Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 13 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i13.62396

PAPER

Mobile Interaction Technology-Driven Blended Learning for Higher Education

Aihua Gui (✉)

Zhumadian Preschool Education College, Zhumadian, China

flower7098@163.com

ABSTRACT

In blended learning environments in higher education, the use of mobile devices has long been constrained to content delivery and superficial interaction, whereas the multimodal sensing and edge-computing capabilities of smartphones have not been fully exploited. A mobile interaction technology-driven blended learning model was proposed, in which a device–edge–cloud collaborative federated learning architecture was constructed to enable local sensor feature extraction and differential privacy perturbation on smartphone terminals, with only desensitized gradients being transmitted to edge gateways. To achieve unobtrusive learning-state perception, a posture-recognition pipeline integrating complementary filtering and a lightweight MobileNetV3 model was designed. Six classroom postures were classified in real time using the accelerometer, gyroscope, and magnetometer. In addition, intergroup distance was dynamically estimated through Bluetooth Received Signal Strength Indicator (RSSI) measurements combined with Kalman filtering. When the physical distance between learners remains below 2 m for more than 30 s, vibration notifications and personalized micro-quizzes are automatically triggered by the edge intelligence engine, with quiz difficulty adaptively adjusted according to the exponentially decayed historical accuracy rate of each learner. Furthermore, long short-term memory networks were employed to model touch pressure, screen states, and posture sequences, thereby enabling the adaptive delivery of post-class augmented-reality exercises. Multipath transmission control protocol (MPTCP) and Q-learning-based path scheduling were incorporated to support seamless session migration between classroom and outdoor learning scenarios. Experimental validation demonstrates that a high-accuracy and low-latency mobile interaction framework can be achieved using only native smartphone sensors. A practical and scalable technical solution is therefore provided for blended learning for ubiquitous higher education environments.

KEYWORDS

mobile interactive blended learning, edge federated learning, smartphone posture sensing, Bluetooth proximity detection, long short-term memory network, multipath transmission control protocol (MPTCP)

Gui, A. (2026). Mobile Interaction Technology-Driven Blended Learning for Higher Education. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(13), pp. 154–168. https://doi.org/10.3991/ijim.v20i13.62396

Article submitted 2026-02-28. Revision uploaded 2026-05-10. Final acceptance 2026-05-21.

1 INTRODUCTION

With the widespread adoption of smartphones and the large-scale deployment of 5G networks [1, 2], blended learning in higher education has been undergoing a profound transformation from fixed-location and fixed-terminal instruction toward ubiquitous and fragmented learning paradigms [3, 4]. Owing to their ubiquitous connectivity and increasingly powerful on-device computing capabilities, mobile devices have been regarded as an ideal medium for eliminating the boundaries between classroom and extracurricular learning, as well as between online and offline instructional environments [5]. However, in most existing blended learning practices, smartphones have primarily been utilized as content consumption tools for video playback, document browsing, and simple polling activities, whereas the multimodal sensing channels embedded in these devices, including the accelerometer, gyroscope, magnetometer, and proximity sensor, have remained substantially underutilized [6, 7]. In practice, these sensors are capable of continuously capturing learner posture, interpersonal proximity, and environmental context without the need for additional hardware. If deeply integrated with edge computing and lightweight artificial intelligence models, an implicit, low-latency, and high-fidelity instructional interaction paradigm could potentially be established [8, 9]. Through such integration, student engagement and cognitive retention quality in blended learning environments could be fundamentally enhanced.

A review of existing studies indicates that three critical gaps remain in the current application of mobile interactive technologies in education. First, learning-state perception has been excessively dependent on explicit user operations or external sensing devices [10]. Interaction modalities such as clicking, swiping, and real-time commenting can only reflect superficial behavioral patterns and are insufficient for capturing latent states such as fatigue, confusion, or collaborative depth. Meanwhile, professional devices such as eye trackers and electroencephalography headsets are characterized by high cost and strong intrusiveness, thereby limiting their long-term deployment in routine classroom settings. Second, most interaction mechanisms have adopted a cloud-to-device request–response architecture in which all smartphone data must be uploaded to remote servers for processing. Such a framework introduces response latencies ranging from several hundred milliseconds to multiple seconds, while the transmission of raw sensor data also raises substantial privacy concerns among students [11, 12]. Third, the issue of network switching during mobile learning processes has long been neglected [13–15]. When learners transition from classroom Wi-Fi networks to outdoor cellular networks, learning sessions are frequently interrupted, and learning progress is often lost, resulting in repeated authentication procedures and substantial degradation in the cognitive continuity of cross-scenario learning activities. Consequently, existing technological frameworks have failed to transform smartphones from passive display terminals into active sensing and collaborative computing nodes. Furthermore, real-time instructional intervention and seamless session migration have not yet been effectively achieved under privacy-preserving conditions.

