ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Research on the Construction of dynamic Teaching Evaluation Model of smart Classroom in Universities Driven by Deep Reinforcement Learning Algorithm

Danxia Li[1,*]

[1] School of Mathematics and Computer Science, Yan'an University, Yan'an 716000, Shaanxi, China

SUMMARY: *Aiming at the problems of static teaching evaluation, lagging results and insufficient utilization of multi-source data in smart classroom of colleges and universities, a dynamic teaching evaluation model driven by deep reinforcement learning was constructed. A dynamic evaluation index system was established around the five elements of teachers 'teaching behavior, students 'classroom participation, the quality of teacher-student interaction, the use of resources and the effect of classroom feedback. Multi-modal data such as classroom video, voice, platform log, interactive text and classroom test were integrated to complete classroom state representation and time series modeling. On this basis, the Actor-Critic structure and PPO strategy optimization mechanism were introduced to realize the dynamic output and feedback update of classroom evaluation. The experiment was carried out based on 8 courses and 96 classroom records, and a total of 41280 state samples were formed. The results show that the accuracy, recall rate and F1 value of the proposed model reach 0.923, 0.909 and 0.915 respectively, the mean absolute error and root mean square error are 0.071 and 0.118 respectively, and the average response delay of a single time window is 0.11 s. The stability index and dynamic adaptability index reach 0.94 and 0.92, respectively, which are better than the baseline models such as AHP, SVM, LSTM and Transformer. The research shows that the model can better support the continuous perception, dynamic evaluation and strategy optimization of the smart classroom teaching process.*

KEYWORDS: *Deep reinforcement learning; Smart classroom; Dynamic teaching evaluation; Multi-modal data fusion*

1 Introduction

With the continuous development of artificial intelligence, learning analytics and multimodal perception technology, university classrooms are shifting from the traditional experiential teaching field to data-driven, interaction traceable and process computable smart teaching scenarios. International research shows that the application of artificial intelligence in education has gradually extended from early results prediction and resource recommendation to teaching process analysis, learning behavior recognition, automatic feedback generation and classroom support optimization. Multi-modal learning analysis has also increasingly become an important direction of educational intelligence research, especially in higher education scenarios. The joint use of heterogeneous data such as video, voice, text, platform log and behavior trajectory provides a technical basis for dynamic identification and evaluation modeling of classroom

*yadxldx@163.com
https://doi.org/10.65102/is2026078
Published: 30 April, 2026

teaching status. However, from the existing research, the classroom teaching evaluation in colleges and universities is still mainly based on after-class summary, single rating or local behavior judgment, lacking an overall grasp of the continuous changes of elements such as teachers 'teaching behavior, students 'classroom participation, the quality of teacher-student interaction and feedback effect, and the evaluation process has problems of static, fragmentation and lag. At the same time, although the existing automated teaching evaluation methods have improved the evaluation efficiency and objectivity to a certain extent, most of them still rely on single modal information or limited evaluation dimensions, and have not yet formed a dynamic update mechanism for complex classroom situations, and it is difficult to realize the effective linkage between evaluation results and teaching regulation. Based on this, this paper takes the smart classroom of colleges and universities as the research scene, tries to introduce the deep reinforcement learning method to construct a dynamic teaching evaluation model, and fuses and represents multi-source data such as classroom video, voice, interactive text, platform log and test record to form a dynamic state space that can reflect the running state of the classroom. In order to realize the real-time generation, continuous update and strategy optimization of classroom evaluation, the teaching evaluation problem is transformed into a sequence decision problem in a continuous situation. This paper focuses on the refinement of classroom evaluation dimensions, the construction of state representation, the modeling and experimental verification of deep reinforcement learning, and forms the overall idea of "multi-source data input-classroom state modeling-strategy learning update-dynamic evaluation output". Compared with the existing research, the innovation of this paper is to promote the classroom teaching evaluation from result-oriented to process-oriented, introduce multimodal data fusion and state space representation into the smart classroom evaluation modeling, and apply deep reinforcement learning to the adaptive optimization of classroom evaluation strategies, so as to enhance the real-time, dynamic and availability of evaluation.

2 Literature Review

In recent years, there has been a significant increase in international research on smart classroom, teaching evaluation and educational artificial intelligence. Diaz and Nussbaum (2024) pointed out that after artificial intelligence entered the teaching and learning scene, the research focus should not stop at simple automation support, but should turn to the construction of "educational intelligence" with teaching implications, so as to enhance the real support ability of AI systems for the teaching process [1]. Cukurova (2025) further proposed that the relationship between learning, analysis and artificial intelligence is not linear substitution, but should move towards the framework of "hybrid intelligence", that is to improve the quality of educational decision-making and learning support through human-computer collaboration [2]. Bautista and Lopez-Costa (2025) pointed out in the systematic review of smart learning space that the construction of smart learning environment should integrate teaching, environment and digital dimensions at the same time, which indicates that smart classroom is not just the superposition of technical equipment, but a comprehensive field that can support the perception of teaching behavior, the recognition of learning state and feedback regulation [3]. Futterer et al. (2025) found in a systematic review in the field of classroom management that machine learning and deep learning have been widely used in attendance tracking, behavior monitoring and participation recognition, which provides a technical basis for classroom evaluation to move from empirical judgment to data-driven analysis [4]. In general, the existing research has clarified the development direction of intelligent teaching evaluation in smart classroom environment from the macro level, but there is still a lack of a unified research framework for how to construct a dynamic, continuous and optimized evaluation model in specific classroom

situations.

In terms of classroom process identification and evaluation evidence acquisition, multimodal learning analysis has become an important research path. Yan et al. (2024) proposed an evidence-based multi-modal learning analysis framework to support feedback and reflection in collaborative learning, emphasizing that the interpretability of evaluation and the effectiveness of feedback should be improved through multi-source evidence [5]. D 'Angelo and Rajarathinam (2024) revealed the association between classroom language intervention and collaborative problem solving through the analysis of the intervention voice of teaching assistants in undergraduate engineering education, and provided a phonetic modality basis for the identification of classroom interaction quality [6]. Moon et al. (2024) used multimodal learning analysis as a formative assessment tool to explore the collaborative dynamics in mathematics teacher education, indicating that classroom evaluation can shift from single result judgment to process behavior tracking [7]. Mohammadi et al. (2025) pointed out in their systematic review that artificial intelligence is promoting multimodal learning analysis from data fusion to intelligent inference, which significantly enhances the modeling ability of complex learning processes [8]. Banihashem, Gašević and Noroozi (2025) reviewed learning analysis research from the perspective of formative assessment, and pointed out that its main progress focused on process feedback and learning support, but at the same time, there were also problems such as insufficient feedback granularity, learner subjectivity and practical usability [9]. Echeverria et al. (2025) proposed a learning analysis dashboard for collaborative reflection, indicating that visual analysis results can enhance learners' understanding and depth of reflection on the interaction process [10]. Suraworachet, Zhou, and Cukurova (2025) investigated the usage perception of multimodal AI collaborative analysis systems from the perspective of college students, indicating that multimodal intelligent systems have application potential in real higher education scenarios, but their acceptability is still closely related to feedback transparency and usage experience [11]. This kind of research provides important enlightenment for the dynamic teaching evaluation of smart classroom, that is, the evaluation model needs to be established on the basis of multi-source data fusion and classroom state representation, in order to reflect the classroom teaching process more truly.

In terms of intelligent feedback and automatic teaching evaluation, related research has further promoted teaching evaluation from static conclusions to process intervention. Tang et al. (2025) conducted empirical analysis on generative artificial intelligence assisted teaching, and discussed the applicability and teaching support value of AI as a teaching assistant [12]. Jiang et al. (2025) pointed out that the use target and interaction mode of automatic feedback would affect the writing results by examining the interaction mode between learners and automatic feedback systems, indicating that feedback is not simply output results, but deeply coupled with the learning process [13]. Ben Zion et al. (2025) compared the relationship between AI teaching evaluation and students' perception, and pointed out that automated teaching evaluation has a certain consistency, but it is still difficult to completely replace students' comprehensive judgment on teaching quality [14]. Er et al. (2025) compared the differences between teacher feedback and AI-generated feedback in student perception and use, indicating that AI feedback has advantages in efficiency and accessibility, but still needs to be improved in authority, situational adaptation and acceptance [15]. Bauer et al. (2025) found through field experiments in higher education that adaptive feedback generated by AI can affect statistical learning skills and learning interest, which indicates that intelligent feedback has a certain intervention effect in teaching improvement [16]. Ba et al. (2025) investigated the influence of ChatGPT's assisted feedback on the dynamics and results of online inquiry discussions, and further proved that intelligent feedback can act on the interaction process rather than only on the result level [17]. Weidlich et al. (2025) compared the effect differences of three

feedback sources, teachers, peers and AI, in higher education, indicating that the functional boundaries of different feedback subjects are not yet stable, and the teaching evaluation system still needs to take into account both automaticity and educaticity [18]. These studies show that intelligent feedback and automatic evaluation have become an important technical direction of teaching evaluation, but the existing work is more focused on feedback comparison, user perception or local task performance, and has not yet formed a dynamic evaluation model for the overall operation of the classroom.

In the aspect of reinforcement learning and educational intelligent decision-making, related research provides a methodological basis for the introduction of deep reinforcement learning in this paper. Memarian and Doleck (2024) pointed out in their review of the scope of reinforcement learning in education that reinforcement learning has begun to enter education links such as teaching, learning, evaluation and feedback, and its advantage is that it can gradually optimize strategies in the process of interaction, but it also has the problems of bias control and insufficient deployment in real scenes [19]. Riedmann, Schaper and Lugrin (2025) further summarized the main application scenarios and methodological challenges of reinforcement learning in education in their systematic review, and pointed out that although this direction has the potential to support adaptive educational decision-making, current research is still mainly focused on personalized learning support and task optimization. System modeling for classroom teaching evaluation is still few [20]. Based on the above literature, it can be found that the existing foreign research has formed rich results in the smart classroom environment, multimodal data analysis, intelligent feedback and reinforcement learning education application, but there are still three obvious gaps. First, most research focuses on learning support, feedback optimization or local behavior analysis, and lacks a dynamic evaluation framework for the overall classroom teaching process. Second, although multimodal research improves the ability of classroom state recognition, the mapping relationship between multimodal research and teaching evaluation index system is still not close enough. Third, reinforcement learning research emphasizes strategy optimization, but rarely integrates with classroom multi-source data, teaching evaluation dimensions and real-time feedback mechanism. Therefore, constructing a teaching evaluation model driven by deep reinforcement learning, integrating multimodal classroom data, and dynamically updating for the whole process of classroom around the smart classroom in colleges and universities has clear research value and expansion space.

In order to further compare the differences between the existing research and the proposed research scheme, it is necessary to summarize and analyze the representative related literature from the dimensions of research object, data basis, method path and dynamic optimization ability. Although the existing research has achieved rich results in multi-modal learning analysis, intelligent feedback, automatic evaluation and reinforcement learning education application, most of them still focus on collaborative learning feedback, local behavior recognition, feedback source comparison or method review. There has not been a dynamic teaching evaluation model for the whole process of smart classroom teaching in colleges and universities, which can integrate multi-modal data and realize continuous optimization through deep reinforcement learning. Based on this, this paper selects several representative literatures and compares them with the model to be constructed in this study, as shown in Table 1.

