Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 15 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i15.62710

PAPER

Design of a Mobile Interactive Distance Education Platform with Student Engagement Assessment

Qifeng Du[1], Sheng Guo[2], Hongwei Zhen[3] (✉)

[1]Handan Open University, Handan, China

[2]Hebei Open University, Shijiazhuang, China

[3]Langfang Open University, Langfang, China

13722669166@163.com

ABSTRACT

The continuity of interaction and the reliability of learning analytics in mobile distance education are substantially constrained by dynamically fluctuating network conditions and the challenges associated with heterogeneous device adaptation. Existing educational systems are often limited by inadequate transmission stability under weak-network environments, privacy risks arising from cloud-centric learning analytics, and inefficient synchronization across multiple terminal devices. To address these limitations, a hierarchical end–edge–cloud microservice architecture for distance education was proposed. An adaptive transmission degradation mechanism based on a dual-protocol stack and integrated bandwidth time-series prediction was developed. In addition, a lightweight on-device engagement assessment model was established using multidimensional implicit interaction features, while a collaborative whiteboard synchronization algorithm optimized for heterogeneous mobile devices was designed. The proposed framework effectively addresses the challenges of dynamic adaptation and real-time learning-state perception in mobile learning environments. Technical support is thereby provided for the development of high-performance, privacy-preserving intelligent mobile education systems.

KEYWORDS

cross-platform mobile development, adaptive protocol transmission, lightweight on-device neural network, operational transformation algorithm, mobile learning engagement assessment

1 INTRODUCTION

The widespread adoption of mobile communication technologies and intelligent terminal devices has facilitated the transition of distance education from conventional fixed-network environments to ubiquitous mobile interactive learning environments [1]. Compared with wired networks, mobile learning environments are characterized by dynamic bandwidth fluctuations, unstable communication links, and pronounced device heterogeneity [2, 3], by which stringent requirements are

Du, Q., Guo, S., Zhen, H. (2026). Design of a Mobile Interactive Distance Education Platform with Student Engagement Assessment. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(15), pp. 22–36. https://doi.org/10.3991/ijim.v20i15.62710

Article submitted 2026-04-22. Revision uploaded 2026-06-20. Final acceptance 2026-06-22.

imposed on the real-time performance and continuity of instructional interactions. Most existing distance education platforms continue to rely on traditional wired-network architectures [4, 5], through which the dynamic transmission characteristics of mobile networks cannot be effectively accommodated. Consequently, under weak-network conditions, service degradation frequently occurs in the form of audio–video stuttering, communication interruptions, and classroom disconnections, resulting in substantial deterioration of instructional interaction quality. With respect to learning-state assessment [6], cloud-based visual analysis frameworks are predominantly employed in current engagement assessment. Continuous uploading of user-related data is therefore required [7], leading to considerable communication overhead and introducing significant privacy and security concerns. In addition, the adaptability of existing collaborative whiteboard technologies to heterogeneous mobile devices remains limited. Under high-concurrency interactive scenarios, synchronization conflicts are frequently generated, thereby restricting the large-scale deployment of mobile teaching applications. To address these challenges, research on instructional system optimization and learning-state perception in mobile learning environments is conducted. Theoretical foundations for adaptive mobile transmission and on-device intelligent perception are expected to be further advanced, while existing limitations in collaborative instruction across heterogeneous mobile devices are intended to be mitigated. Engineering guidance can thereby be provided for the development of highly stable and privacy-preserving intelligent education platforms, offering both significant theoretical contributions and practical value.

Although extensive efforts have been devoted to transmission optimization for mobile learning, learning-state assessment, and multi-terminal collaborative interaction, substantial limitations remain within existing technological frameworks, thereby restricting their applicability to complex mobile teaching environments. In the area of media-stream transmission, most mainstream distance education platforms rely on a single Web Real-Time Communication (WebRTC) protocol [8, 9]. Network adaptation can therefore only be performed reactively after communication degradation has occurred, while capabilities for network time-series prediction and proactive adaptation remain largely absent. As a consequence, service continuity is frequently compromised under weak-network conditions. Furthermore, a quantifiable framework for dynamic protocol adaptation has not yet been established. With respect to student engagement assessment, conventional cloud-based visual monitoring approaches are associated with considerable privacy leakage risks [10, 11] and are increasingly incompatible with the privacy compliance requirements of end-user devices.

Existing lightweight on-device assessment methods are primarily constructed on explicit learning data, whereas implicit interaction behavior features are often neglected. Consequently, assessment dimensions remain limited, evaluation accuracy is constrained, and a closed-loop mechanism integrating learning-state feedback and instructional intervention has yet to be realized. In the field of multi-terminal collaboration, traditional operational transformation algorithms have not been adequately adapted to the heterogeneous resolution characteristics of mobile devices. As a result, interaction-coordinate deviations and inconsistencies in interface states are frequently introduced. Moreover, optimization strategies for high-concurrency interactions remain insufficiently developed, leading to substantial increases in synchronization latency and system failure rates when large-scale instructional activities are conducted [12, 13]. At the architectural level, most existing distance education platforms continue to employ cloud-centric centralized deployment models [14, 15]. Task scheduling has generally not been performed according to the heterogeneous computational capabilities available across end, edge, and cloud layers. In addition, quantitative latency-constrained design criteria specifically tailored for mobile interactive learning environments remain lacking. Consequently, the requirements for instructional real-time performance and overall system stability cannot be simultaneously satisfied within existing frameworks.