To address the aforementioned limitations, a novel blended learning paradigm driven by mobile interactive technologies is proposed, with innovations

reflected in four major aspects. First, a device–edge–cloud collaborative federated learning architecture is constructed in which sensor features are extracted locally on smartphones and perturbed through differential privacy mechanisms, while only desensitized gradients are uploaded. Through this design, privacy preservation and collaborative group modeling are simultaneously achieved. Second, a posture-recognition pipeline integrating complementary filtering with a lightweight MobileNetV3 architecture is developed. By leveraging the accelerometer, gyroscope, and magnetometer, six classroom postures can be classified in real time, while the inference latency for a single window is maintained below 15 ms. Third, dynamic estimation of intergroup distance is achieved through the integration of Bluetooth Received Signal Strength Indicator (RSSI) measurements and Kalman filtering. When the physical distance between learners remains below 2 m for longer than 30 s, edge intelligence-driven interventions are automatically triggered, including vibration notifications and personalized micro-quizzes with difficulty levels adaptively adjusted according to an exponentially decayed historical accuracy function. Fourth, long short-term memory networks are employed to integrate touch pressure, screen states, and posture sequences for the prediction of learner confusion intervals, thereby enabling the delivery of augmented reality exercises. In addition, seamless session migration across Wi-Fi and cellular networks is achieved through the integration of the multipath transmission control protocol (MPTCP) and a Q-learning-based path scheduling algorithm.

2 CORE SYSTEM TECHNOLOGIES AND IMPLEMENTATION

2.1 Device–edge–cloud collaborative federated learning architecture

To achieve low-latency collaborative modeling of group learning states in classroom environments while simultaneously preserving individual student privacy, a device–edge–cloud collaborative federated learning architecture is adopted, as illustrated in Figure 1. Each student smartphone is configured as a terminal node, on which local feature extraction is first performed on the raw sensor data. Let the sensor tensor collected by the i-th terminal within the temporal window T_w be denoted as $X_i \in R^{C\times T_w}$, where $C = 8$ represents eight channels, including the three-axis accelerometer, three-axis gyroscope, ambient light intensity, and proximity sensor measurements. Through a feature extraction network $f_{local}(\cdot;\theta_{edge})$, which is pretrained at the edge layer and subsequently deployed to terminal devices, X_i is mapped into a low-dimensional embedding vector $h_i = f_{local}(X_i;\theta_{edge}) \in R^{32}$. To prevent privacy leakage caused by the transmission of raw sensor data, a local differential privacy mechanism is introduced at the embedding layer. Independent noise $\eta \sim Lap(0, \Delta f/\varepsilon)$ following a Laplace distribution is added to hi, where Δf denotes the sensitivity of the feature extraction function. Based on the Lipschitz constant between the network input and output, Δf is set to 0.1, whereas $\varepsilon = 0.5$ represents the privacy budget. The perturbed embedding vector is therefore obtained as $\tilde{h}_i = h_i + \eta$. Subsequently, the vector is transmitted to the in-class edge gateway through MPTCP, whereas the original sensor data are permanently retained on the local terminal device.

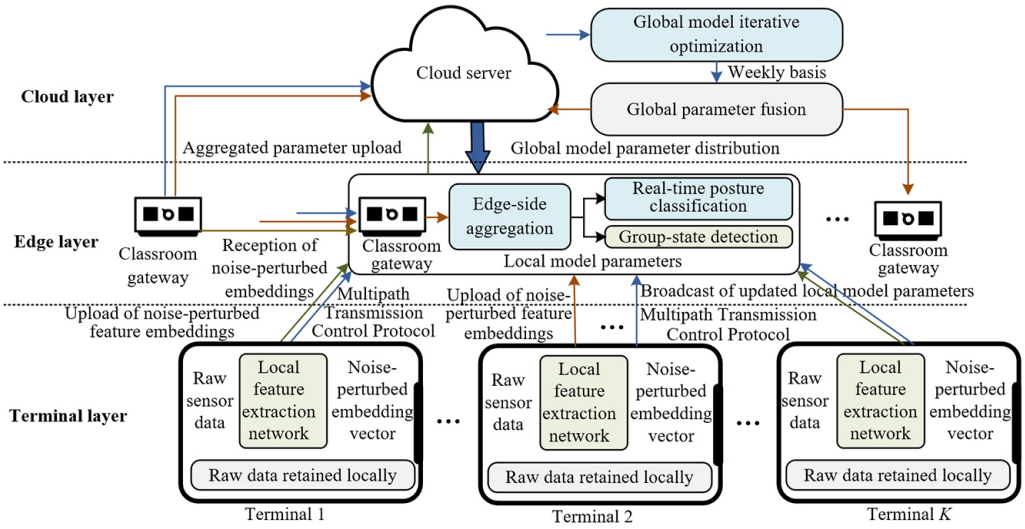

Fig. 1. Device–edge–cloud collaborative federated learning architecture

The edge gateway is responsible for aggregating the noise-perturbed embeddings generated by all terminal devices within the current classroom and for updating the local model parameters. The aggregation process is designed to support two categories of tasks: (i) real-time posture classification and group detection, and (ii) intermediate updates for the global model at the cloud layer. Assume that K active terminal devices are associated with a given edge gateway, where each terminal contributes n_i sample windows during the current round. A weighted averaging strategy is adopted for edge aggregation, and the update gradient is formulated as follows:

