reached 0.935 in the dimension of participation persistence, indicating that the process records in this type of task were relatively complete. The dimension of operation specification of equipment maintenance reached 0.918, reflecting that the sequence of tool use and step confirmation had a strong impact on the evaluation results. The collaborative contribution dimension of community service reaches 0.921, indicating that interaction log and peer review can better capture collaborative performance. Manual production reached 0.913 in the quality dimension of results, and green plant maintenance reached 0.894 in the reflective matching dimension, and the results were basically consistent with the task attributes.

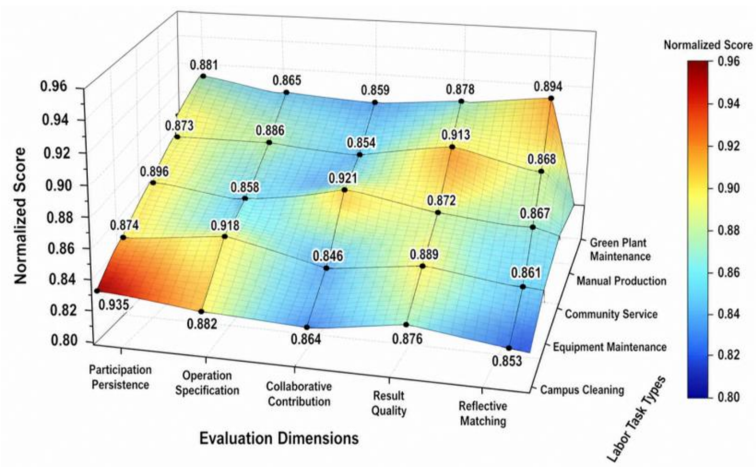

Figure 7: Multimodal labor education evaluation dimension heat matrix

To analyze the agreement between the intelligent evaluation ratings and the teacher review labels, Fig. 8 uses a confusion matrix to present the recognition distribution of the four categories of grades excellent, good, qualified, and need improvement. The confusion matrix can show the correct recognition proportion of samples at each level and the misjudgment of adjacent levels, which is suitable for observing the classification stability of the model on boundary samples. The recognition accuracy of excellent grade was 94.1%, good grade was 92.8%, qualified grade was 91.3%, and needed improvement grade was 88.6%. The misjudgment of the level to be improved is mainly concentrated near the qualified level, because the operation log of some samples is relatively complete, but the reflection text is not sufficiently corresponding to the result evidence, and the model determines it as a boundary state.

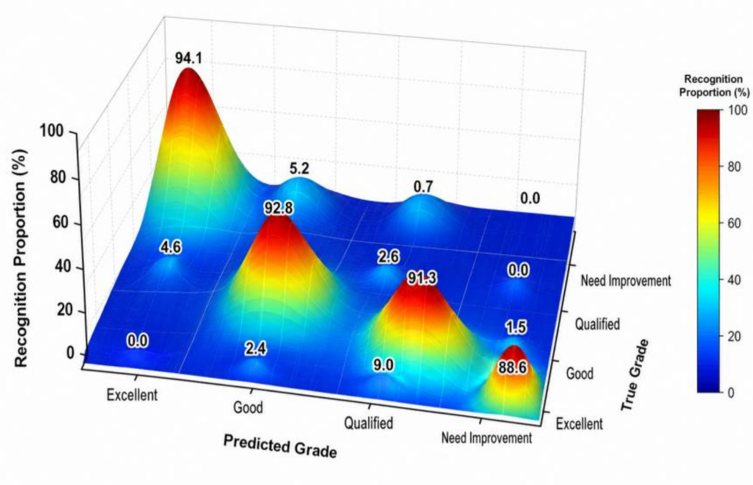

Figure 8: Confusion matrix of labor education evaluation grades

From the dimension distribution and level recognition results, the proposed model can better distinguish real participation, short-term operation, process make-up and reflection disconnection. Image sequences provide action basis, log sequences record task progress, text semantics supplement reflection depth, and mutual evaluation records correct collaboration performance. After the four types of evidence enter the evaluation together, the grade results output by the system are closer to the teacher's review label, and can also explain the source of students' scores. This result shows that the multimodal labor education evaluation can not only generate the final grade, but also retain the process evidence behind the evaluation.

4.4 Multi-modal module ablation and evaluation stability analysis

Multimodal module ablation experiments are used to examine the contribution of different input branches and computational units to the evaluation performance. The experiment was completed on the same training set, validation set and test set, and the visual branch, log branch, text branch, mutual evaluation branch, state modeling unit and confidence calibration unit were removed in turn, and compared with the complete model. The evaluation metrics are Accuracy, F1-score, MAE and stability index. The stability index is converted according to the standard deviation of three repeated experiments, and a higher value indicates a more concentrated model output. This experiment is able to judge the actual role of each module in labor process identification, evaluation error control and result stability.