To address the aforementioned research gaps, four principal contributions are presented. First, an end–edge–cloud hierarchical microservice architecture tailored for heterogeneous mobile devices is established, and a full-stack quantitative latency-constrained framework is developed. Second, by integrating network-state estimation and time-series prediction, a transmission mechanism based on a dual-protocol stack and proactive degradation strategy is proposed, thereby achieving dynamic adaptation and seamless protocol switching of instructional communication under weak-network conditions. Third, a lightweight on-device neural network model is designed on the basis of multidimensional implicit interaction features. Real-time engagement detection and adaptive instructional intervention are thereby enabled without the transmission of privacy-sensitive user data. Finally, the conventional operational transformation algorithm is optimized for high-concurrency mobile learning scenarios. Through the incorporation of coordinate normalization and batch-processing strategies, synchronization efficiency and concurrent processing capacity across heterogeneous mobile devices are substantially improved.

2 SYSTEM DESIGN AND CORE METHODS

2.1 Overall architecture of the end–edge–cloud hierarchical microservice system

An end–edge–cloud hierarchical microservice architecture is established to support heterogeneous mobile teaching environments. System deployment is achieved through the integration of cross-platform mobile terminals, edge protocol gateways, and cloud-based microservice clusters, while a time-series database is employed for the standardized storage and management of instructional behavioral data. The overall architecture and core operational mechanisms of the end–edge–cloud collaborative mobile distance education system are illustrated in Figure 1. Based on the heterogeneous computational resources available across the end, edge, and cloud layers, hierarchical task scheduling and tiered data-flow processing are implemented throughout the architecture. Lightweight feature extraction and local preprocessing of user interaction behaviors are performed independently on mobile devices, thereby preventing bandwidth consumption and privacy risks associated with the transmission of raw user data. Edge nodes continuously monitor network conditions and execute protocol adaptation and transmission-degradation decisions in real time, ensuring dynamic adaptation of instructional interactions under varying network environments. The cloud layer is primarily responsible for iterative model training and long-term analysis of instructional behaviors, thereby supporting the continuous evolution of the system's intelligent capabilities.

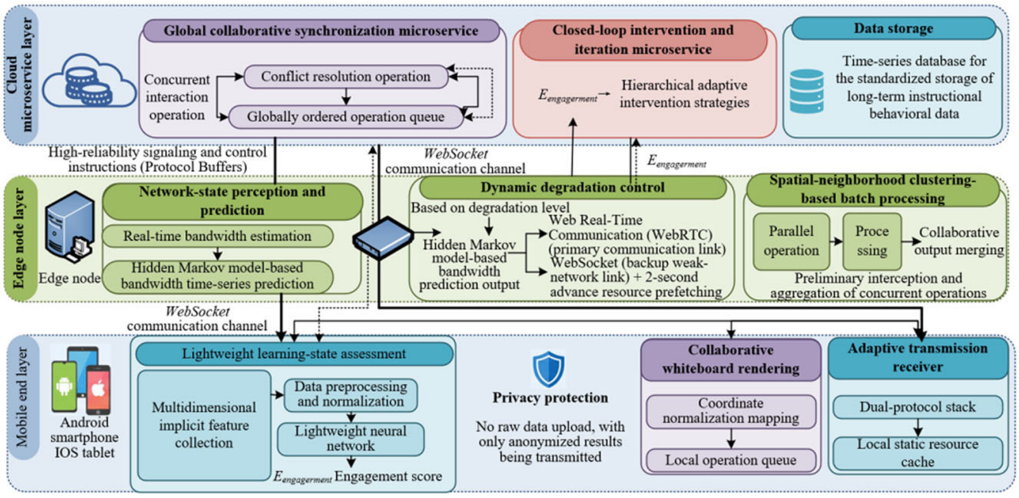

Fig. 1. Overall architecture and core mechanisms of the end–edge–cloud collaborative mobile distance education system

To satisfy the stringent real-time interaction requirements of mobile learning environments, a full-stack quantitative latency-constrained framework is established. End-to-end latency for control signaling is strictly maintained below 100 ms, while media-stream transmission latency is constrained to less than 400 ms under normal network conditions. Under weak-network conditions characterized by significant network fluctuations, the upper latency threshold for media-stream transmission is limited to 800 ms. These latency constraints provide standardized quantitative criteria for the subsequent design and optimization of the adaptive transmission mechanism, the learning-state assessment model, and the collaborative synchronization algorithm.