Table 1: Comparison between representative related studies and the model of this study

Representative Literature	Research Object	Data Basis	Methodological Path	Whether Dynamic Optimization Is Emphasized	Limitations Compared with This Study
Yan et al. (2024) [5]	Feedback and reflection in collaborative learning	Multimodal learning data	Multimodal learning analytics	Partially emphasized	Focuses on feedback in collaborative learning rather than taking college classroom teaching evaluation as the core
Moon et al. (2024) [7]	Formative assessment in mathematics teacher education	Multimodal collaborative data	MMLA + formative assessment	Partially emphasized	Pays more attention to dynamic identification in collaboration, but lacks a unified teaching evaluation model
Memarian and Doleck (2024) [19]	Review of reinforcement learning research in education	Literature data	Review of reinforcement learning scope	Emphasizes strategy optimization	Mainly a review and does not develop a specific model for smart classroom teaching evaluation
Riedmann et al. (2025) [20]	Review of reinforcement learning application systems in education	Literature data	Systematic review of reinforcement learning	Emphasizes strategy optimization	Identifies potential and challenges, but does not integrate multimodal classroom data with evaluation outputs
Ben Zion et al. (2025) [14]	Relationship between AI-based teaching evaluation and student perceptions	Evaluation texts and student perception data	AI-based automatic evaluation	Weakly emphasized	Focuses more on consistency comparison of results, while lacking classroom process state modeling
Er et al. (2025) [15]	Comparison between teacher feedback and AI feedback	Learner feedback usage data	Comparative study of feedback	Weakly emphasized	Focuses on differences between feedback providers and does not involve overall classroom teaching evaluation
Weidlich et al. (2025) [18]	Comparison of the effects of teacher, peer, and AI feedback	Higher education feedback data	Comparative study of feedback effects	Weakly emphasized	Focuses on differences in feedback sources and does not construct a dynamic evaluation decision-making mechanism
This study	Dynamic teaching evaluation in smart classrooms in higher education	Multimodal data including classroom video, audio, text, logs, and quizzes	Evaluation indicator system + state-space modeling + deep reinforcement learning	Yes	Covers the whole classroom process and emphasizes the integration of evaluation generation, strategy updating, and feedback-loop closure

3 Design of dynamic teaching evaluation model of smart classroom in Colleges and Universities driven by deep reinforcement learning

3.1 Dynamic teaching evaluation elements and indicators design

The dynamic teaching evaluation model of smart classroom in colleges and universities driven by deep reinforcement learning is built on the basis of perceptive classroom process status, extractable teaching behavior characteristics, computable evaluation indicators, and representable time changes. Dynamic teaching evaluation is not a one-time static judgment for classroom results, but a real-time identification and dynamic characterization of teachers 'teaching behaviors, students 'classroom participation, the quality of teacher-student interaction, the use of teaching resources and classroom feedback effects for the continuous state changes in the process of classroom operation. The design aims to transform the experiential description of classroom teaching activities into computable and updatable digital expressions, and provide structured input for subsequent state space construction and strategy learning.

In the smart classroom environment, classroom video, voice, teaching platform log, interactive text, classroom test record and device call data constitute the data basis of dynamic teaching evaluation. This kind of data has significant heterogeneity in time granularity, structure form and information type. It is necessary to use computer vision, automatic speech recognition, natural language processing, educational data mining and event log analysis to complete feature extraction and unified representation. Based on this, the design of dynamic teaching evaluation elements no longer stays in the empirical dimension division in traditional teaching evaluation, but emphasizes the mappable relationship between teaching activities and data characteristics, so that each type of evaluation elements can correspond to specific data sources, analysis techniques and calculation methods.

Starting from the input requirements of the model, this paper divided the dynamic teaching evaluation elements into five dimensions: teachers 'teaching behavior, students 'classroom participation, the quality of teacher-student interaction, the use of teaching resources and the effect of classroom feedback. Teachers' teaching behaviors were used to characterize the state of classroom organization, content advancement and language control. Student classroom participation was used to describe students' presence, concentration and task response. The quality of teacher-student interaction was used to identify the frequency, depth and timeliness of classroom dialogue. The use of teaching resources is used to describe the matching degree between digital resources retrieval and content presentation. The effect of classroom feedback was used to reflect the real-time test performance, emotional response changes, and learning improvement trends. The five dimensions together constitute the core observation variables of classroom operation state, which not only cover the main components of smart classroom teaching activities, but also have a strong foundation for computational implementation. Figure 1 shows the multi-modal design framework of dynamic teaching evaluation elements and indicators for smart classroom.

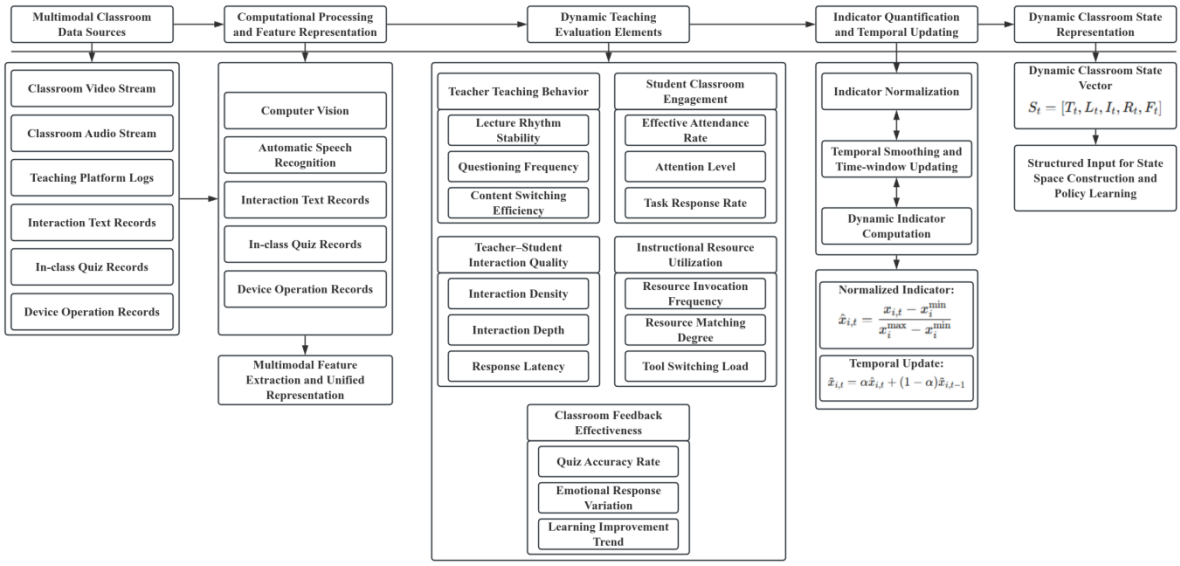

Figure 1: Multimodal design framework of dynamic teaching evaluation elements and indicators in smart classrooms

In order to enhance the cohesion between the index system and the deep reinforcement learning model, the above five types of evaluation factors are further transformed into dynamic state variables on time slice t. Let the state vector of the class at time t be:

$$S_t = [T_t, L_t, I_t, R_t, F_t] \tag{1}$$

Among them, Tt represents the feature set of teachers 'teaching behavior, Lt represents the feature set of students 'classroom participation, It represents the feature set of teacher-student interaction quality, Rt represents the feature set of teaching resources use, and Ft represents the feature set of classroom feedback effect. Any original index xi,t can be expressed as follows after range normalization:

$$\hat{x}_{i,t} = \frac{x_{i,t} - x_i^{min}}{x_i^{max} - x_i^{min}} \tag{2}$$

Considering the continuous evolution characteristics of the classroom state, the time smoothing mechanism is introduced to the standardized index, and the dynamic index representation is obtained:

$$\tilde{x}_{i,t} = \alpha \hat{x}_{i,t} + (1-\alpha)\tilde{x}_{i,t-1} \tag{3}$$

Here, $\alpha \in [0,1]$ is the state update coefficient, which is used to control the weight distribution between the current observation and the historical state. This processing method can weaken the instantaneous noise interference while retaining the real-time change information of the classroom, and improve the stability and robustness of the state representation.

Based on the above design logic, this paper constructs the dynamic teaching evaluation elements and index system, which mainly includes five first-level dimensions: teachers 'teaching behavior, students 'classroom participation, the quality of teacher-student interaction, the use of teaching resources and the effect of classroom feedback. In terms of teachers 'teaching behavior, the stability of teaching rhythm, the frequency of questions and the efficiency of

content switching were selected to reflect teachers 'teaching rhythm, the density of question initiation and the fluency of teaching content conversion respectively. In terms of students 'classroom participation, the class effective rate, attention level and task response rate are selected to describe students 'class attendance, classroom concentration and task participation performance. In terms of the quality of teacher-student interaction, interaction density, interaction depth and response delay are set to reflect the frequency and level of classroom interaction and the timeliness of student response. In the use of teaching resources, resource calling frequency, resource matching degree and tool switching load are set to characterize the impact of digital resource use intensity, resource content adaptation and switching on teaching continuity. In the aspect of classroom feedback effect, the correct rate of in-class test, the change of emotional response and the trend of learning improvement are set to measure the instant knowledge mastery, classroom emotional fluctuation and the change of learning performance before and after. Each index is obtained by multi-modal data such as classroom speech, video, platform log, test record and interactive text, and is calculated by speech recognition, computer vision, text analysis, log mining and time series modeling.

3.2 Multi-source Data Acquisition and state representation in Smart Classroom

The effective construction of dynamic teaching evaluation model in smart classroom relies on the continuous acquisition, unified coding and time state expression of multi-source classroom data. Classroom teaching activity is not a static process driven by a single data stream, but a coupled evolution process of teacher teaching behavior, student participation behavior, teacher-student interaction behavior, resource retrieval behavior and learning feedback behavior in a unified spatio-temporal environment. It is difficult to completely describe the dynamic change characteristics of classroom running status by only relying on questionnaires, grades or single-modal text information. Based on this, this paper constructs a multi-source data acquisition and state representation mechanism for smart classroom scenarios, integrates classroom videos, voice, platform logs, interactive texts, classroom test records and equipment operation records into a unified data processing framework, and uses computer vision, automatic speech recognition, natural language processing, time series modeling and multi-modal fusion technology. The heterogeneous classroom data is mapped into a state vector that can be called by the deep reinforcement learning model.