To show the influence of each module on multiple indicators after ablation, Fig. 9 uses matrix score plots to present the changes of different model Settings in Accuracy, F1-score, MAE, and stability index. The row direction corresponds to different ablation Settings, the column direction corresponds to the four evaluation indicators, the specific values are marked in the cells, and the color depth indicates the range of performance change. After removing the state modeling, the Accuracy is reduced to 88.6%, the F1-score is reduced to 0.864, and the stability index is reduced to 0.884. After removing the visual branch, the MAE rose to 0.236, indicating that the scene action evidence had a strong supporting effect on the labor process evaluation. The complete model maintains the optimal combination of the four indicators, reflecting the collaborative relationship between the multimodal modules.

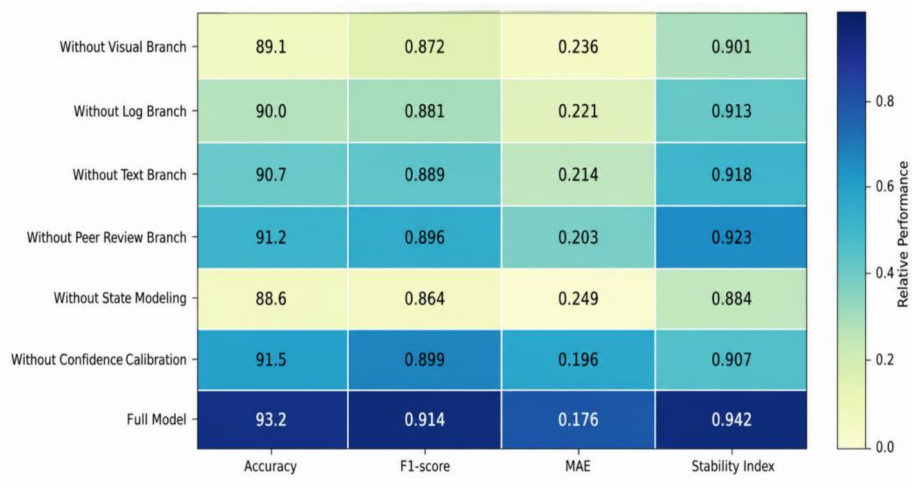

Figure 9: Matrix score plot of ablation experiments for multimodal modules

To further give specific values for ablation experiments, Table 2 lists the evaluation performance under different model Settings. It can be seen from the table that the performance decreases most obviously after removing state modeling, which indicates that the labor education evaluation needs to maintain the continuous relationship between behavioral segments. After removing the visual branch, the operation specification and action quality recognition are greatly affected. After removing the text branch, the error of the reflective matching dimension increases. After removing the confidence calibration, the Accuracy does not decrease much, but the stability index is significantly reduced.

Table 2: Results of ablation experiments for multimodal modules

Model Setting	Accuracy/%	F1-score	MAE	Stability Index
Without visual branch	89.1	0.872	0.236	0.901
Without log branch	90.0	0.881	0.221	0.913
Without text branch	90.7	0.889	0.214	0.918
Without peer-evaluation branch	91.2	0.896	0.203	0.923
Without state modeling	88.6	0.864	0.249	0.884
Without confidence calibration	91.5	0.899	0.196	0.907
Full model	93.2	0.914	0.176	0.942

Ablation results show that state modeling and visual branch contribute greatly to labor process recognition, log and text branch have complementary effects on task advancement and reflective matching, and mutual evaluation branch can correct collaborative performance judgment. Although confidence calibration has little impact on Accuracy, it is able to reduce the output fluctuation of boundary samples. The complete model combines multi-source evidence, state continuity and uncertainty control, and the evaluation output is more stable, which is also more suitable for multi-task evaluation scenarios in university labor education.

5 Conclusion

This paper focuses on the evaluation of labor education in colleges and universities, and constructs a multimodal data-driven intelligent evaluation system. The system encodes image sequence, operation log, achievement record, mutual evaluation trajectory and reflection text. After behavior state modeling, multi-source feature fusion and confidence calibration, evaluation levels and evidence fragments are generated. Based on 4280 valid samples from 126 undergraduates, the proposed model achieves 93.2% accuracy, 0.914 F1 score and 0.176 mean absolute error on the test set, which are better than CNN-LSTM, Text-BERT and Late-FusionNet. The experimental results show that visual evidence supports action specification recognition, the log sequence presents the task progress process, and text semantics and mutual evaluation records supplement the depth of reflection and collaboration performance. After the multi-source evidence enters the model, the consistency between the evaluation results and the teacher's review label is stable, which facilitates the traceability of the score source.

The limitations are mainly reflected in three aspects. The sample is from a single university and covers five types of labor tasks, but the data distribution under different majors, site conditions and organizational methods still needs to be extended and verified. Image data is easily affected by occlusion, illumination and shooting Angle, and logs and text records are also affected by students 'submission habits. The model relies on supervised label training, and there is still room for expansion of the transfer ability to new labor items and small sample scenarios. Subsequent research can build a multi-modal data set of cross-school labor education, introduce lightweight reasoning, incremental learning and privacy protection training mechanisms at the edge, and write the teacher review results back to the model update process, so that the system can maintain stable operation in a larger platform.