2.2 Dual-protocol stack adaptive degradation transmission architecture based on hidden Markov model-based bandwidth prediction

A dual-protocol stack transmission architecture is employed to achieve dynamic adaptation to mobile network conditions. Two communication channels, namely WebRTC and WebSocket, are deployed to support differentiated instructional interaction tasks. The WebRTC channel serves as the primary communication link and is responsible for real-time audio–video transmission under high-bandwidth conditions. In contrast, the WebSocket channel functions as a backup communication link for instructional interactions conducted under weak-network conditions characterized by limited bandwidth and severe network disturbances. Real-time bandwidth estimation is performed using the Kalman filter embedded within the Google congestion control framework, and an instantaneously available bandwidth value, denoted as $\hat{B}$, is obtained. By combining the packet loss rate (p) and the round-trip time (RTT), a standardized channel-switching decision model is established. Optimal threshold values are determined through multiple pilot experiments, enabling adaptive and dynamic selection of the transmission channel. The decision rule is defined as follows:

$$S(t) = \begin{cases} \text{WebRTC}, & \hat{B} \geq B_{th} \wedge p \leq p_{th} \wedge RTT \leq RTT_{th} \\ \text{WebSocket}, & otherwise \end{cases} \tag{1}$$

where, the threshold parameters in the system are specified as B_{th} = 300 kbps, p_{th} = 10%, and RTT_{th} = 400 ms. These values correspond to the minimum bandwidth requirement for instructional video encoding, the critical packet-loss rate for maintaining communication stability, and the maximum tolerable latency for real-time instructional interaction in mobile learning environments. Consequently, a quantifiable and reproducible basis for transmission-channel switching is provided.

To overcome the transient service interruptions caused by reactive channel-switching strategies in conventional approaches, a hidden Markov model is introduced to capture the temporal evolution characteristics of network bandwidth and enable proactive transmission degradation. A continuous bandwidth observation sequence over the most recent 10-second interval, denoted as $[\hat{B}_{t-9},\ldots,\hat{B}_t]$, is continuously collected by the system. By estimating state-transition probabilities, the fluctuation patterns of mobile network conditions are modeled, and the bandwidth at the next time step, $\hat{B}_{t+1}$, is predicted. When the predicted bandwidth satisfies the condition $\hat{B}_{t+1} < 1.2 \times B_{th}$, the network is identified as approaching a weak-network state. Consequently, a proactive resource-prefetching mechanism is activated two seconds in advance. Critical static instructional resources, including courseware snapshots and exercise content, are cached locally on the mobile device. As a result, interface blanking and interaction interruptions that commonly occur during protocol switching are effectively mitigated, thereby substantially improving service continuity in weak-network learning environments.

To ensure interaction-state consistency throughout the protocol-switching process, interaction commands are encapsulated using Protocol Buffers-based binary serialization. The size of each interaction command is maintained within the range of 40–120 bytes, resulting in a substantially lower transmission overhead than that associated with conventional media-stream frames under weak-network conditions. In addition, an ordered operation queue mechanism based on monotonically increasing sequence numbers is established. A globally unique temporal identifier is generated by the client for each local interaction event, while command execution and synchronization broadcasting are performed by the server in strict accordance with the assigned temporal order. Furthermore, a retransmission timeout mechanism of 500 ms is implemented. When command loss is detected due to packet loss, retransmission requests are automatically triggered. If command recovery is not completed within the specified timeout interval, the corresponding command is regarded as invalid, and the client state is synchronously updated. Through this mechanism, operational conflicts and state inconsistencies arising during dual-protocol switching are effectively eliminated.

2.3 Lightweight real-time engagement assessment model driven by implicit interaction features on mobile devices

To address the privacy risks and limited behavioral representation capabilities associated with existing learning-state assessment approaches, a lightweight engagement assessment model operating entirely on mobile devices is developed. Throughout the assessment process, no facial images, raw interaction records, or other privacy-sensitive information are transmitted, thereby eliminating communication overhead and mitigating information leakage risks at the data source. A 60-second sliding temporal window is employed to capture users' implicit classroom interaction behaviors, and a 12-dimensional quantitative feature framework is established to comprehensively characterize learning-state dynamics. To achieve standardized quantitative representation of behavioral patterns, the principal interaction indicators are computed using statistical measures. The key feature formulations are defined as follows:

$$CV_{click} = \frac{\sigma_{click}}{\mu_{click}} \tag{2}$$

$$R_{ans} = \frac{T_{response}}{T_{expected}} \tag{3}$$

$$\rho_{attention} = \frac{1}{12}\sum_{i=1}^{12} a_i \tag{4}$$

$$H = -\sum_{j} p(s_j)\log p(s_j) \tag{5}$$

where, σ_{click} and μ_{click} denote the standard deviation and mean of click intervals, respectively; $T_{response}$ and $T_{expected}$ represent the actual response time and the expected response time for instructional tasks; a_i denotes the screen-attention state indicator; and $p(s_j)$ represents the probability distribution associated with user interface dwell states. Prior to model input, all features are normalized using the Z-score normalization method to eliminate biases arising from differences in measurement scales and feature magnitudes. The normalization procedure is expressed as:

$$\tilde{x}_i = \frac{x_i - \mu_i}{\sigma_i} \tag{6}$$

where, μ_i and σ_i denote the global mean and standard deviation of the corresponding feature, respectively, estimated from large-scale offline instructional interaction logs.