Suppose the classroom observation period is [0,T], and it is divided into discrete time Windows of length Δ. Then the time interval corresponding to the TTH time slice is $[t\Delta,(t+1)\Delta)$, where t=1,2,...,N, and $N=\lfloor T/\Delta \rfloor$. For any mode $m \in M$, the original data sequence can be expressed as follows:

$$X^{(m)} = \{x_1^{(m)}, x_2^{(m)}, \dots, x_{n_m}^{(m)}\} \tag{4}$$

Here, M={v,a,l,n,q,d} represent six categories of modalities: video, audio, log, text, quiz, and device operation, respectively. After time window segmentation, the mode subsequence corresponding to the TTH time slice is as follows:

$$G_t^{(m)} = \{x_\tau^{(m)} \mid \tau \in [t\Delta, (t+1)\Delta)\} \tag{5}$$

This definition transforms classroom activities from continuous flow data into discrete state observation units, and establishes a time index basis for subsequent feature extraction and state update.

The key to multi-source data acquisition is to establish a uniform timestamp alignment

mechanism. Different modalities have significant differences in sampling frequency and recording granularity. For example, video streams are usually sampled at frame rate fv, speech streams are recorded at sampling rate fa, and log and test data are distributed in an event-triggered manner. In order to achieve cross-modal synchronization, this paper takes the classroom server time as the baseline to resample and map the modal timestamps. Let the observed value recorded by mode m at the original timestamp $t_k^{(m)}$ be $x_k^{(m)}$, then its aggregated value over the uniform window t be expressed as follows:

$$\bar{x}_t^{(m)} = \frac{1}{|G_t^{(m)}|} \sum_{x_k^{(m)} \in G_t^{(m)}} x_k^{(m)} \tag{6}$$

If the corresponding mode produces no observations in window t, the mask variable $b_t^{(m)}$ is introduced as follows:

$$b_t^{(m)} = \begin{cases} 1, & |G_t^{(m)}| > 0 \\ 0, & |G_t^{(m)}| = 0 \end{cases} \tag{7}$$

A hybrid strategy of forward filling and local mean is used to complete the missing data:

$$\hat{x}_t^{(m)} = b_t^{(m)} \bar{x}_t^{(m)} + (1 - b_t^{(m)}) \left(\lambda \hat{x}_{t-1}^{(m)} + (1 - \lambda) \mu^{(m)} \right) \tag{8}$$

where μ(m) is the training set mean of mode m and $\lambda \in [0,1]$ is the history retention coefficient. This processing method can reduce the destruction of state continuity caused by sparse event data.

To eliminate dimensional differences between features at different scales, all observed variables are normalized:

$$z_{i,t}^{(m)} = \frac{\hat{x}_{i,t}^{(m)} - \mu_i^{(m)}}{\sigma_i^{(m)} + \varepsilon} \tag{9}$$

Here, $\mu_i^{(m)}$ and $\sigma_i^{(m)}$ represent the mean and standard deviation of the feature of mode i, respectively, and ε is a smoothing term to prevent the denominator from being zero. The standardized multi-modal features can enter the unified coding stage.

From the perspective of computer technology implementation, different modes adopt differentiated feature extraction paths. Video modality is mainly used to recognize teacher posture, blackboard writing switch, student head orientation, leaving seat behavior and classroom emotion representation, and its encoding process can be expressed as follows.

$$h_t^{(v)} = \phi_v \left(G_t^{(v)} \right) = \mathrm{BiLSTM}(\mathrm{CNN}(F_t^{(v)})) \tag{10}$$

Here, $F_t^{(v)}$ represents the set of video frames in window t, CNN(·) is used to extract spatial visual features, and BiLSTM(·) is used to model inter-frame temporal dependencies. Audio modality mainly represents the teacher's speaking speed, pause structure, intonation change and classroom response intensity, and its coding method is as follows:

$$h_t^{(a)} = \phi_a\left(G_t^{(a)}\right) = \mathrm{GRU}(\mathrm{MFCC}(G_t^{(a)})) \tag{11}$$

where MFCC(·) represents the Mel-frequency cepstral coefficient extraction function. Text modality includes bullet screen in class, discussion record, question text and automatic voice transcribed text, which are used to identify semantic topics, emotional polarity and interaction depth, and can be expressed as follows:

$$h_t^{(n)} = \phi_n\left(G_t^{(n)}\right) = \mathrm{BERT}(\mathrm{Token}(G_t^{(n)})) \tag{12}$$

Log and device operation modes mainly reflect resource call frequency, interface switching event, platform activity and function call path, and their coding results are denoted as:

$$h_t^{(l)} = \phi_l\left(G_t^{(l)}\right), \quad h_t^{(d)} = \phi_d\left(G_t^{(d)}\right) \tag{13}$$

The test mode directly reflects the immediate feedback and knowledge mastery status of the classroom, and its feature vector can be written as follows:

$$h_t^{(q)} = \phi_q\left(G_t^{(q)}\right) = [r_t, c_t, e_t] \tag{14}$$

Here, r_t is the response rate, c_t is the correct rate, and e_t is the error distribution entropy.In order to facilitate the fusion of different modalities in a unified space, this paper uses linear mapping to project the features of each modality into the same dimensional latent space:

$$u_t^{(m)} = W_m h_t^{(m)} + b_m, \quad m \in \mathcal{M} \tag{15}$$

Here, W_m and bm represent the projection matrix and bias term of mode m, respectively. The projected modal representation $\{u_t^{(m)}\}$ generates a unified classroom state embedding through the attention fusion mechanism. The modal attention weights are defined as follows:

$$\alpha_t^{(m)} = \frac{\exp\left(w^\top \tanh(Uu_t^{(m)} + c)\right)}{\sum_{j\in\mathcal{M}} \exp\left(w^\top \tanh(Uu_t^{(j)} + c)\right)} \tag{16}$$

Then the integrated multimodal representation of the classroom is as follows:

$$H_t = \sum_{m\in\mathcal{M}} \alpha_t^{(m)}\, u_t^{(m)} \tag{17}$$

This design enables the model to adaptively adjust the contribution degree of each mode to the state representation according to the information density in different stages of the classroom. For example, the weights of video and audio modes are higher in the stage of teacher teaching, and the weights of test and log modes are relatively enhanced in the stage of classroom practice or voting.

In order to maintain consistency with the dynamic evaluation index system in Section 3.1, this paper does not directly input the fusion representation Ht into the deep reinforcement learning model, but decomposes it into five state components through structured mapping: teacher teaching behavior, student classroom participation, teacher-student interaction quality,

teaching resource use and classroom feedback effect. Let the mapping function be $\psi(\cdot)$, then:

$$[T_t, L_t, I_t, R_t, F_t] = \psi(H_t) \tag{18}$$

Among them, $T_t = W_T H_t + b_T$, $L_t = W_L H_t + b_L$, $I_t = W_I H_t + b_I$, $R_t = W_R H_t + b_R$, $F_t = W_F H_t + b_F$.Further, the complete state vector of the class in time slice t can be obtained as follows:

$$S_t = [T_t, L_t, I_t, R_t, F_t] \tag{19}$$

If we further concatenate the five categories of states into a d-dimensional vector, we can write:

$$S_t \in \mathbb{R}^d, \quad d = d_T + d_L + d_I + d_R + d_F \tag{20}$$

Here, d_T,d_L,d_I,d_R,d_F represent the dimensions of the five types of state subvectors, respectively. Considering the continuity of the classroom state in time, it is difficult to reflect the stage evolution in the teaching process by only using the current window information, so this paper introduces the gated update mechanism to carry out the time recursion of the state:

$$\tilde{S}_t = \gamma_t \odot S_t + (1 - \gamma_t) \odot \tilde{S}_{t-1} \tag{21}$$

Here, $\odot$ denotes the Hadamard product and γt is the time-gated vector:

$$\gamma_t = \sigma(W_g[S_t, \tilde{S}_{t-1}] + b_g) \tag{22}$$

The recursive mechanism not only retains the current classroom change information, but also absorbs the historical state memory, so as to enhance the ability of state representation to describe the evolution of classroom rhythm and local fluctuations. In order to avoid the accumulation of state noise in a long sequence, a regularization term can also be introduced to constrain the change amplitude of adjacent states:

$$\mathcal{L}_{\text{smooth}} = \sum_{t=2}^{N} \| \tilde{S}_t - \tilde{S}_{t-1} \|_2^2 \tag{23}$$

The term helps to keep the smoothness and interpretability of the state sequence.

According to the above modeling process, the technical process of multi-source data acquisition and state representation in smart classroom can be summarized as follows: multi-source classroom data synchronous collection, time window segmentation, modal-level preprocessing and standardization, differentiated feature coding, attention fusion, evaluation dimension mapping, and temporal state update. The whole process makes the heterogeneous classroom data achieve structured expression in a unified latent space, and finally forms the state input suitable for deep reinforcement learning.

In order to enhance the engineering practicability of the process, this paper summarizes the data structures, acquisition methods and coding techniques of the main modalities as shown in Table 2.

Table 2: Multi-source data acquisition and coding design for smart classroom

Data Modality	Raw Data Form	Acquisition Method	Typical Sampling Granularity	Main Features	Encoding Technology
Video Data (v)	Classroom video streams of teachers and students	Cameras, lecture recording devices	15–25 fps	Posture, gaze, leaving-seat behavior, facial expressions, board-writing transitions	CNN, pose estimation, BiLSTM
Audio Data (a)	Classroom speech streams	Microphones, lecture recording systems	16–44.1 kHz	Speech rate, pauses, intonation, loudness, response intensity	MFCC, GRU, ASR
Platform Logs (l)	Clickstreams, dwell time, page transitions	Teaching platforms, learning terminals	Event-triggered	Resource invocation frequency, activity level, operation paths	Log mining, sequence encoding
Interactive Text (n)	Bullet comments, discussions, questions, ASR-transcribed text	Interaction systems, ASR transcription	Sentence-level / turn-level	Semantic topics, interaction depth, emotional polarity	BERT, text classification, sentiment analysis
Quiz Records (q)	Answer results, response time, error distribution	Clickers, online quiz platforms	Item-level / task-level	Accuracy rate, response rate, response latency, error entropy	Educational data mining, statistical modeling
Device Records (d)	Screen switching, screen casting, whiteboard and software operation events	Teaching terminals, classroom IoT devices	Event-triggered	Tool switching load, resource usage paths	Event log analysis, time-series analysis

After the completion of the modal-level coding, the state variables need to be structurally mapped with the aforementioned evaluation elements. Considering that the dynamic teaching evaluation model emphasizes the decision logic of "state-action-reward", the state representation should not only be an abstract hidden vector, but also have both teaching semantics and computational expression. To this end, this paper further refines the state structure of the five categories of evaluation dimensions, as shown in Table 3.