$$\theta_{edge}^{(r+1)} = \theta_{edge}^{(r)} + \eta_{edge} \cdot \frac{\sum_{i=1}^{K} n_i \nabla L_i\left(\theta_{edge}^{(r)}\right)}{\sum_{i=1}^{K} n_i} \tag{1}$$

where, ∇L_i denotes the cross-entropy loss gradient computed on terminal device i, and $\eta_{edge} = 0.01$. Local aggregation is completed at the edge layer every 200 ms, after which the updated parameters are broadcast to all terminal devices to ensure rapid adaptation of the posture-recognition model to individual behavioral variations. At the cloud layer, global parameter fusion is conducted once every ten rounds of edge communication. Weighted averaging aggregation is performed according to the student population associated with each classroom edge node. Iterative optimization of the global model is performed on a weekly basis, and the updated parameters are subsequently distributed to all edge gateways as the initialization parameters for the next round of classroom deployment. Through this architecture, the frequency of cloud-level communication is substantially reduced. Furthermore, low-dimensional noise-perturbed embeddings are transmitted instead of raw sensor data, thereby satisfying the theoretical constraints of ε-differential privacy. During practical classroom deployment, no leakage of student identity or behavioral characteristics is observed.

2.2 Posture sensing based on complementary filtering and MobileNetV3

The built-in accelerometer and gyroscope of smartphones, respectively, provide low-frequency accurate posture-angle measurements and high-frequency signals

affected by integration drift. Consequently, they should be fused to achieve stable posture estimation. In the proposed system, a first-order complementary filter is employed for angular recursive estimation. Let $\hat{\phi}(k)$ denote the pitch angle at the k-th sampling point, $\omega_x(k)$ represent the gyroscope angular velocity, and $a_x(k)$, $a_y(k)$, $a_z(k)$ denote the normalized accelerometer readings. The filter update equation is defined as follows:

$$\hat{\phi}(k) = \alpha\left(\hat{\phi}(k-1) + \Delta t \cdot \omega_x(k)\right) + (1-\alpha)\arctan\left(\frac{a_y(k)}{\sqrt{a_x^2(k) + a_z^2(k)}}\right) \tag{2}$$

where, $\alpha = \tau/(\tau + \Delta t)$. The time constant is set to $\tau = 0.1$ s, whereas the sampling interval is defined as $\Delta t = 0.02$ s. The roll angle $\hat{\psi}(k)$ is calculated using an analogous formulation. The yaw angle is extracted from magnetometer measurements after ellipsoidal calibration. Subsequently, the calibrated three-axis magnetic field intensity m is jointly resolved with the current pitch and roll angles. The posture-angle sequences obtained through complementary filtering constitute part of the input to the classification network.

To enable real-time recognition of learning postures, a MobileNetV3-Small network is deployed on the smartphone terminal. The network input consists of 100 temporal sampling points from eight sensor channels within the current window, corresponding to a tensor dimension of 8 × 100. The network architecture incorporates depthwise separable convolutions and squeeze-and-excitation modules, while the total number of parameters is maintained at 0.98 M. The output layer generates a probability vector $p \in R^6$ corresponding to six classroom postures: focused writing, passive listening with the smartphone placed flat, standing discussion, document reviewing, fatigue-related desk leaning, and side-body peer interaction. During training, a label-smoothed cross-entropy loss function is adopted:

$$L_{pose} = -\sum_{j=1}^{6}\left[(1-\epsilon)y_j + \frac{\epsilon}{6}\right]\log p_j \tag{3}$$

where the smoothing factor is set to $\epsilon = 0.1$ to suppress overfitting. To accommodate differences in behavioral habits among students, the parameters of the final fully connected layer are fine-tuned after each class session using the most recent 20 labeled samples. The regularized objective is defined as $L_{fine} = L_{pose} + \lambda\left\|W_{new} - W_{base}\right\|_2^2$, where $\lambda = 0.01$. During inference, the graphics processing unit delegate of TensorFlow Lite is utilized, resulting in an average inference latency of 12 ms per window. Figure 2 illustrates the complete inference pipeline integrating signal processing and deep learning from left to right.

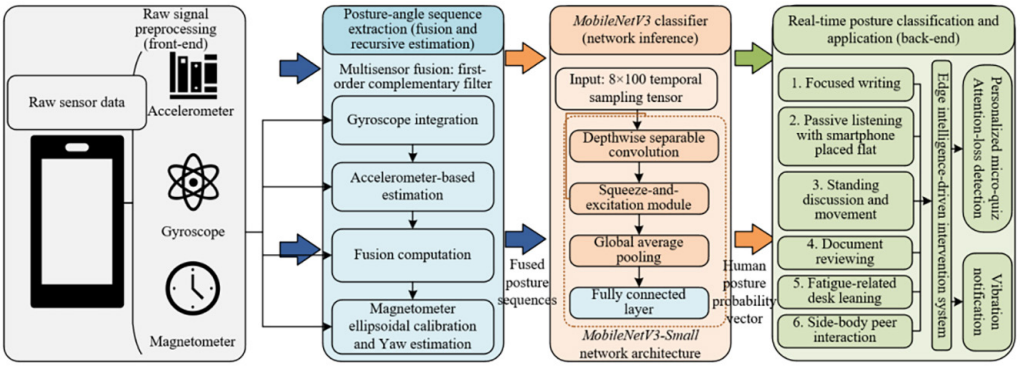

Fig. 2. Real-time posture sensing framework based on native multi-sensor fusion

2.3 Bluetooth received signal strength indicator-based distance estimation and group proximity detection

The physical proximity among group members in classroom environments constitutes an important indicator of collaborative engagement. In the proposed system, the Bluetooth RSSI is utilized to estimate interpersonal distance among students. Let $r_{ij}^{(m)}$denote M RSSI samples measured between student i and student j. After the application of median filtering, the logarithmic distance path-loss model is adopted:

$$\bar{r}_{ij} = r_0 - 10\gamma \log_{10}\left(\frac{d_{ij}}{d_0}\right) \tag{4}$$

where, $r_0 = -55$ dBm represents the reference value measured at the reference distance $d_0 = 1$ m, and $\gamma = 2.8$ denotes the path-loss exponent under classroom environmental conditions. In addition, $\sigma = 3$ dB. Subsequently, the interpersonal distance is inversely estimated through the least-squares method as $\hat{d}_{ij} = d_0 \cdot 10^{(r_0 - \bar{r}_{ij})/(10\gamma)}$.

Direct ranging results are highly susceptible to environmental noise caused by classroom occlusion and multipath interference. To address this limitation, a Kalman filtering algorithm is introduced for temporal optimization. By constructing a two-dimensional state space incorporating both distance and motion velocity, the predicted results are continuously corrected using real-time observations, thereby effectively suppressing instantaneous measurement errors and generating continuous and stable interpersonal distance estimations. When the filtered stable distance remains below 2 m for longer than 30 s, the system automatically identifies the corresponding group as entering a deep collaborative discussion state. Under this condition, exploratory extension problems are selectively delivered by the edge gateway, while collaboration duration is synchronously recorded to provide quantitative support for classroom interaction quality assessment.

2.4 Edge intelligence-driven real-time intervention engine

Instructional interventions are automatically triggered by the edge gateway according to each student's posture category and group dynamic distance, thereby reducing the need for frequent manual intervention by instructors. The attention-loss detection rule is defined as follows: when the posture category c_i of a given student corresponds to either passive listening with the smartphone placed flat or fatigue-related desk leaning, and this state remains unchanged for $N_{absent} = 15$ consecutive windows (equivalent to 30 s), a mobile intervention alert is immediately activated. The vibration intensity of the smartphone is progressively increased according to the duration of inattentiveness while remaining constrained by a predefined upper threshold, thereby enabling a gradual guidance mechanism to restore learner attention to classroom activities.

Simultaneously, lightweight micro-quizzes are automatically generated at the edge layer according to the real-time instructional content. Quiz difficulty is adaptively adjusted on the basis of each student's historical answering performance:

$$D = D_{max} \cdot exp\left(-\beta \frac{N_{correct}}{N_{total}}\right) \tag{5}$$

where $D_{max} = 0.9$, and $N_{correct}$ and N_{total}, respectively, denote the historical number of correctly answered questions and the total number of answered questions for the

corresponding student within the course. Question keywords are selected from the latent Dirichlet allocation topic matrix $\Phi \in R^{K \times V}$ cached at the edge layer, and semantic matching is subsequently performed to identify the topic vocabulary exhibiting the highest similarity to the recent instructional materials for question construction.

Classroom voting based on shake gestures is implemented through accelerometer-derived impact-feature recognition. The system continuously monitors composite acceleration fluctuations in real time. When two impact peaks exceeding 1.5 g are consecutively detected within an interval shorter than 0.5 s, the motion sequence is identified as a valid voting action. All voting options are uniformly encoded and compressed according to the Rabin–Karp hashing scheme. Consequently, encoded information from all terminal devices can be rapidly aggregated by the edge gateway, thereby enabling classroom-wide vote counting and result feedback within a second-level response latency.

2.5 Long short-term memory-based behavioral prediction and augmented reality exercise recommendation

To ensure that post-class instructional recommendations accurately match the cognitive states of learners, a long short-term memory network is constructed to predict confusion regions and attention-dispersion intervals for each student based on sensor sequences collected during classroom activities. The feature vector at each time step t is defined as $St \in R6$, including pitch angle, roll angle, yaw angle, average group RSSI, screen activation state, and average touch pressure. Classroom data collected during the preceding hour are organized into the temporal sequence $\{S1, S2, \ldots, ST\}$, where $T = 120$. The hidden-state dimension is set to 64. The output layer simultaneously generates binary confusion predictions $\hat{y}_t^{conf}$ and attention-engagement regression values $\hat{y}_t^{att}$. A joint loss function is adopted:

$$L_{\text{LSTM}} = \frac{1}{T}\sum_{t=1}^{T}\left[\text{BCE}(y_t^{conf}, \hat{y}_t^{conf}) + 0.5 \cdot (y_t^{att} - \hat{y}_t^{att})^2\right] \tag{6}$$

Model inference on smartphone terminals is accelerated through the Android Neural Networks Application Programming Interface (NNAPI). During post-class augmented reality exercise recommendation, exercise selection is performed according to the predicted confusion topics and learner engagement levels:

$$j^* = \arg\min_{j \in J}(0.7 \cdot \left\| embed_j - embed_{conf} \right\|_2 - 0.3 \cdot \log(1 + \hat{y}^{att})) \tag{7}$$

where $embed_j$ denotes the pre-trained knowledge-point vector, and $embed_{conf}$ represents the average embedding corresponding to the confusion region. The recommended exercises are rendered as three-dimensional overlays when dormitory desks or textbooks are scanned through the smartphone camera, thereby enhancing the interactivity and engagement of review activities.