Author's Profile

Mi Ma was born in in Xi 'an, Shaanxi Province, P.R. China, in 1986. She obtained a bachelor's degree from China Women’s University.I am currently studying at the Department of General Courses,XI’AN Traffic Engineering University.My main research direction is Educational Sociology and Social Work.

Funding

This work was supported by Research Project on Teaching Reform of Undergraduate and Continuing Education of Shaanxi Provincial Department of Education in 2025.（25BZ132）

References

[1] Yağcı M. Educational data mining: prediction of students' academic performance using machine learning algorithms[J]. Smart Learning Environments, 2022, 9(1): 11.

[2] Baashar Y, Alkawsi G, Mustafa A, et al. Toward predicting student’s academic performance using artificial neural networks (ANNs)[J]. Applied Sciences, 2022, 12(3): 1289.

[3] Sghir N, Adadi A, Lahmer M. Recent advances in predictive learning analytics: A

decade systematic review (2012–2022)[J]. Education and information technologies, 2023, 28(7): 8299-8333.

[4] Khalil M, Prinsloo P, Slade S. The use and application of learning theory in learning analytics: A scoping review[J]. Journal of Computing in Higher Education, 2023, 35(3): 573-594.

[5] Prinsloo P, Slade S, Khalil M. The answer is (not only) technological: Considering student data privacy in learning analytics[J]. British Journal of Educational Technology, 2022, 53(4): 876-893.

[6] Mutimukwe C, Viberg O, Oberg L M, et al. Students' privacy concerns in learning analytics: Model development[J]. British Journal of Educational Technology, 2022, 53(4): 932-951.

[7] Nazaretsky T, Ariely M, Cukurova M, et al. Teachers' trust in AI-powered educational technology and a professional development program to improve it[J]. British journal of educational technology, 2022, 53(4): 914-931.

[8] Ouhaichi H, Spikol D, Vogel B. Research trends in multimodal learning analytics: A systematic mapping study[J]. Computers and Education: Artificial Intelligence, 2023, 4: 100136.

[9] Giannakos M, Cukurova M. The role of learning theory in multimodal learning analytics[J]. British Journal of Educational Technology, 2023, 54(5): 1246-1267.

[10] Prinsloo P, Slade S, Khalil M. Multimodal learning analytics—In-between student privacy and encroachment: A systematic review[J]. British Journal of Educational Technology, 2023, 54(6): 1566-1586.

[11] Paulsen L, Lindsay E. Learning analytics dashboards are increasingly becoming about learning and not just analytics-A systematic review[J]. Education and Information Technologies, 2024, 29(11): 14279-14308.

[12] Palanci A, Yılmaz R M, Turan Z. Learning analytics in distance education: A systematic review study[J]. Education and Information Technologies, 2024, 29(17): 22629-22650.

[13] Sakr A, Abdullah T. Virtual, augmented reality and learning analytics impact on learners, and educators: A systematic review[J]. Education and Information Technologies, 2024, 29(15): 19913-19962.

[14] Khairy D, Alharbi N, Amasha M A, et al. Prediction of student exam performance using data mining classification algorithms[J]. Education and Information Technologies, 2024, 29(16): 21621-21645.

[15] Fazil M, Rísquez A, Halpin C. A novel deep learning model for student performance prediction using engagement data[J]. Journal of Learning Analytics, 2024, 11(2): 23-41.

[16] Patidar P, Ngoon T J, Vogety N, et al. Edulyze: Learning Analytics for Real-World Classrooms at Scale[J]. Journal of Learning Analytics, 2024, 11(2): 297-313.

[17] Haas M R, Caprani C, Van Beurden B. Bayesian generative modelling of student results in course networks[J]. Journal of Learning Analytics, 2023, 10(3): 135-152.

[18] Schneider B, Sung G. Is seeing the instructor’s face or gaze in online videos helpful for learning?[J]. Journal of Learning Analytics, 2024, 11(3): 210-223.

[19] Kotorov I, Krasylnykova Y, Pérez-Sanagustín M, et al. Supporting decision-making for promoting teaching and learning innovation[J]. Journal of Learning Analytics, 2024, 11(1): 21-36.

[20] Freitas E, Fonseca F, Garcia V, et al. MMALA: Developing and evaluating a maturity model for adopting learning analytics[J]. Journal of Learning Analytics, 2024, 11(1): 67-86.

[21] De Vreugd L, Van Leeuwen A, Jansen R, et al. Learning Analytics Dashboard Design and Evaluation to Support Student Self-Regulation of Study Behaviour[J]. Journal of Learning Analytics, 2024, 11(3): 249-262.