Table 3: Dynamic teaching evaluation state variable structure

State Sub-vector	State Meaning	Main Source Modalities	Representative Variables	Example Dimension
(T_t)	Teacher instructional behavior state	Video, audio, text	Lecture pace, questioning frequency, content-switching efficiency	($d_T = 6$)
(L_t)	Student classroom participation state	Video, logs, quizzes	Effective attendance rate, attention level, task response rate	($d_L = 6$)
(I_t)	Teacher–student interaction quality state	Audio, text, logs	Interaction density, interaction depth, response latency	($d_I = 5$)
(R_t)	Teaching resource usage state	Logs, device records	Resource invocation frequency, resource matching degree, tool-switching load	($d_R = 5$)
(F_t)	Classroom feedback effectiveness state	Quizzes, video, text	Quiz accuracy, emotional response change, learning improvement trend	($d_F = 5$)
(S_t)	Overall dynamic classroom state	Fusion of all modalities	($[T_t, L_t, I_t, R_t, F_t]$)	($d = 27$)

The dimension Settings in Table 3 are only sample configurations in the model implementation stage, and can be adjusted according to the class size, sensing device conditions, and computing resources in practical applications. If the state vector is directly input into the deep reinforcement learning policy network, the state space can be denoted as follows:

$$\mathcal{S} = \{\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_N\} \tag{24}$$

The corresponding state transition can be expressed as follows:

$$P(\tilde{S}_{t+1} \mid \tilde{S}_t, a_t) \tag{25}$$

Here, a_t is the evaluation action or feedback strategy selected by the subsequent evaluation model at time slice t. Therefore, the multi-source data acquisition and state representation not only serve the digital reconstruction of classroom information, but also directly lay the foundation of the environmental state of the deep reinforcement learning model.

In summary, the design of multi-source data acquisition and state representation in smart classroom does not focus on simply expanding data sources, but on transforming original classroom data into a state space with teaching semantic constraints and time continuity characteristics through unified time window, modal level coding, cross-modal fusion and structured mapping. This state space not only retains the fine-grained ability of multimodal data to describe the classroom process, but also meets the modeling requirements of deep reinforcement learning for high-dimensional state input, so as to lay a computable data foundation for action design, reward function construction and strategy learning in the subsequent dynamic teaching evaluation model.

3.3 Construction of Dynamic Teaching Evaluation Model for Deep Reinforcement Learning

After the multi-source data acquisition and state representation of smart classroom are completed, the dynamic teaching evaluation problem can be further formulated as a sequential decision-making problem that continues to evolve with the classroom process. Because the classroom operation state is not a static snapshot of a single moment, but the coupling result of teachers 'teaching behavior, students 'participation state, teacher-student interaction quality, resource invocation method and feedback effect in continuous time, the traditional evaluation model based on fixed weight or single classification is difficult to effectively describe the dynamic fluctuation characteristics of classroom quality. Based on this, this paper modeled the dynamic teaching evaluation of university smart classroom as a finite-time domain Markov decision process driven by deep reinforcement learning. Let the evaluation environment be:

$$\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle \tag{26}$$

Here, S is the classroom state space, A is the evaluation action space, P is the state transition probability, R is the reward function, and $\gamma \in (0,1)$ is the discount factor. For any time slice t, the comprehensive state of the classroom obtained in Section 3.2 is denoted as $\tilde{S}_t \in R_d$, then the model needs to generate an evaluation action $a_t \in A$ according to the current state, and obtain an immediate reward rt through environmental feedback at the next moment, thus forming a transition chain of state-action-reward-new state:

$$(\tilde{S}_t, a_t, r_t, \tilde{S}_{t+1}) \tag{27}$$

Considering the significant time dependence of classroom evaluation, it is difficult to fully reflect the evolution law of teaching rhythm, interaction density and feedback response by only using a single moment state. Therefore, this paper introduces a sliding state window with length h to construct a time series input tensor:

$$X_t = [\tilde{S}_{t-h+1}, \tilde{S}_{t-h+2}, \dots, \tilde{S}_t] \in \mathbb{R}^{h \times d} \tag{28}$$

In terms of network structure design, the model adopted a dual-branch architecture of "temporal state encoder-policy network-value network". The state encoder is responsible for extracting the high-order dynamic representation of the classroom process from the sliding window, the policy network is responsible for generating the evaluation action probability distribution, and the value network is responsible for estimating the long-term payoff of the current state. In order to enhance the ability to depict the classroom evolution trend, this paper combines the gated recurrent unit with the attention mechanism to encode the input sequence. Let the encoder output be:

$$z_t = \mathrm{Attn}(\mathrm{GRU}(X_t)) \tag{29}$$

Among them, $\mathrm{GRU}(\cdot)$ is used to capture the temporal dependence of the state sequence, and $\mathrm{Attn}(\cdot)$ is used to highlight the critical time slices that are more sensitive to the evaluation decision. If the hidden state of GRU at time k is denoted as hk, the attention weight can be expressed as follows:

$$\beta_k = \frac{\exp(q^\top \tanh(W_h h_k + b_h))}{\sum_{j=t-h+1}^{t} \exp(q^\top \tanh(W_h h_j + b_h))} \tag{30}$$

As a result, the time-series aggregate representation is as follows:

$$z_t = \sum_{k=t-h+1}^{t} \beta_k h_k \tag{31}$$

This representation not only contains the local fluctuation information of the classroom state, but also retains the stage trend characteristics, which can be used as a compact input for subsequent strategy generation.

In order to avoid the abstract semantics of evaluation action, this paper defines the action space as a discrete combination form of "evaluation level-warning mark-feedback strength". Let the set of evaluation levels be:

$$\mathcal{G} = \{1,2,3,4,5\} \tag{32}$$

Corresponding to "poor, low, medium, good, excellent" five levels of teaching status; The set of warning markers is as follows:

$$\mathcal{W} = \{0,1\} \tag{33}$$

where 0 means no warning is triggered and 1 means warning is triggered. The feedback strength set is as follows:

$$\mathcal{F} = \{0,1,2\} \tag{34}$$

where 0 represents weak feedback, 1 represents moderate feedback, and 2 represents strong feedback. The single-step action can then be written as follows:

$$a_t = (g_t, w_t, f_t), \quad g_t \in \mathcal{G}, w_t \in \mathcal{W}, f_t \in \mathcal{F} \tag{35}$$

The action space scale is thus obtained as follows:

$$|\mathcal{A}| = |\mathcal{G}| \times |\mathcal{W}| \times |\mathcal{F}| = 5 \times 2 \times 3 = 30 \tag{36}$$

The discrete action design makes the classroom evaluation result no longer a single score output, but also has the functions of level judgment, risk identification and feedback regulation. The policy network generates the action distribution according to the coded representation zt:

$$\pi_\theta(a_t|\tilde{S}_t) = \mathrm{Softmax}(W_p z_t + b_p) \tag{37}$$

Here, θ is the policy network parameter. The corresponding scalar output of classroom assessment can be further mapped as follows:

$$\hat{y}_t = \omega_1 g_t + \omega_2 w_t + \omega_3 f_t \tag{38}$$

Here,ω_1,ω_2,ω_3 are action semantic mapping coefficients, which are used to unify the dimensions of different action components. In the value estimation branch, the model estimates the expected cumulative return of the current state through the value network:

$$V_\phi(\tilde{S}_t) = W_v z_t + b_v \tag{39}$$

Here, ϕ is the value network parameter. The role of value function is not simply to fit the immediate evaluation results, but to describe the long-term benefits of a certain evaluation action in the subsequent evolution of classroom states, which is particularly important for dynamic teaching evaluation, because some short-term conservative evaluation strategies may bring more stable classroom feedback and higher consistency of teaching quality in the subsequent periods.

The construction of reward function directly determines the optimization direction of the model. If only the classification accuracy or scoring error is taken as the target, the model is easy to degenerate into a static predictor, which cannot reflect the essence of dynamic evaluation and feedback regulation. Therefore, this paper constructs a composite reward function from four aspects: evaluation consistency, feedback gain, response delay and state smoothness:

$$r_t = \lambda_1 R_t^{acc} + \lambda_2 R_t^{imp} - \lambda_3 R_t^{delay} - \lambda_4 R_t^{osc} \tag{40}$$

Here,$\lambda_1,\lambda_2,\lambda_3,\lambda_4 \geq 0$ are the reward weights. The evaluation consistency term is defined as follows:

$$R_t^{acc} = 1 - \frac{|\hat{y}_t - y_t^*|}{Y_{max}} \tag{41}$$

where y_t^* represents the reference label given by expert annotation, teacher mutual evaluation or benchmark evaluation system, and Ymax is the upper bound of evaluation scale. The feedback gain term is used to characterize the positive improvement effect of model output on subsequent classroom status:

$$R_t^{imp} = \eta_1(L_{t+1} - L_t) + \eta_2(I_{t+1} - I_t) + \eta_3(F_{t+1} - F_t) \tag{42}$$

Here, L_t,I_t,F_t represent student participation status, interaction quality status, and feedback effect status, respectively. This term describes the promotion effect of evaluation action on classroom participation improvement, interaction improvement and feedback optimization. The response delay term penalizes the evaluation output lag:

$$R_t^{delay} = \frac{\tau_t}{\tau_{max}} \tag{43}$$

Here, τ_t is the actual delay from the completion of state collection to the generation of evaluation results. The fluctuation penalty term is used to suppress the abnormal oscillation of evaluation results in adjacent time slices:

$$R_t^{osc} = \|\hat{y}_t - \hat{y}_{t-1}\|_2^2 \tag{44}$$

Through the above design, the optimization goal of the model is extended from simply pursuing local prediction accuracy to simultaneously taking into account evaluation timeliness, intervention effectiveness and sequence stability.

In the model training stage, the proximal policy optimization method based on the Actor-Critic framework is used. Let the old policy be $\pi\theta$old:, then the probability ratio is defined on the sample trajectory as follows:

$$\rho_t(\theta) = \frac{\pi_\theta(a_t|\tilde{S}_t)}{\pi_{\theta_{old}}(a_t|\tilde{S}_t)} \tag{45}$$

To reduce the risk of oscillation in the policy update process, a truncated objective function is used:

$$L^{clip}(\theta) = \mathbb{E}_t[\min(\rho_t(\theta)A_t, clip(\rho_t(\theta),1-\epsilon,1+\epsilon)A_t)] \tag{46}$$

Here, ϵ is the truncation threshold and At is the dominance function. The dominance function is obtained by generalized dominance estimation as follows:

$$A_t = \sum_{l=0}^{L-1} (\gamma\kappa)^l \delta_{t+l}, \quad \delta_t = r_t + \gamma V_\phi(\tilde{S}_{t+1}) - V_\phi(\tilde{S}_t) \tag{47}$$

Here, κ is the dominant attenuation factor. The value network loss is defined as follows:

$$L^V(\phi) = \mathbb{E}_t\left[\left(V_\phi(\tilde{S}_t) - \hat{R}_t\right)^2\right] \tag{48}$$

$\hat{R}_t$ is the cumulative discounted return from time t. In order to enhance the exploration ability and avoid premature convergence of the policy, the entropy regularization term is also introduced:

$$L^H(\theta) = \mathbb{E}_t\left[-\sum_{a\in\mathcal{A}} \pi_\theta(a|\tilde{S}_t)\log\pi_\theta(a|\tilde{S}_t)\right] \tag{49}$$

Therefore, the total model loss can be written as follows:

$$L(\theta,\phi) = L^{clip}(\theta) - c_1 L^V(\phi) + c_2 L^H(\theta) \tag{50}$$

where c_1,c_2 are the balance coefficients. Through this joint goal, the model can improve the adaptability of evaluation strategies to complex classroom situations while maintaining the stability of strategy update.