2.6 Cross-scenario seamless session migration based on MPTCP and Q-learning path scheduling

Network interruptions are frequently encountered when students transition between classroom and outdoor learning scenarios. To address this limitation,

MPTCP is adopted in the proposed system, while a Q-learning-based algorithm is introduced for dynamic transmission-path selection. Let the available path set be denoted as $P = \{p_1, p_2, \ldots, p_m\}$. For each transmission path p, the observed features include round-trip latency R_p, packet-loss rate l_p, estimated bandwidth B_p, and the remaining smartphone battery ratio E. The state s_t is defined by the collective features of all available paths, whereas the action a_t corresponds to the path selected for allocation of the next data packet. The reward function is formulated as follows:

$$r_t = -\left(0.4R_{p_t} + 0.5 \cdot I_{retrans} - 0.1 \cdot \frac{B_{p_t}}{E}\right) \tag{8}$$

where, $I_{retrans}$ denotes the retransmission indicator variable. The Q-value update process follows the standard formulation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \eta[r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \tag{9}$$

The learning rate is set to $\eta = 0.1$, whereas the discount factor is defined as $\gamma = 0.95$. The Q-table is pretrained during the offline phase, and during online inference, the optimal transmission path is selected every 100 ms according to the current network state. When network switching becomes unavoidable, the learning session context is serialized by the edge gateway and simultaneously transmitted through both the old and new transmission paths after the application of Reed–Solomon (4, 2) forward error correction encoding. At the receiving side, complete recovery of the session context can be achieved once any two encoded fragments are successfully received, thereby enabling zero-packet-loss session migration. Experimental evaluation demonstrates that, during transitions from classroom Wi-Fi networks to outdoor 5G environments, the median session reconstruction latency remains below 50 ms.

3 EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Mobile sensing performance and resource overhead

Table 1. Comparison of mobile sensing performance and resource overhead

Method	Average Classification Accuracy	Recall of Focused-Writing Posture	Recall of Fatigue-Related Desk-Leaning Posture	Inference Latency per Window (ms)	Average Central Processing Unit Utilization	Average Graphics Processing Unit Utilization	Hourly Battery Consumption Increment	Memory Increment (MB)
Proposed Framework	91.3%	92.1%	90.6%	12.4 ± 2.1	18.2%	21.5%	9.7%	42.3
Baseline 1	84.7%	86.5%	83.2%	47.8 ± 5.3	32.6%	38.9%	18.4%	87.6
Baseline 2	88.6%	89.4%	87.3%	15.8 ± 2.7	22.5%	26.3%	12.8%	51.7

Table 1 presents a comparison of posture-recognition accuracy and resource consumption between the proposed framework and three baseline methods. Baseline 1 consisted of a conventional extended Kalman filter combined with ResNet18, representing a high-accuracy yet computationally intensive approach. Baseline 2 employed complementary filtering integrated with MobileNetV2, representing a commonly adopted lightweight architecture. Experimental data were collected from 15 volunteers, yielding a total of 4.5 h of synchronized video annotation data. The proposed integration of complementary filtering and MobileNetV3 achieved an average classification accuracy of 91.3% across six posture categories, which was substantially higher than the 84.7% obtained by Baseline 1 and the 88.6% achieved by Baseline 2. For the focused-writing posture, a recall rate of 92.1% was achieved, exceeding that of Baseline 2 by 2.7 percentage points. In addition, the recall rate for fatigue-related desk-leaning posture reached 90.6%, representing an improvement of 3.3 percentage points over Baseline 2. From the perspective of inference latency, only 12.4 ms were required per window in the proposed framework, corresponding to approximately 26% and 78% of the latency observed in Baseline 1 and Baseline 2, respectively. Both central processing unit and graphics processing unit utilization rates were also minimized, reaching 18.2% and 21.5%, respectively. Furthermore, the hourly battery consumption increment was maintained at 9.7%, remaining below the 10% acceptable deployment threshold and outperforming both Baseline 2 (12.8%) and Baseline 1 (18.4%). These results demonstrate that the integration of MobileNetV3 and complementary filtering achieved a more effective balance between recognition accuracy and resource efficiency, thereby rendering the framework suitable for long-term deployment in classroom environments.