A three-layer fully connected neural network is constructed to achieve deep fusion of implicit features and regression of engagement scores. The network architecture is intentionally streamlined to accommodate the computational constraints of embedded mobile devices. The forward propagation process is defined as follows:

$$h_1 = ReLU(W_1\tilde{x} + b_1), \quad h_2 = ReLU(W_2h_1 + b_2), \quad y = \sigma(W_3^T h_2 + b_3) \tag{7}$$

Through two successive nonlinear transformations, high-dimensional features are refined and compressed. An engagement score, denoted as $E_{engagement}$, is subsequently generated through a Sigmoid activation function, yielding an output within the interval [0, 1]. The model is converted into a lightweight format using TensorFlow Lite. In addition, graphics processing unit acceleration is enabled through the mobile hardware delegate mechanism. Deployment on mainstream mobile devices demonstrates that the single inference latency is limited to approximately 15 ms, thereby consistently satisfying the low-latency requirements of real-time learning-state assessment in instructional environments.

To establish a dynamic closed-loop framework integrating learning-state perception and instructional regulation, a hierarchical adaptive intervention strategy is developed on the basis of continuous temporal engagement assessment results. Differentiated instructional guidance mechanisms are activated according to predefined quantitative thresholds. During the intervention process, the difficulty level of adaptive exercises is dynamically adjusted according to students' historical learning performance. The difficulty adaptation function is defined as:

$$Diff = 0.3 + 0.5 \times (1 - accuracy) \tag{8}$$

Intervention levels are determined according to engagement scores observed over two consecutive sampling windows. Differentiated instructional actions are executed, including interface reminders, adaptive exercise recommendations, and temporary instructional review sessions. Following each intervention, subsequent user interaction behaviors are incorporated into the feature-calculation framework in real time and are utilized for continuous refinement of the learning-state assessment results. Consequently, a dynamic closed-loop mechanism is established, enabling continuous refinement of learning-state perception and instructional intervention.

2.4 Improved operational transformation-based collaborative whiteboard synchronization algorithm for heterogeneous mobile devices

To ensure global state consistency during collaborative whiteboard interactions across heterogeneous mobile devices, an operational transformation mechanism is employed to resolve concurrent writing conflicts in instructional environments. All whiteboard interaction events are represented using a standardized structured format, in which each interaction operation is defined as a quadruple $O = \langle t, u, type, data \rangle$. For any two concurrent interaction operations, denoted as O_a and O_b, operation alignment and adaptation are achieved through a conflict resolution operation. The unified transformation rule is expressed as:

$$transform(O_a, O_b) = (O'_a, O'_b) \tag{9}$$

In conventional operational transformation algorithms, conflict resolution is directly performed using device-specific screen coordinates. However, the heterogeneous screen resolutions and display dimensions of Android and iOS devices are generally not considered, resulting in trajectory deviations and rendering inconsistencies. To address this limitation, a global coordinate normalization mapping model is established. All interaction coordinates are first transformed into a unified normalized coordinate space in accordance with:

$$x = \frac{x_{raw}}{W}, y = \frac{y_{raw}}{H} \tag{10}$$

where, x_{raw} and y_{raw} denote the original touch coordinates recorded on the source device, and W and H represent the screen width and height, respectively. At the receiving device, inverse coordinate mapping is performed according to local hardware parameters. Through this mechanism, synchronization deviations arising from heterogeneous display configurations are fundamentally eliminated.

3 EXPERIMENTAL DESIGN AND RESULTS ANALYSIS

3.1 Experimental environment and dataset configuration

The experimental platform was composed of multiple mainstream mobile devices and a server-side cluster infrastructure. The mobile-device environment included heterogeneous terminals equipped with representative chipsets, including Qualcomm Snapdragon 865, MediaTek Dimensity 1100, and Apple A14 processors, thereby covering both Android and iOS platforms. On the server side, a Kubernetes-based microservice cluster was deployed on rack-mounted servers, while a high-performance time-series database was utilized for the storage of instructional data. The software environment was established using Flutter to implement a cross-platform mobile client. WebRTC was employed to provide the underlying media-transmission capability, while the NetEm network emulator was utilized to simulate dynamic weak-network conditions commonly encountered in mobile environments. TensorFlow Lite was adopted to support lightweight deployment of the proposed model on mobile devices and to enable hardware-accelerated inference.

Two self-constructed datasets were utilized during experimental evaluation. The first dataset consisted of 20 hours of manually annotated classroom interaction data collected from authentic teaching sessions. The dataset contained implicit interaction samples collected under diverse network conditions and learning behaviors and was employed for model training and accuracy validation of the engagement model. The second dataset was designed for collaborative whiteboard evaluation and contained concurrent interaction records generated by 200 mobile devices. Large-scale operation samples collected from devices with heterogeneous screen resolutions were included. This dataset was employed to evaluate the performance of the collaborative synchronization algorithm.

3.2 Comparative evaluation of dual-protocol stack transmission robustness under weak-network conditions

This experiment was conducted to evaluate the stability and real-time performance of instructional communication under weak-network conditions. The simulated network bandwidth was fixed at 384 kbps, corresponding to the typical transmission characteristics of a 3G mobile network. Five packet-loss levels, namely 0%, 5%, 10%, 15%, and 20%, were configured to represent progressively degraded network conditions. Two transmission schemes were evaluated. The baseline scheme consisted of a conventional WebRTC-only instructional transmission framework, whereas the experimental scheme employed the proposed dual-protocol stack adaptive degradation transmission architecture based on hidden Markov model-based bandwidth prediction. The evaluation metrics included session availability, average media transmission latency, average command transmission latency, and protocol-switching latency. The transmission performance of the two schemes under different weak-network conditions is summarized in Table 1.