In order to clearly show the structural relationship within the model from state encoding to policy generation, and then to value estimation and reward return, the network framework shown in Figure 2 is constructed. The diagram should focus on the coupling relationship between the multimodal state input, the timing encoder, the policy output branch, the value evaluation branch, and the reward feedback loop.

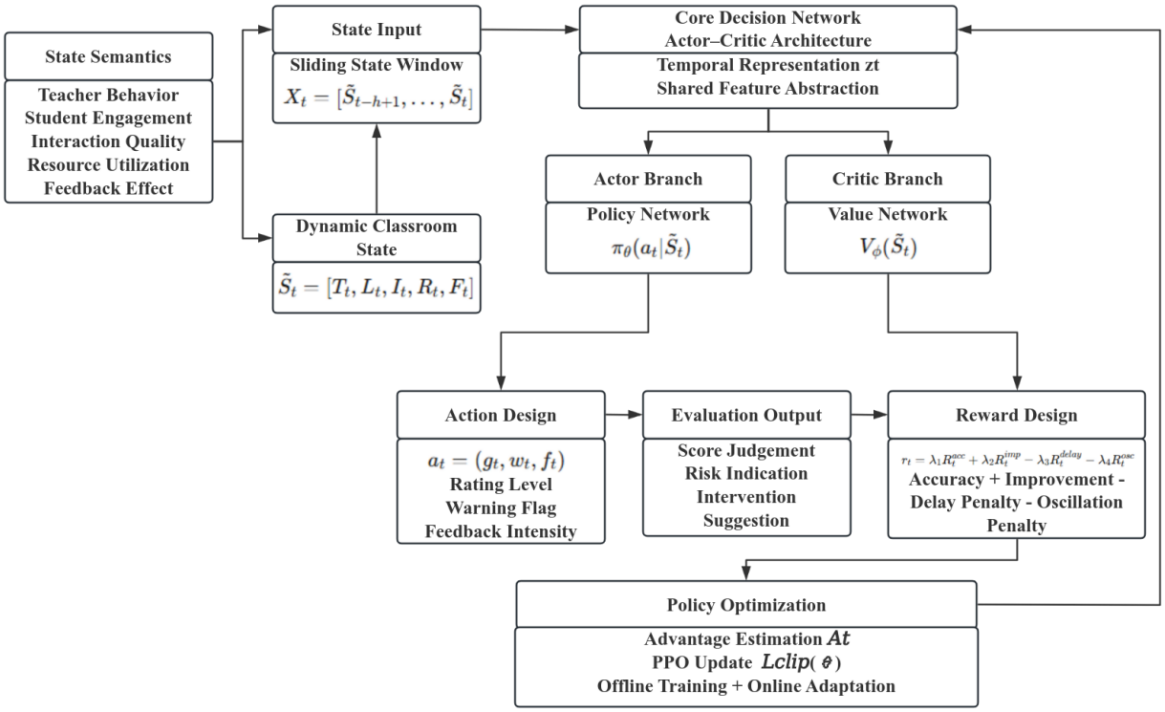

Figure 2: Deep reinforcement learning architecture for dynamic teaching evaluation in smart classrooms

At the engineering implementation level, the model operation is divided into two stages: offline training and online inference. In the offline stage, a trajectory sample set is constructed based on historical classroom data:

$$\mathcal{D} = \{(\tilde{S}_t, a_t, r_t, \tilde{S}_{t+1})\}_{t=1}^{N} \tag{51}$$

And the parameters are updated by mini-batch sampling. If the batch size is B, the learning rate is α, and the number of single round updates is E, the policy parameter iteration can be written as follows:

$$\theta \leftarrow \theta + \alpha \nabla_\theta L^{clip}(\theta), \quad \phi \leftarrow \phi - \alpha \nabla_\phi L^{V}(\phi) \tag{52}$$

In the online phase, the system receives the multi-source observation data of the smart classroom in real time, outputs the evaluation results at the current time after state coding and strategy reasoning, and feeds back the observation and reward at the next time back to the training buffer, realizing the closed loop of "observation-evaluation-feedback-update". This mechanism ensures that the model can not only use historical data to form a stable initial strategy, but also adapt gradually according to new observations in the real classroom environment. Figure 3 shows the flow of model training and online inference.

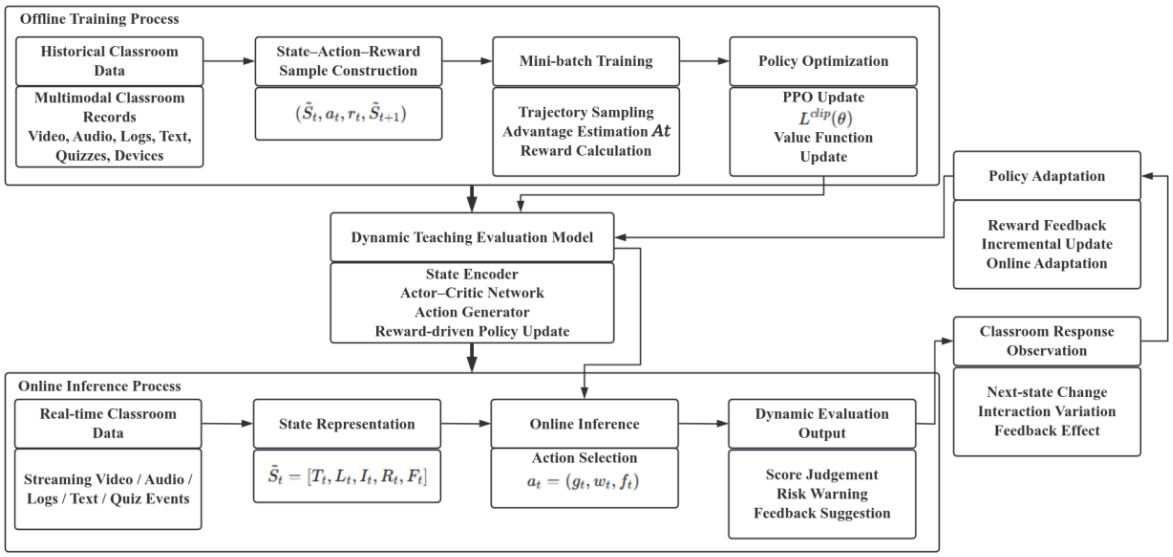

Figure 3: Training and online inference process of the dynamic teaching evaluation model

In conclusion, the dynamic teaching evaluation model of deep reinforcement learning is not a simple classification of classroom states, but a computational framework that can continuously perceive classroom evolution, dynamically output evaluation results, and continuously adjust strategies according to feedback is constructed through multimodal state representation, temporal coding, discrete action decision-making, composite reward modeling, and strategy-value collaborative optimization. The model not only maintains the semantic interpretability of teaching evaluation, but also has the adaptive optimization ability for complex smart classroom scenarios, which provides a complete technical implementation basis for subsequent experimental verification and model performance analysis.

3.4 Model operation process and evaluation output mechanism

The operation of the dynamic teaching evaluation model of deep reinforcement learning consists of four steps of "state access, strategy reasoning, result generation, feedback update", and its essence is to realize the collaborative promotion of evaluation decision-making and strategy optimization in continuous classroom situations. When the model runs, the multi-source data such as video, voice, log, text and test in the smart classroom first enter the state building module, and form the classroom state S~t at the current moment after time window segmentation, multi-modal feature extraction and state mapping. This state is not a single rating variable, but a comprehensive representation of teachers 'teaching behavior, students 'classroom participation, the quality of teacher-student interaction, resource usage and feedback effect, so it can provide structured input for subsequent evaluation decisions.

In the policy inference stage, the model reads the state sequence of the latest period of time, generates the corresponding time series representation, and inputs it into the Actor-Critic decision network. The Actor branch is responsible for selecting the current best evaluation action in the discrete action space, and the Critic branch synchronously estimates the long-term value of the action. After the action output, the system maps it into interpretable evaluation results, including teaching status grade, risk warning mark and feedback suggestion strength. In order to ensure the application readability of the results, the evaluation output does not stay at the probability layer of the network directly, but is converted into semantic results for classroom scenarios, such as normal classroom operation, insufficient interaction, slow feedback or frequent resource switching. Figure 4 shows the overall operation process of the

model.

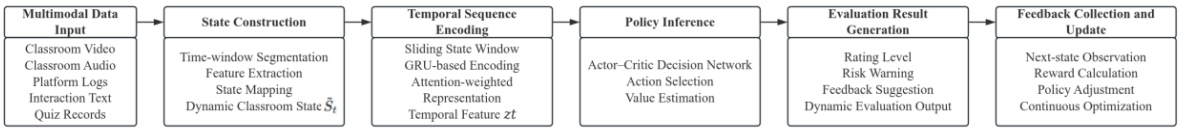

Figure 4: Operational flow of the dynamic teaching evaluation model

After the evaluation output, the system continued to monitor the changes of classroom status in the next time slice, and generated reward signals based on information such as increased participation, improved interaction, feedback response, and evaluation stability. On the one hand, the reward is used to measure the effectiveness of the current evaluation action, and on the other hand, the reward is transmitted back to the network parameters through the policy update module to realize the continuous adaptive adjustment of the model. Therefore, the model forms a closed-loop operation mechanism of "state perception-action generation-evaluation output-feedback correction". The final output includes three kinds of results: dynamic evaluation grade curve, risk early warning results and feedback suggestion set. The former is used to describe the time evolution of classroom quality, and the latter two are used to support teachers 'regulation and teaching improvement. Figure 5 shows the evaluation result generation and feedback update mechanism.

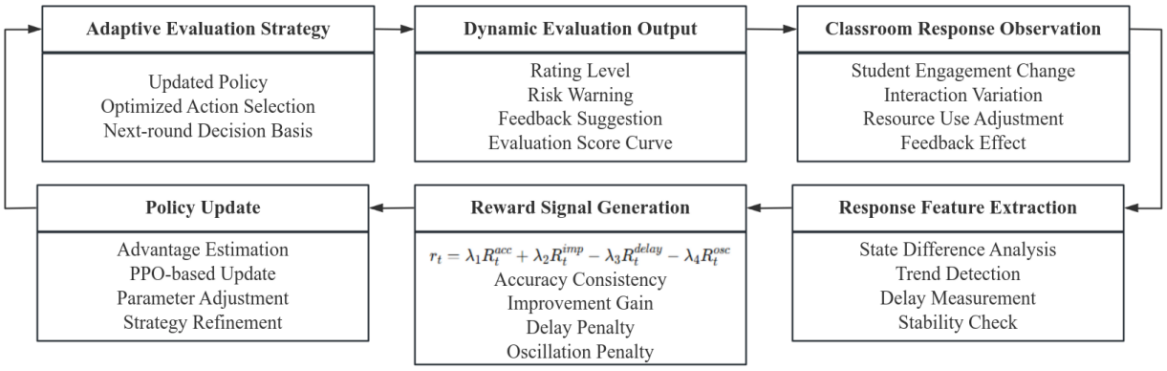

Figure 5: Evaluation output and feedback update mechanism of the model

4 Experiment design and result analysis

4.1 Experimental environment and data sources

The experiment was completed in the mixed environment of Windows 11 and Ubuntu 22.04, the development language was Python 3.10, the deep learning framework used PyTorch 2.1, combined with CUDA 12.1 to realize model training and inference. The main hardware configuration is Intel Xeon Silver 4310 processor, 64 GB memory, and NVIDIA RTX 4090 24 GB graphics card. The sample courses selected 8 smart classroom courses in a university, covering three types of course forms such as computer foundation, program design and data analysis. A total of 96 classroom teaching records were collected, and the duration of a single class was 45-90 min, and the total collection period was 16 weeks. The original data included 428 hours of classroom video, 428 hours of audio, 316,000 platform logs, 28,400 interactive texts, 11,700 classroom test records, and 69,200 equipment operation events. In the process of

dataset construction, multi-modal alignment and state segmentation were completed in a time window of 30 s, forming a total of 41280 state samples, which were divided into training set, validation set and test set according to 7:1:2. In order to reduce the noise interference, the video, voice, log and text data were processed by missing completion, anomaly removal, standardization and time synchronization respectively, and finally the multi-source time series data set for dynamic teaching evaluation was constructed.