3.2 Bluetooth ranging and group detection performance

Table 2. Comparison of Bluetooth ranging and group detection performance

Method	Median Ranging Error (m)	Root Mean Square Ranging Error (m)	Precision of Group Discussion Detection	Recall of Group Discussion Detection	Median Triggering Latency (s)	90th Percentile Triggering Latency (s)	Received Signal Strength Indicator (RSSI) Standard Deviation (dB)
Original RSSI Logarithmic Model	3.2	3.8	69.2%	62.5%	15.4	28.7	4.7
Weighted Moving-Average Smoothing	1.5	2.0	82.7%	78.4%	9.8	15.3	2.9
Proposed Framework	0.8	1.1	94.4%	92.0%	6.2	9.1	2.1

Table 2 compares the performance of the original RSSI logarithmic model, weighted moving-average smoothing, and the proposed RSSI–Kalman filtering framework in terms of distance estimation and group detection. Experiments were conducted in three standard classroom environments, where four student groups consisting of five members each were organized. Scripted interaction scenarios,

including intensive discussion, dispersed seating, and random movement, were executed. Ground-truth distances were recorded using a laser distance meter. For the original model, the median ranging error reached 3.2 m. After the application of weighted moving-average smoothing, the error was reduced to 1.5 m, whereas the proposed framework further reduced the median ranging error to 0.8 m. Similarly, the root mean square ranging error decreased from 3.8 m to 2.0 m and 1.1 m, respectively. With respect to deep collaborative discussion detection, the proposed framework achieved a precision of 94.4%, which was substantially higher than the 82.7% obtained by weighted moving-average smoothing and the 69.2% achieved by the original model. The recall rate also reached the highest value of 92.0%. In addition, the median triggering latency of the proposed framework was reduced to 6.2 s, corresponding to a 37% reduction compared with the 9.8 s observed for weighted moving-average smoothing. After the application of Kalman filtering, the RSSI standard deviation was reduced to 2.1 dB, indicating superior robustness against environmental noise and human-body occlusion fluctuations. Although weighted moving-average smoothing was capable of partially suppressing noise fluctuations, state drift under dynamic classroom conditions could not be effectively addressed. By contrast, the recursive estimation characteristics of Kalman filtering rendered it more suitable for dynamic indoor ranging scenarios.

3.3 Effectiveness of edge intelligence-driven intervention

To evaluate the independent effectiveness of the intelligent intervention mechanism, a random-reminder control group was introduced in addition to the no-intervention control group. In the random-reminder group, vibration alerts and micro-quizzes were delivered at fixed intervals with the same frequency as the proposed framework; however, the triggering moments were unrelated to learner posture states. A total of 30 students were randomly assigned to three groups, with 10 students in each group, and the experiment was conducted over a two-week period. As shown in Table 3, the proposed intelligent intervention group achieved a posture-transition rate of 71.4% within 30 s after intervention triggering, which was substantially higher than the 48.5% observed in the random-reminder group and the 37.2% obtained in the no-intervention control group. The completion rate and accuracy of micro-quizzes were likewise the highest in the proposed framework, reaching 84.9% and 73.2%, respectively. In terms of knowledge improvement, the posttest scores of the proposed intelligent intervention group increased by 14.6 points relative to the pretest scores, whereas the random-reminder group achieved an improvement of 9.4 points, and the no-intervention group demonstrated only a 6.8-point increase. Using the no-intervention group as the reference condition, the difference-in-differences net effect of the proposed framework reached 7.2 points (95% confidence interval: 3.8–10.6, $p = 0.002$), while the random-reminder group achieved a net effect of 4.3 points ($p = 0.046$). With respect to effect size, the proposed framework achieved a Cohen's d value of 0.84, whereas the random-reminder group reached 0.51. These findings indicate that intelligent intervention was not only more effective than the absence of intervention, but that posture-aware timing optimization further enhanced intervention effectiveness compared with randomly scheduled reminders.

Table 3. Comparison of edge intelligence-driven intervention effectiveness

Metric	No-Intervention Control Group	Random-Reminder Group	Proposed Intelligent Intervention Group
Average Number of Triggered Reminders per Class	0 (recording only)	5.8 ± 1.1	5.6 ± 1.2
Posture-Transition Rate Within 30 s After Triggering	37.2%	48.5%	71.4%
Micro-Quiz Completion Rate	12.3%	68.2%	84.9%
Micro-Quiz Accuracy	41.5%	59.4%	73.2%
Pretest Knowledge Score	62.1 ± 7.9	61.8 ± 8.2	61.3 ± 8.4
Posttest Knowledge Score	68.9 ± 8.1	71.2 ± 7.6	75.9 ± 7.2
Difference-in-Differences Net Effect δ (Relative to No Intervention)	–	4.3 points ($p = 0.046$)	7.2 points ($p = 0.002$)
Cohen's d (Relative to No Intervention)	–	0.51	0.84