Table 1. Performance comparison of transmission schemes under different packet-loss rates

Packet Loss Rate	Scheme	Session Availability (%)	Media Transmission Latency (ms)	Command Transmission Latency (ms)	Protocol-Switching Latency (ms)
0%	Conventional WebRTC scheme	98.6	182	52	Not applicable
0%	Proposed dual-protocol stack scheme	98.9	180	50	28
5%	Conventional WebRTC scheme	81.2	356	68	Not applicable
5%	Proposed dual-protocol stack scheme	97.5	312	49	30
10%	Conventional WebRTC scheme	65.7	589	85	Not applicable
10%	Proposed dual-protocol stack scheme	94.1	426	51	29
15%	Conventional WebRTC scheme	58.3	792	102	Not applicable
15%	Proposed dual-protocol stack scheme	92.3	615	53	31
20%	Conventional WebRTC scheme	41.5	985	136	Not applicable
20%	Proposed dual-protocol stack scheme	90.6	782	55	30

As shown in Table 1, the conventional WebRTC-only scheme exhibited pronounced sensitivity to network degradation. As the packet-loss rate increased, session availability declined substantially. Under the 15% packet-loss condition, session availability decreased to only 58.3%. Under the most severe condition, media transmission latency approached 1,000 ms. Such performance levels are insufficient to satisfy the real-time requirements of interactive mobile learning environments. In contrast, the proposed dual-protocol stack scheme, supported by time-series bandwidth prediction and a proactive transmission-degradation mechanism, consistently maintained a session availability rate exceeding 90% across all weak-network scenarios with varying packet-loss rates. As a result, the occurrence of instructional interruptions under mobile weak-network conditions was substantially reduced. Furthermore, the lightweight command-transmission mechanism implemented through the backup communication channel maintained command transmission latency below 55 ms across all experimental conditions. Protocol-switching latency remained stable at approximately 30 ms, indicating that transmission adaptation could be completed with negligible perceptible disruption to instructional interactions. Consequently, both service continuity and real-time responsiveness were effectively preserved under complex mobile network environments.

3.3 Evaluation of engagement assessment accuracy and on-device inference performance

To evaluate both the predictive accuracy and embedded deployment suitability of the proposed on-device lightweight engagement assessment model, two representative baseline methods were selected for comparison. Baseline Model 1 consisted of a cloud-based convolutional neural network facial-vision assessment model, while Baseline Model 2 employed a conventional support vector machine-based learning analytics model. The experimental model corresponded to the proposed lightweight neural network driven by multidimensional implicit features. A self-constructed classroom dataset annotated manually was utilized as the test benchmark. Performance was evaluated using four metrics: mean absolute error of the engagement score, three-class classification accuracy, single-sample inference latency on mobile devices, and peak memory consumption. The comparative results obtained for the three engagement assessment models are summarized in Figure 2 and Table 2.

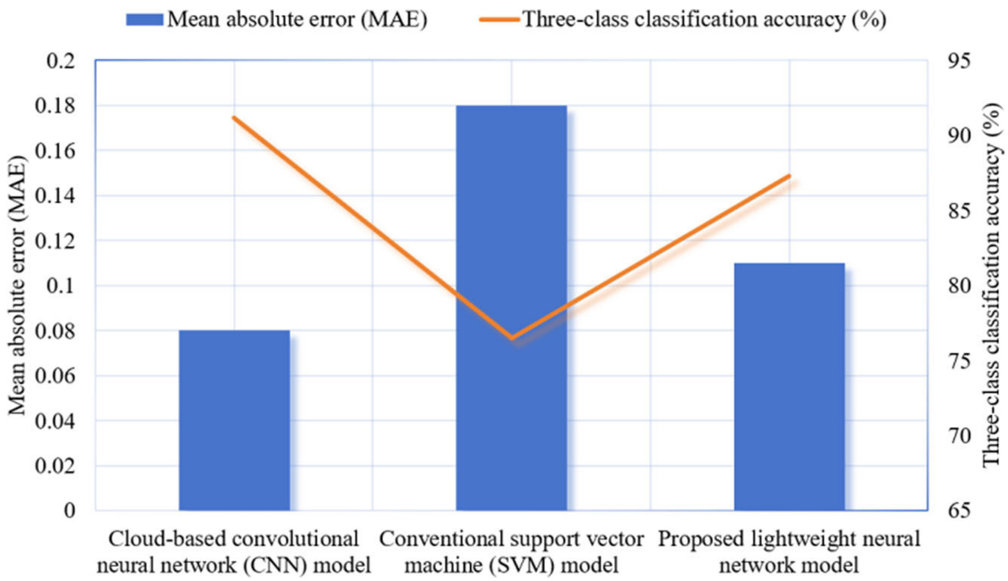

Fig. 2. Comparison of error and accuracy among different engagement assessment models