4.2 Compare models and evaluation metrics

In order to verify the effectiveness of the deep reinforcement learning dynamic teaching evaluation model constructed in this paper, the traditional evaluation method is set up as a comparison with other intelligent models. The traditional methods select analytic Hierarchy process and weighted comprehensive evaluation method, which mainly complete the classroom evaluation based on the preset index weight, and can be used to test the application boundary of the static rule-driven method in the dynamic scene of smart classroom. In terms of intelligent models, BP neural network, support vector machine, LSTM and Transformer are selected as baseline models. Among them, BP neural network is used to characterize the general nonlinear mapping ability, support vector machine is used to characterize the classification performance under small and medium sample conditions, LSTM is used to reflect the time series modeling ability, and Transformer is used to test the effect of attention mechanism in classroom multimodal state recognition. At the same time, in order to highlight the role of "reinforcement learning driven" in the proposed method, a deep feedforward network without policy update mechanism is also set as an ablation control to compare the differences between dynamic decisions and static predictions.

The evaluation index is constructed from three levels: classification performance, error control and operation efficiency. The classification performance index includes accuracy rate, recall rate and F1 value. The accuracy rate is used to measure the accuracy of the model's overall judgment of the classroom teaching state, the recall rate is used to reflect the recognition ability of the model for key states such as low participation, insufficient interaction or abnormal feedback, and the F1 value is used to comprehensively evaluate the balance between accuracy and completeness of the model. The mean absolute error and root mean square error are selected as error indicators to measure the degree of deviation between the model output evaluation results and the expert annotation results. The mean absolute error reflects the overall deviation level, and the root mean square error is more sensitive to large errors, which is more suitable for testing the stability of the model. In terms of operational efficiency, the average response time delay of a single time window and the processing throughput per unit classroom are selected as evaluation indicators to measure the response speed and continuous processing ability of the model in online scenarios, respectively. Through the above comparison model and evaluation index design, the comprehensive effect of the model in this paper can be systematically tested from four dimensions of evaluation accuracy, dynamic recognition ability, error control level and real-time operation performance.

4.3 Model training and parameter setting

In order to ensure the convergence stability and online response ability of the dynamic teaching evaluation model of deep reinforcement learning in the complex scene of smart classroom, this paper uses the strategy optimization method based on PPO to complete the model training. In the training phase, the multi-source temporal state samples constructed in Section 3.2 are taken as the input, the evaluation action and reward function defined in Section 3.3 are taken as the optimization objective, and the parameter learning is completed under the mechanism of combination of offline trajectory sampling and online policy update. Considering that the

classroom state sequence has the characteristics of short-term fluctuation and stage evolution at the same time, the state window length is set to 8, and the batch size is set to 64, so as to balance the training efficiency and the time dependence modeling ability. The training rounds were set to 200 epochs, which could cover the whole process of the model from initial exploration to stable convergence in the later stage.

In terms of optimization parameter Settings, the learning rate of policy network was set to 3×10^{-4}, the learning rate of value network was set to 1×10^{-3}, and the discount factor γ was set to 0.95. The attenuation coefficient of dominance estimation is set to 0.95, the PPO truncation coefficient is set to 0.2, and the entropy regularization term coefficient is set to 0.01. The capacity of the experience pool is set to 4096 state transition samples to ensure that enough classroom trajectory information can be retained during the training process, and to avoid the interference of redundant samples caused by too large cache size. In order to suppress gradient oscillation and improve training stability, the gradient clipping threshold is set to 0.5, the weight attenuation coefficient is set to 1×10^{-5}, and the parameter update frequency is set to be performed every 4 epochs. The above key parameter Settings are shown in Table 4. On the whole, this set of parameter configurations takes into account the convergence speed of the strategy, the stability of the value estimation, and the actual needs of the classroom time series data modeling.

Table 4: Key parameter Settings for model training

Parameter Name	Symbol / Description	Value
Number of Training Epochs	Epoch	200
Batch Size	Batch size	64
State Window Length	Sequence length	8
Actor Learning Rate	Actor learning rate	(3×10^{-4})
Critic Learning Rate	Critic learning rate	(1×10^{-3})
Discount Factor	(\gamma)	0.95
Advantage Estimation Decay Factor	GAE factor	0.95
PPO Clipping Coefficient	(\epsilon)	0.2
Entropy Regularization Coefficient	Entropy coefficient	0.01
Replay Buffer Size	Replay buffer size	4096
Gradient Clipping Threshold	Gradient clipping	0.5
Weight Decay Coefficient	Weight decay	(1×10^{-5})
Parameter Update Frequency	Update interval	Every 4 epochs

Under the above parameter conditions, the model training process shows good convergence characteristics. As shown in Figure 6, Actor Loss decreases rapidly in the early stage of training, gradually decreases from the initial high level and stabilizes in the lower interval in the later stage, indicating that the policy network can continuously improve the evaluation action selection. The Critic Loss also showed a continuous downward trend and leveled off in the middle and late stages, indicating that the estimation of long-term returns by the value network was gradually stable. At the same time, the standardized reward value continued to rise with the increase of training rounds and remained in a high interval after about 150 epochs, indicating that the evaluation strategy learned by the model had good effectiveness in classroom state recognition and feedback optimization. It can be seen that the parameters listed in Table 4 can support the model to complete stable training and provide a reliable parameter basis for the analysis of subsequent experimental results.

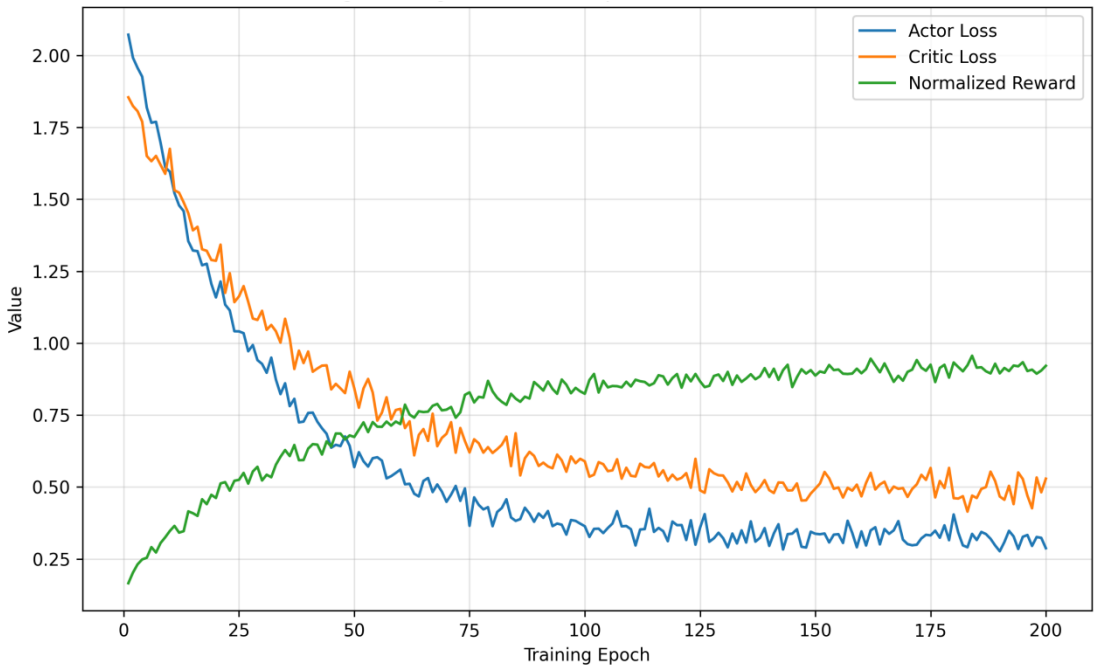

Figure 6: Training convergence curve of the proposed dynamic teaching evaluation model

4.4 Analysis of experimental results

In order to test the comprehensive performance of the proposed model in the dynamic teaching evaluation task of smart classroom, this paper compares the proposed deep reinforcement learning model with the baseline methods such as analytic hierarchy process, weighted comprehensive evaluation method, BP neural network, support vector machine, LSTM and Transformer, and the results are shown in Table 5.

Table 5: Comparison of experimental results of different models in dynamic teaching evaluation task

Model	Accuracy	Recall	F1-score	MAE	RMSE	Response Delay (s/window)	Stability Index	Dynamic Adaptability Index
AHP	0.742	0.701	0.718	0.214	0.287	0.41	0.71	0.66
Weighted	0.758	0.716	0.733	0.203	0.275	0.38	0.73	0.68
BP	0.811	0.784	0.796	0.161	0.228	0.23	0.79	0.74
SVM	0.826	0.798	0.811	0.149	0.214	0.19	0.81	0.76
LSTM	0.872	0.851	0.860	0.108	0.169	0.16	0.87	0.83
Transformer	0.891	0.868	0.878	0.096	0.154	0.15	0.89	0.86
Proposed	0.923	0.909	0.915	0.071	0.118	0.11	0.94	0.92

On the whole, the performance of traditional rule-based methods in dynamic classroom scenarios is relatively weak, and the accuracy of AHP and weighted comprehensive evaluation method are only 0.742 and 0.758, respectively, indicating that it is difficult to effectively describe the temporal changes and state transitions in the classroom process by only relying on fixed weights and static indicators. BP and SVM have improved accuracy and error control compared with traditional methods, but their performance is still significantly lower than that of time series models due to the lack of modeling ability for long time series dependence and continuous feedback mechanism. LSTM and Transformer have achieved good results in

dynamic evaluation tasks relying on sequence modeling ability, but the model in this paper still performs best on each core index, indicating that integrating temporal state representation, evaluation action generation and reward-driven strategy update into a unified framework can further improve the effectiveness of dynamic classroom teaching evaluation.