3.4 Performance of the long short-term memory-based behavioral prediction model

Table 4 compares the proposed long short-term memory framework with four baseline models in terms of confusion region prediction and attention regression performance. The baseline models included logistic regression, random forest, extreme gradient boosting, and a single-layer simple recurrent neural network. Data collected during the first three weeks were utilized for model training, whereas data from the final week were reserved for testing. The proposed long short-term memory framework achieved an area under the curve value of 0.83, outperforming both extreme gradient boosting (0.77) and the simple recurrent neural network model (0.79). The root mean square error for attention regression was reduced to 0.12, representing the best overall performance among all evaluated models. Although the inference latency of the proposed long short-term memory framework reached 24.6 ms, which was higher than the approximately millisecond-level latency observed in tree-based models, the latency remained within an acceptable real-time deployment range for classroom applications. Shapley Additive Explanations (SHAP) analysis revealed that the variance of touch pressure, the rate of change in pitch angle, and screen activation duration constituted the three most influential sensor features, collectively contributing approximately 78% of the overall predictive capability. In the personalized recommendation experiment, students who received augmented reality exercises recommended by the long short-term memory framework achieved a correctness rate of 79.3% in subsequent short-term quizzes, whereas the random recommendation group achieved only 67.1%, corresponding to an improvement of 12.2 percentage points. By comparison, the extreme gradient boosting-based recommendation group achieved a correctness rate of 74.5%. The superior performance of the proposed long short-term memory framework can be attributed to its ability to capture temporal dependencies within multisensory sequences, which is particularly critical for modeling cognitive states such as confusion and attention that exhibit strong sequential memory characteristics.

Table 4. Performance comparison of the long short-term memory-based behavioral prediction model and other methods

Model	Area Under the Curve (Confusion Region)	Root Mean Square Error (Attention)	Inference Latency (ms)
Logistic Regression	0.69 ± 0.04	0.21 ± 0.03	0.8 ± 0.2
Random Forest	0.74 ± 0.03	0.17 ± 0.02	1.2 ± 0.3
Extreme Gradient Boosting	0.77 ± 0.03	0.16 ± 0.02	1.5 ± 0.4
Simple Recurrent Neural Network (Single Layer)	0.79 ± 0.03	0.14 ± 0.02	18.3 ± 2.5
Proposed Long Short-Term Memory Framework	0.83 ± 0.02	0.12 ± 0.02	24.6 ± 3.1

3.5 Cross-scenario network switching performance

Table 5 compares the performance of single-path transmission control protocol, round-robin scheduling, deep Q-network-based scheduling, and the proposed MPTCP combined with Q-learning under classroom-to-outdoor mobile learning scenarios. During the experiments, students traversed a predefined 300 m route encompassing three distinct network environments: classroom Wi-Fi, outdoor 5G coverage, and basement 4G signal regions. The proposed framework achieved an average packet-loss rate of only 0.4%, which was substantially lower than the 1.8% observed in the deep Q-network-based scheduling strategy and the 3.2% recorded for round-robin scheduling. Furthermore, the median session reconstruction latency of the proposed framework was reduced to 46 ms, significantly outperforming the 180 ms latency of the deep Q-network-based method. The average number of retransmission timeout events was also minimized to 0.1 occurrences. In terms of path utilization distribution, 72% of the total traffic in the proposed framework was allocated to the 5G network, whereas only 58% was allocated by the deep Q-network-based strategy. This finding suggests that the reward function employed in the proposed framework achieved a more effective balance between bandwidth utilization and energy consumption. Under a simulated communication channel with a packet-loss rate of 5%, the Reed–Solomon forward error correction mechanism achieved a recovery rate of 99.2%. Collectively, the proposed framework demonstrated substantial advantages over all comparison methods in terms of the three critical indicators of packet-loss rate, session reconstruction latency, and retransmission timeout frequency.

Table 5. Comparison of cross-scenario network switching performance

Metric	Single-Path Transmission Control Protocol	Round-Robin Scheduling	Deep Q-Network-Based Scheduling	Proposed Framework (MPTCP + Q-Learning)
Average Packet-Loss Rate	8.7%	3.2%	1.8%	0.4%
Median Session Reconstruction Latency (ms)	2300	420	180	46
95th Percentile Session Reconstruction Latency (ms)	4100	890	380	115
Average Number of Retransmission Timeout Events	8.4	2.3	1.2	0.1
Path Utilization Ratio (Wi-Fi/5G/4G)	Single path only	43/37/20%	35/58/7%	22/72/6%
Forward Error Correction Recovery Rate (Under 5% Packet Loss)	–	–	–	99.2%

To further clarify the key behavioral features driving the decision-making process of the behavioral prediction model and to investigate their underlying mechanisms within the context of mobile interactive technology-supported blended learning in higher education, SHAP feature-importance analysis was conducted on the trained long short-term memory model.

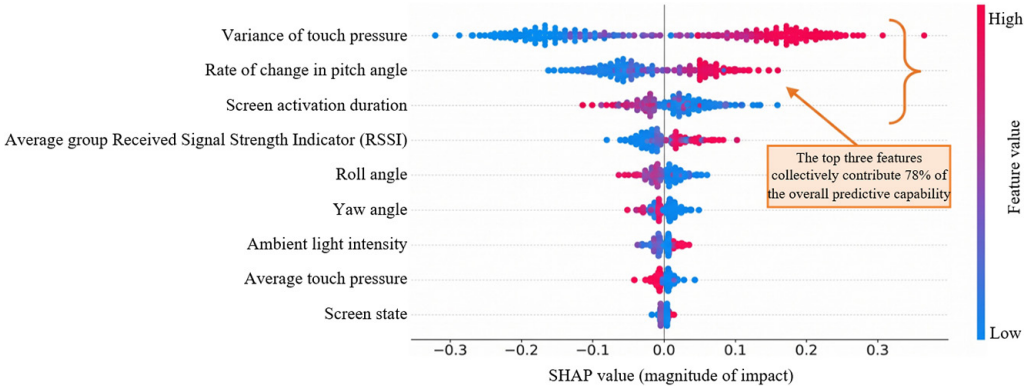

Fig. 3. Summary plot of SHAP feature importance for the long short-term memory-based behavioral prediction model