From the perspective of accuracy, the cloud-based convolutional neural network model achieved the highest performance owing to its ability to exploit large-scale visual features. However, this approach relies on the transmission of privacy-sensitive data and cloud-computing resources, rendering it unsuitable for privacy-preserving and lightweight on-device deployment scenarios. In contrast, the conventional support vector machine model was constructed using a limited set of explicit learning features. As a consequence, the restricted feature representation capability resulted in relatively large prediction errors, and the classification accuracy remained below 80%. The proposed model was developed using a 12-dimensional implicit interaction feature framework. The mean absolute error was reduced to 0.11, while the three-class classification accuracy reached 87.3%. These results indicate that substantially higher assessment accuracy was achieved compared with the conventional machine-learning approach, while performance remained highly comparable to that of the cloud-based visual model. From the perspective of deployment performance, the proposed model was optimized through TensorFlow Lite quantization and graphics processing unit acceleration. As a result, the single-sample inference latency was limited to only 15.1 ms, which is considerably lower than the commonly accepted real-time threshold of 100 ms. Furthermore, peak memory consumption was restricted to 10.5 MB, representing the lowest resource utilization among all evaluated approaches. These results demonstrate that the proposed model fully satisfies the constraints of mobile-edge deployment, including limited computational resources, low power consumption, and enhanced privacy protection, while maintaining high evaluation accuracy. Consequently, an effective balance between accuracy and practical engineering applicability is achieved.

Table 2. Performance comparison of different engagement assessment models

Model	Single-Sample Inference Latency (ms)	Peak Memory Consumption (MB)
Cloud-based convolutional neural network model	Cloud-side computation	28.6 (device-side)
Conventional support vector machine model	8.2	12.3
Proposed lightweight neural network model	15.1	10.5

3.4 Performance evaluation of whiteboard synchronization under high-concurrency heterogeneous device environments

This experiment was conducted to evaluate the performance of the improved operational transformation-based collaborative synchronization algorithm under heterogeneous device configurations and high-concurrency interaction scenarios. Four concurrency levels, corresponding to 50, 100, 150, and 200 simultaneous participants, were configured. The baseline method consisted of the native Jupiter operational transformation algorithm, whereas the experimental method employed the proposed improved operational transformation algorithm incorporating coordinate normalization and batch-processing optimization strategies. Performance was evaluated using three metrics: operation throughput (operations per second), average synchronization latency, and the occurrence rate of interface inconsistency failures. The comparative results obtained under different concurrency levels are presented in Table 3.

Table 3. Performance comparison of collaborative whiteboard synchronization algorithms under high-concurrency conditions

Number of Concurrent Users	Scheme	Operation Throughput (Operations/s)	Average Synchronization Latency (ms)	Interface Inconsistency Rate (%)
50	Native operational transformation algorithm	492	126	2.1
50	Improved operational transformation algorithm	1286	85	0.3
100	Native operational transformation algorithm	485	189	4.5
100	Improved operational transformation algorithm	1572	112	0.5
150	Native operational transformation algorithm	468	253	7.8
150	Improved operational transformation algorithm	1845	146	0.8
200	Native operational transformation algorithm	451	328	11.2
200	Improved operational transformation algorithm	2013	192	1.1

The experimental results demonstrate that the native operational transformation algorithm was constrained by its operation-by-operation serial processing mechanism and the lack of effective adaptation to heterogeneous coordinate systems. Consequently, throughput remained at approximately 500 operations per second, while synchronization latency and interface inconsistency rates increased markedly as the number of concurrent users grew. Under the 200-participant concurrency scenario, synchronization latency exceeded 300 ms, and the failure rate surpassed 11%, indicating that the native algorithm was unable to support large-scale classroom interaction effectively. By contrast, the proposed algorithm completely resolved cross-device resolution adaptation issues through a global coordinate normalization strategy. In addition, server-side computational overhead was substantially reduced through the incorporation of a regional operation batch-processing mechanism. As a result, a throughput of 2,013 operations per second was achieved under the 200-participant high-concurrency scenario, while synchronization latency was consistently maintained below 200 ms. Furthermore, the globally unified coordinate framework effectively prevented display misalignment and state-conflict issues, resulting in an overall failure rate of less than 1.2% across all experimental conditions. These findings demonstrate that the proposed algorithm substantially enhances the stability of large-scale collaborative teaching applications deployed on heterogeneous mobile devices.

3.5 Effectiveness evaluation of the hierarchical closed-loop engagement intervention strategy

To quantitatively evaluate the instructional benefits of the proposed hierarchical adaptive intervention mechanism, students from the same class were equally divided into two groups. A control group without intervention and an experimental group employing the proposed intervention strategy were established. A comparative instructional experiment was subsequently conducted over eight consecutive class sessions. Performance was assessed using two indicators: the average engagement score over the entire instructional period and the proportion of time spent in a low-engagement state. A low-engagement state was defined as any instructional period during which the engagement score remained below 0.4. The statistical results of classroom learning-state data obtained from the two groups are presented in Figure 3.