In terms of evaluation accuracy, the accuracy, recall and F1 value of the proposed model reach 0.923, 0.909 and 0.915, respectively, which are higher than all baseline models. Compared with the state-of-the-art static deep model Transformer, the proposed model improves the accuracy, recall rate and F1 value by 3.2, 4.1 and 3.7%, respectively. Compared with LSTM, the improvement reaches 5.1, 5.8 and 5.5%, respectively. Figure 7 further gives the visual comparison results of classification performance. It can be seen that as the model gradually transitions from the traditional method to the time series model and the reinforcement learning driven model, the three indicators show a stable upward trend as a whole, while the proposed model maintains the highest level in the three dimensions. This indicates that the evaluation framework based on deep reinforcement learning can not only identify classroom states more accurately, but also improve the recognition ability of key scenarios such as low participation, insufficient interaction, and abnormal feedback.

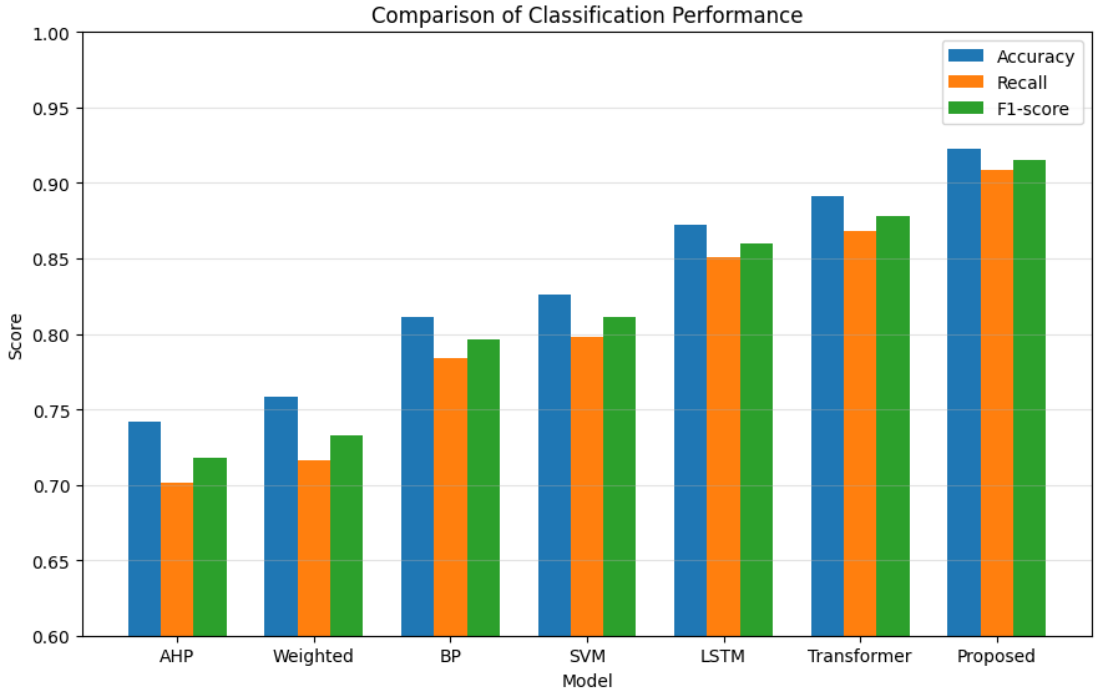

Figure 7: Comparison of Classification Performance

From the perspective of error indicators, the MAE and RMSE of the proposed model are reduced to 0.071 and 0.118, respectively, which are significantly better than those of the baseline models. Compared with Transformer, MAE is decreased by about 26.0% and RMSE is decreased by about 23.4%. Compared with LSTM, MAE and RMSE decrease by about 34.3% and 30.2%, respectively. Figure 8 shows that the proposed model has stronger advantages in error control, indicating a higher consistency between its output results and expert annotations or reference evaluations. This result shows that the deep reinforcement learning model can not only improve the accuracy of classification judgment, but also reduce the overall deviation of the evaluation results when conducting dynamic teaching evaluation, thus enhancing the interpretability and credibility of the evaluation results.

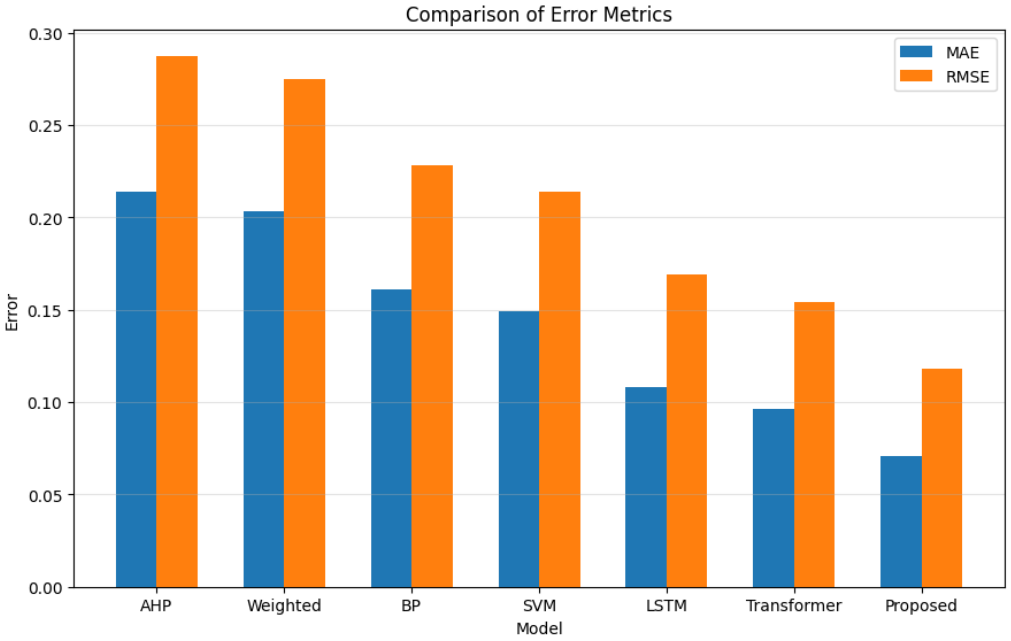

Figure 8: Comparison of Error Metrics

In terms of timeliness, the average response delay of the proposed model in a single time window is 0.11 s, which is significantly lower than that of the traditional method and other intelligent models. Figure 9 shows that although AHP and weighting methods are simple to implement, the overall response efficiency is not dominant due to the combination of dependent rule calculation and post-processing. LSTM and Transformer have significantly improved the accuracy, but the proposed model still compresses the average response delay by about 31.3% and 26.7%, respectively. This result shows that the proposed model does not significantly increase the cost of online reasoning while maintaining high accuracy, and can better meet the real-time requirements of dynamic evaluation in smart classroom.

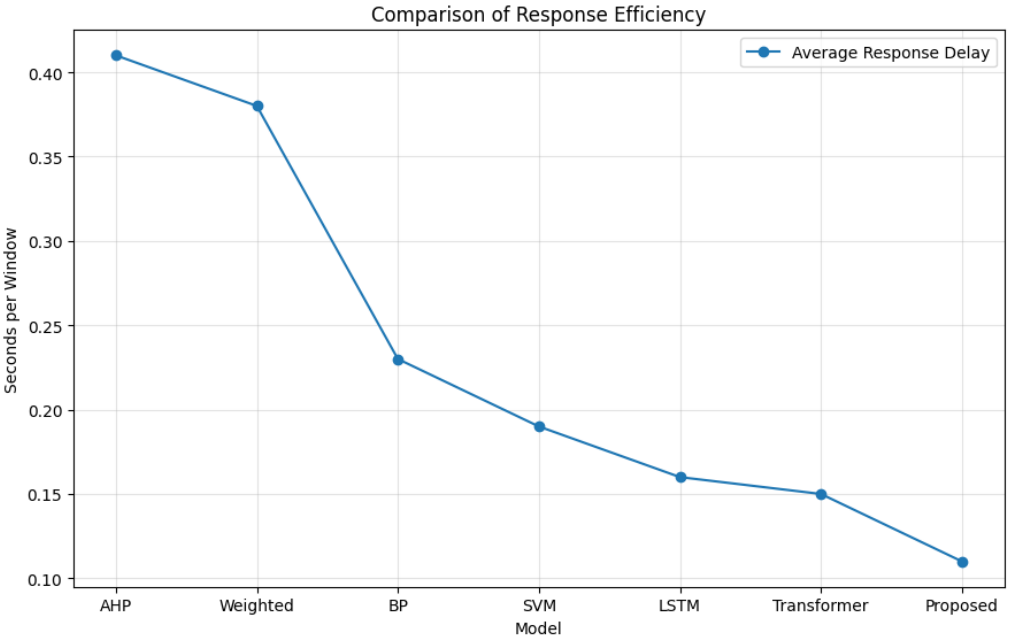

Figure 9: Comparison of Response Efficiency

In terms of stability and dynamic adaptive ability, the stability index and dynamic adaptive ability index of the proposed model reach 0.94 and 0.92, respectively, which are higher than other baseline models. Figure 10 shows that the output results of traditional methods are more prone to discontinuity and adaptation lag in the face of classroom rhythm fluctuations, interaction frequency changes, and feedback state disturbances. By relying on the reward feedback and strategy update mechanism, the model in this paper can still maintain high stability under the condition of state change, and make more timely adjustments to the dynamic changes of the classroom. Compared with Transformer, the stability index of the proposed model is increased by 0.05, and the dynamic adaptive ability index is increased by 0.06. Compared with LSTM, the improvement reaches 0.07 and 0.09, respectively. It shows that the advantages of the model in this paper are not only reflected in the static accuracy level, but also reflected in the adaptive evaluation ability of the continuously changing classroom situation.

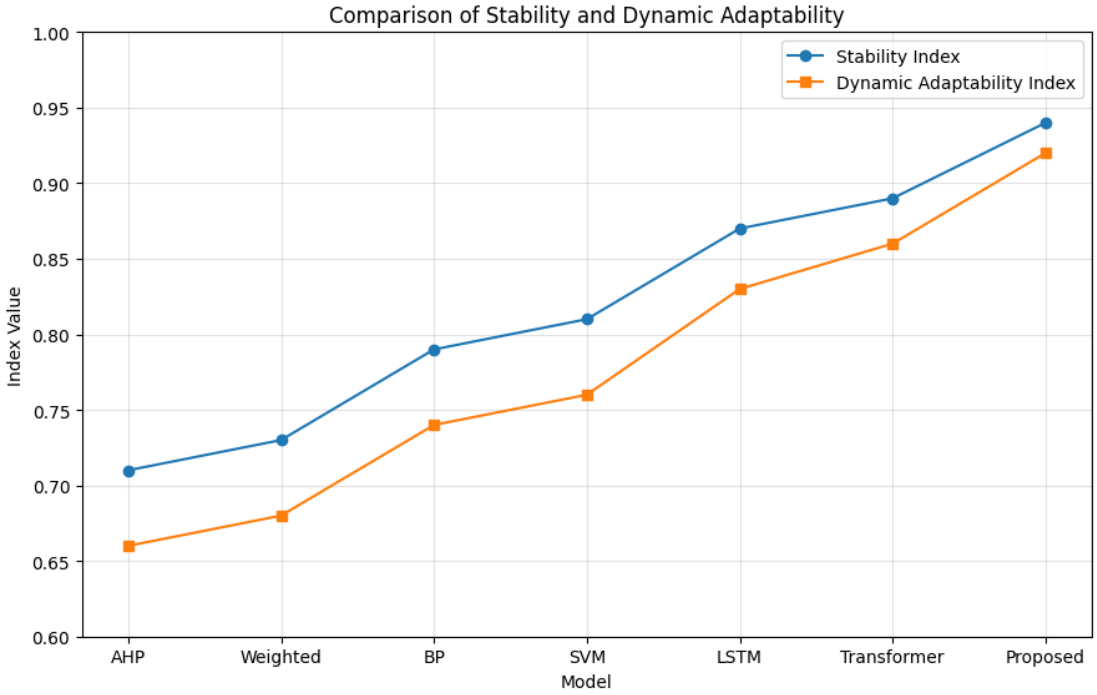

Figure 10: Comparison of Stability and Dynamic Adaptability

It can be seen from Table 5 and Figure 7-Figure 10 that the proposed model performs best in four aspects: evaluation accuracy, timeliness, stability and dynamic adaptability. This result verifies the effectiveness of the above model design, that is, through multi-source state representation, actor-critic decision structure and reward-driven strategy update, the smart classroom teaching evaluation can be promoted from static result judgment to dynamic, continuous and optimized evaluation oriented to the whole process of classroom.

4.5 Ablation experiments and discussion

In order to further test the contribution of the three key modules of multimodal fusion, state modeling and reward optimization to the model performance, this paper sets up three groups of ablation experiments based on the complete model, and the results are shown in Table 6. After removing the multi-modal fusion, the model only depended on the single channel state input, the accuracy was reduced from 0.923 to 0.881, and the F1 value was reduced to 0.869, indicating that the joint representation of classroom video, voice, log and text information could significantly enhance the integrity of evaluation evidence. After removing state modeling, the

model no longer used sliding window and temporal coding, and only based on single moment state for reasoning, its accuracy was further reduced to 0.862, and the stability index was reduced to 0.84, indicating that classroom teaching evaluation had obvious temporal dependence, and the lack of state evolution modeling would directly weaken the model's ability to recognize classroom rhythm changes and interactive fluctuations. After removing the reward optimization, the model degenerates into a static depth prediction framework. Although the basic accuracy is still higher than that of the traditional method, the dynamic adaptability decreases to 0.85, indicating that the reward feedback mechanism plays a key role in continuously adjusting the evaluation strategy and improving the adaptability of the classroom situation.

Table 6: Comparison of the results of ablation experiments

Model Variant	Accuracy	F1-score	MAE	Stability Index	Dynamic Adaptability Index
Full model	0.923	0.915	0.071	0.94	0.92
Without multimodal fusion	0.881	0.869	0.103	0.88	0.86
Without state modeling	0.862	0.851	0.118	0.84	0.82
Without reward optimization	0.889	0.874	0.096	0.89	0.85

From the perspective of applicability, the proposed model is suitable for deployment in a smart classroom environment with classroom sensing devices, teaching platform logs and interaction records, especially in scenarios that need to continuously monitor classroom participation, interaction quality and feedback effect. However, its limitations are also obvious. On the one hand, the quality of multimodal data is greatly affected by the acquisition equipment and classroom organization. On the other hand, the reward function still relies on the reference evaluation and state change construction, which may need to be recalibed when transferring across courses and disciplines. Therefore, future research still needs to be further optimized around lightweight deployment, cross-scenario generalization and reward design robustness.

5 Conclusion

Focusing on the problem of dynamic teaching evaluation of smart classroom in colleges and universities, this paper constructs a dynamic teaching evaluation model driven by multi-source data and supported by deep reinforcement learning, and verifies its effectiveness in experiments. The research showed that the evaluation framework based on five elements of teachers 'teaching behavior, students 'classroom participation, teacher-student interaction quality, resource usage and feedback effect could completely cover the core states of the smart classroom teaching process. Through the unified representation of multi-modal data such as video, voice, log, text and test, the heterogeneous information in the classroom process can be transformed into computable and updatable dynamic state input. On this basis, the Actor-Critic structure and PPO strategy optimization mechanism were introduced to realize the transformation of classroom evaluation from static judgment to dynamic decision-making. The experimental results show that the accuracy, recall and F1 value of the proposed model in the dynamic teaching evaluation task reach 0.923, 0.909 and 0.915 respectively, the mean absolute error and root mean square error are reduced to 0.071 and 0.118 respectively, and the average response delay of a single time window is 0.11 s. The stability index and dynamic adaptive ability index reach 0.94 and 0.92, respectively, which are better than the baseline models such as AHP, weighted comprehensive evaluation method, BP, SVM, LSTM and Transformer. Ablation

experiments further show that multimodal fusion, state modeling and reward optimization are key factors to improve the performance of the model.

The research contribution of this paper is mainly reflected in two aspects. On the one hand, starting from the teaching process of smart classroom, this paper integrated the dynamic teaching evaluation element refinement, multi-source data representation, state space construction, evaluation action design and reward feedback mechanism into a unified model framework, and formed a dynamic teaching evaluation method for the whole process of classroom. On the other hand, computer technologies such as computer vision, automatic speech recognition, natural language processing, temporal modeling and deep reinforcement learning were introduced into the teaching evaluation scene, which promoted the development of educational evaluation from static result analysis to the intelligent direction of data-driven, continuous evolution and strategy optimization.

At the same time, there are still some shortcomings in this paper. The number and collection range of sample courses are still limited. The current experiment is mainly based on 8 courses and 96 classroom records, and the generalization ability of the model in cross-course types and interdisciplinary scenarios still needs to be further tested. Multi-modal data acquisition relies on smart classroom equipment and platform logs, and still faces the problems of computing power consumption and system integration in real-time deployment. Although the model performs well in prediction performance and dynamic adaptation ability, there is still room for improvement in the interpretability of the strategic decision-making process and reward allocation mechanism. The follow-up research can focus on expanding the sample size, enhancing the ability of cross-scenario transfer, optimizing the lightweight deployment scheme, and introducing an interpretable reinforcement learning mechanism, so as to further improve the engineering applicability and promotion value of the dynamic teaching evaluation model in smart classroom.

Funding

This work was supported by Natural Science Basic Research Program of Shaanxi (Grant No. 2024JCYBMS576)

About the Author

Danxia Li was born in Yan’an, Shaanxi, P.R. China, in 1980. She received her Ph.D. from Beijing Institute of Technology. She is currently affiliated with the School of Mathematics and Computer Science at Yan’an University. Her primary research focuses on Machine Learning and higher education research.

References

[1] Díaz B, Nussbaum M. Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence[J]. Computers & Education, 2024, 217: 105071.

[2] Memarian B, Doleck T. A scoping review of reinforcement learning in education[J]. Computers and Education Open, 2024, 6: 100175.

[3] Yan L, Echeverria V, Jin Y, et al. Evidence-based multimodal learning analytics for feedback and reflection in collaborative learning[J]. British Journal of Educational

Technology, 2024, 55(5): 1900-1925.

[4] D'Angelo C M, Rajarathinam R J. Speech analysis of teaching assistant interventions in small group collaborative problem solving with undergraduate engineering students[J]. British Journal of Educational Technology, 2024, 55(4): 1583-1601.

[5] Moon J, Yeo S, Banihashem S K, et al. Using multimodal learning analytics as a formative assessment tool: Exploring collaborative dynamics in mathematics teacher education[J]. Journal of Computer Assisted Learning, 2024, 40(6): 2753-2771.

[6] Cukurova M. The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence[J]. British Journal of Educational Technology, 2025, 56(2): 469-488.

[7] Riedmann A, Schaper P, Lugrin B. Reinforcement learning in education: A systematic literature review[J]. International Journal of Artificial Intelligence in Education, 2025: 1-55.

[8] Mohammadi M, Tajik E, Martinez-Maldonado R, et al. Artificial intelligence in multimodal learning analytics: A systematic literature review[J]. Computers and Education: Artificial Intelligence, 2025, 8: 100426.

[9] Banihashem S K, Gašević D, Noroozi O. A critical review of using learning analytics for formative assessment: Progress, pitfalls and path forward[J]. Journal of Computer Assisted Learning, 2025, 41(3): e70056.

[10] Tang Q, Deng W, Huang Y, et al. Can generative artificial intelligence be a good teaching assistant?—an empirical analysis based on generative ai-assisted teaching[J]. Journal of Computer Assisted Learning, 2025, 41(3): e70027.

[11] Jiang Y, Klebanov B B, Hao J, et al. Unveiling patterns of interaction with automated feedback in Writing Mentor and their relationships with use goals and writing outcomes[J]. Journal of Computer Assisted Learning, 2025, 41(2): e70014.

[12] Echeverria V, Nieto G F, Zhao L, et al. A learning analytics dashboard to support students' reflection on collaboration[J]. Journal of Computer Assisted Learning, 2025, 41(1): e13088.

[13] Suraworachet W, Zhou Q, Cukurova M. University Students' Perceptions of a Multimodal AI System for Real-World Collaboration Analytics: Lessons Learned From a Case Study[J]. Journal of Computer Assisted Learning, 2025, 41(5): e70103.

[14] Bautista G, López-Costa M. Smart learning spaces considering the integration of the pedagogical, environmental and digital dimensions: a systematic review[J]. Learning Environments Research, 2025, 28(3): 455-472.

[15] Zion Y B, Yakov S, Abramovitch E, et al. AI-Based Teaching Evaluations: How Well Do They Reflect Student Perceptions?[J]. Computers and Education: Artificial Intelligence, 2025: 100448.

[16] Fütterer T, Goldberg P, Bühler B, et al. Artificial intelligence in classroom management:

A systematic review on educational purposes, technical implementations, and ethical considerations[J]. Computers and Education: Artificial Intelligence, 2025: 100483.

[17] Er E, Akçapınar G, Bayazıt A, et al. Assessing student perceptions and use of instructor versus AI-generated feedback[J]. British Journal of Educational Technology, 2025, 56(3): 1074-1091.

[18] Bauer E, Richters C, Pickal A J, et al. Effects of AI-generated adaptive feedback on statistical skills and interest in statistics: A field experiment in higher education[J]. British Journal of Educational Technology, 2025, 56(5): 1735-1757.

[19] Ba S, Zhan Y, Huang L, et al. Investigating the impact of ChatGPT-assisted feedback on the dynamics and outcomes of online inquiry-based discussion[J]. British Journal of Educational Technology, 2025, 56(5): 1710-1734.

[20] Weidlich J, Gotsch F, Schudel K, et al. Teacher, peer, or AI? Comparing effects of feedback sources in higher education[J]. Computers and Education Open, 2025: 100300.