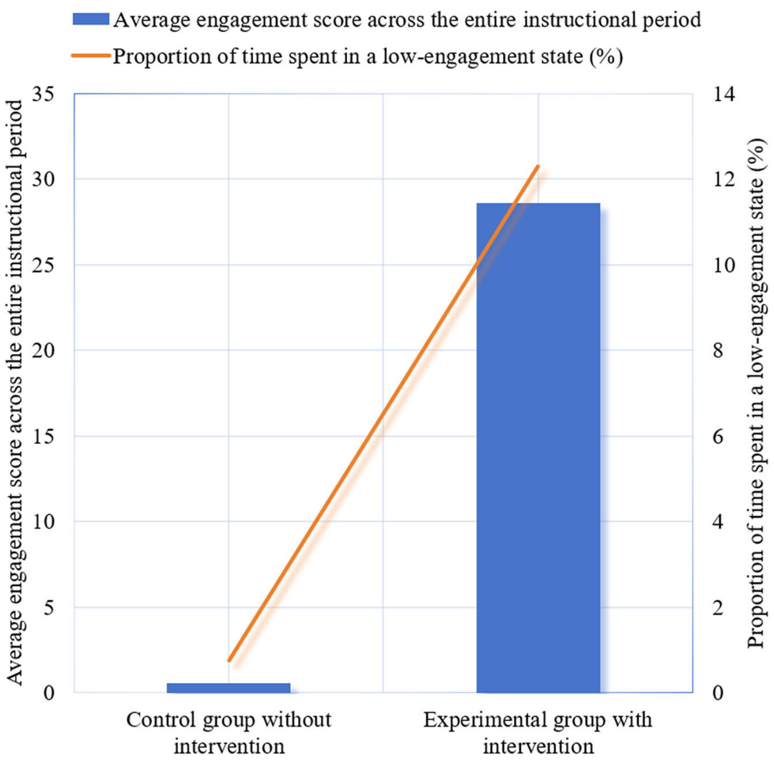

Fig. 3. Comparison of the effectiveness of instructional intervention strategies

As indicated by the statistical results, relatively low levels of student engagement were observed under natural instructional conditions in the absence of external intervention. Approximately one-third of the total instructional duration was spent in a low-engagement state. Through the proposed hierarchical closed-loop intervention mechanism, differentiated instructional regulation strategies were dynamically selected according to students' real-time learning states. Student disengagement behaviors were corrected in a timely manner through lightweight interface reminders, adaptive exercise recommendations, and guided instructional review activities. Following continuous closed-loop intervention across multiple instructional sessions, a substantial increase in the average student engagement score was observed in the experimental group. Simultaneously, the proportion of time spent in low-engagement states was markedly reduced. These findings demonstrate that dynamic intervention strategies driven by on-device learning-state assessment can effectively enhance classroom learning states. Furthermore, an intelligent instructional closed-loop framework integrating perception, regulation, and iteration was successfully established. Consequently, the proposed approach exhibits strong practical applicability for real-world educational environments.

4 CONCLUSION

The challenges associated with dynamic network fluctuations, heterogeneous device adaptation, insufficient privacy-preserving learning-state perception, and inefficient multi-device synchronization in mobile distance education environments were addressed through the development of an intelligent instructional interaction and perception framework tailored for mobile learning scenarios. An end–edge–cloud hierarchical microservice architecture was established, and a standardized latency-constrained framework was developed to support hierarchical task scheduling and efficient processing of instructional data. By integrating a time-series bandwidth prediction mechanism, a dual-protocol stack adaptive transmission strategy was designed, through which service interruptions under weak-network conditions were effectively mitigated. Furthermore, a lightweight neural network model driven by implicit interaction features collected on mobile devices was constructed to enable real-time engagement assessment and adaptive instructional closed-loop intervention without the transmission of privacy-sensitive user data. In addition, the collaborative synchronization algorithm was enhanced through coordinate normalization and batch-processing optimization, thereby resolving state inconsistency issues arising during high-concurrency interactions across heterogeneous mobile devices.

The effectiveness and technical advantages of the proposed framework were systematically validated through multiple comparative experiments. Relative to conventional distance education approaches, substantial improvements were achieved in session availability under weak-network conditions, on-device learning-state assessment accuracy, and collaborative interaction performance across multiple devices. Stable operation was consistently maintained under realistic classroom environments characterized by dynamic mobile network conditions and heterogeneous hardware configurations. The proposed framework contributes to the advancement of theoretical and technical foundations for adaptive instructional transmission, on-device intelligent learning-state perception, and large-scale collaborative interaction among mobile terminals. Both methodological innovation and practical engineering value are demonstrated. Consequently, effective support can be provided for the large-scale deployment of mobile learning applications in scenarios such as rural distance education and mobile vocational training. Furthermore, a reliable technical paradigm is established for the design and optimization of high-performance, privacy-preserving intelligent mobile education systems.

5 REFERENCES

[1] Heriyanto, Y. Y. Prasetyawan, and I. Krismayani, "Distance learning information literacy: Undergraduate students experience distance learning during the COVID-19 setting," *Information Development*, vol. 37, no. 3, pp. 458–466, 2021. https://doi.org/10.1177/02666669211018248

[2] W. Zhao, "Driving the integration of mobile learning and blended learning models in higher education," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 5, pp. 45–59, 2025. https://doi.org/10.3991/ijim.v19i05.54529

[3] J. Wang, "Intelligent education based on mobile learning: Transitioning from traditional classrooms to adaptive learning environments," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 11, pp. 51–65, 2025. https://doi.org/10.3991/ijim.v19i11.56057

[4] A. Adem, E. Çakıt, and M. Dağdeviren, "Selection of suitable distance education platforms based on human–computer interaction criteria under fuzzy environment," *Neural Computing and Applications*, vol. 34, no. 10, pp. 7919–7931, 2022. https://doi.org/10.1007/s00521-022-06935-w

[5] H. Gao, "Online AI-guided video extraction for distance education with applications," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–7, 2022. https://doi.org/10.1155/2022/5028726

[6] M. Händel, N. Naujoks-Schober, and M. Dresel, "Metacognitive monitoring via strategies and judgments," *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, vol. 55, nos. 2–3, pp. 67–76, 2023. https://doi.org/10.1026/0049-8637/a000273

[7] R. Hu, Y. Guo, H. Li, Q. Pei, and Y. Gong, "Personalized federated learning with differential privacy," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9530–9539, 2020. https://doi.org/10.1109/jiot.2020.2991416

[8] S. Islam, M. Welzl, and T. Fladby, "Real-life implementation and evaluation of coupled congestion control for WebRTC media and data flows," *IEEE Access*, vol. 10, pp. 95046–95066, 2022. https://doi.org/10.1109/access.2022.3206041

[9] K. Wajda, A. Chodorek, and R. R. Chodorek, "The possibility of improving the range and quality of air-to-ground WebRTC-based IoT communication in emergency situations by employing an external transceiver," *Sensors*, vol. 24, no. 20, p. 6533, 2024. https://doi.org/10.3390/s24206533

[10] U. Župerl, K. Stepien, G. Munđar, and M. Kovačič, "A cloud-based system for the optical monitoring of tool conditions during milling through the detection of chip surface size and identification of cutting force trends," *Processes*, vol. 10, no. 4, p. 671, 2022. https://doi.org/10.3390/pr10040671

[11] F. J. Meyer *et al.*, "HydroSAR: A cloud-based service for the monitoring of inundation events in the Hindu Kush Himalaya," *Remote Sensing*, vol. 16, no. 17, p. 3244, 2024. https://doi.org/10.3390/rs16173244

[12] Y. Park and Y. Shin, "A block-based interactive programming environment for large-scale machine learning education," *Applied Sciences*, vol. 12, no. 24, p. 13008, 2022. https://doi.org/10.3390/app122413008

[13] Y. Watanabe, Y. Kawamoto, and N. Kato, "A novel routing control method using federated learning in large-scale wireless mesh networks," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9291–9300, 2023. https://doi.org/10.1109/twc.2023.3269785

[14] K. Tatavarty, M. Johnson, and B. Rubinsky, "Cloud-based personalized sEMG classification using lightweight CNNs for long-term haptic communication in deaf-blind individuals," *Bioengineering*, vol. 12, no. 11, p. 1167, 2025. https://doi.org/10.3390/bioengineering12111167

[15] C. McRoy *et al.*, "Radiology education in the time of COVID-19: A novel distance learning workstation experience for residents," *Academic Radiology*, vol. 27, no. 10, pp. 1467–1474, 2020. https://doi.org/10.1016/j.acra.2020.08.001

6 AUTHORS

Qifeng Du graduated from Yanshan University with a master's degree in Marxist Philosophy. He is currently an Associate Professor at the School of Open Education, Handan Open University, Handan 056300, China, and has been recognized as a Youth Talent of Handan City. His research focuses on Marxist Educational Philosophy and Open Education Management. He has presided over one Handan City Philosophy and Social Science Project and completed one teaching reform and curriculum ideological-political education project at Handan Open University. He has published eight academic papers and served as the deputy editor of the textbook Teaching Reference for Introduction to the Basic Principles of Marxism (E-mail: 1669081337@qq.com).

Sheng Guo graduated from the Hebei University of Science and Technology with a Master degree, majoring in computer science (master degree). He is currently a Lecturer at the School of Community Education, Hebei Open University, Shijiazhuang 050081, China. His research interests include open distance education, computing. He presided over one Scientific Research of Higher Education Institutions project in Hebei Province, presided over and completed one research project on social science development in Hebei Province, published more than 6 papers, and edited 2 monographs (E-mail: emailguo@163.com).

Hongwei Zhen graduated from Hebei Normal University with a bachelor's degree. She currently serves as an Associate Professor at the Community Education Department, Langfang Open University, Langfang 065000, China, a special researcher for the Hebei Provincial Ideological and Political Work Research Association and a model teacher in Langfang City, majoring in English education and ideological and political education. Her teaching quality has been evaluated as "Excellent" for consecutive years. In terms of academic output, she has presided over 2 municipal-level projects in Langfang City, published more than 10 academic papers, authored 1 personal academic monograph, and has made proactive contributions to teaching practice in her field (E-mail: 13722669166@163.com).