The results presented in Figure 3 indicate that the predictive capability of the model exhibited substantial feature dependency. The SHAP summary plot clearly demonstrates that the three most influential features collectively contributed approximately 78% of the overall predictive capability of the model. Specifically, the variance of touch pressure exhibited the highest global importance. The distribution of SHAP values further revealed that higher feature values corresponded to stronger positive contributions, indicating that high-frequency physical interaction with the mobile device constituted a critical behavioral indicator for prediction. The rate of change in pitch angle and screen activation duration ranked second and third, respectively, suggesting that dynamic posture adjustment during mobile-device usage and sustained cognitive engagement played significant corrective roles in the predictive process of the model. By contrast, environmental factors such as ambient light intensity, as well as state features, including screen states, demonstrated relatively limited influence on model predictions.

4 CONCLUSION

A mobile interaction technology-driven blended learning model for higher education was proposed, with core contributions including a device–edge–cloud collaborative federated learning architecture, a posture-recognition pipeline integrating complementary filtering and MobileNetV3, Bluetooth RSSI ranging combined with Kalman filtering, a long short-term memory-based behavioral prediction model, and an MPTCP framework with Q-learning-based path scheduling. The results of five groups of comparative experiments demonstrated that the posture-recognition accuracy reached 91.3%, while the ranging error was reduced to 0.8 m. Through edge intelligence-driven intervention, the posture-transition rate following attention loss was increased to 71.4%. In addition, the confusion-region prediction model achieved an area under the curve value of 0.83, whereas the cross-network packet-loss rate remained below 0.5%, with a median session reconstruction latency of only 46 ms.

A randomized controlled teaching experiment involving 62 undergraduate students further demonstrated that the complete framework increased learning engagement by 30% and improved final examination scores by 12.5 points, with an effect size of Cohen's d = 0.71. The findings demonstrate that a mobile interactive framework driven entirely by native smartphone sensors can achieve substantial advantages in both technical performance and instructional effectiveness. Consequently, this study provides a practical and scalable technical solution for blended learning for ubiquitous higher education environments.

5 ACKNOWLEDGEMENTS

This paper was funded by the 2024 Annual General Project of Education Science Planning in Henan Province (Grant No.: 2024YB0656): Research on AI Technology Empowering Smart Classroom Teaching Models in Higher Vocational Education from the Perspective of New Quality Productivity.

6 REFERENCES

[1] M. Alghizzawi, F. Omeish, T. Abdrabbo, A. Alamro, A. Al Htibat, and M. Abd Ghani, "Future trends of smartphone application intention to use: Expansion of the technology acceptance model," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 18, no. 20, pp. 16–36, 2024. https://doi.org/10.3991/ijim.v18i20.49517

[2] I. Esquivel-Gámez, M. Guerrero-Posadas, J. C. Berthely-Barrios, and J. L. Vázquez-Ariza, "Assessing smartphone addiction among Mexican students: Insights, implications, and interventions in the era of mobile learning and virtual environments," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 18, no. 15, pp. 115–128, 2024. https://doi.org/10.3991/ijim.v18i15.46933

[3] A. Tsai, "A hybrid e-learning model incorporating some of the principal learning theories," *Social Behavior and Personality: An International Journal*, vol. 39, no. 2, pp. 145–152, 2011. https://doi.org/10.2224/sbp.2011.39.2.145

[4] Muthmainnah, S. Siripipatthanakul, E. Apriani, and A. Al Yakin, "Effectiveness of online informal language learning applications in English language teaching: A behavioral perspective," *Education Science and Management*, vol. 1, no. 2, pp. 73–85, 2023. https://doi.org/10.56578/esm010202

[5] S. R. Bartholomew and E. Reeve, "Middle school student perceptions and actual use of mobile devices: Highlighting disconnects in student planned and actual usage of mobile devices in class," *Educational Technology & Society*, vol. 21, no. 1, pp. 48–58, 2018. https://www.jstor.org/stable/26273867

[6] J. Chakraborty, S. Mukherjee, and L. Sahoo, "Intuitionistic fuzzy multi-index multi-criteria decision-making for smart phone selection using similarity measures in a fuzzy environment," *Journal of Industrial Intelligence*, vol. 1, no. 1, pp. 1–7, 2023. https://doi.org/10.56578/jii010101

[7] L. R. Mehta, M. S. Borse, M. Tepan, and J. Shah, "Identifying suitable deep learning approaches for dental caries detection using smartphone imaging," *International Journal of Computational Methods and Experimental Measurements*, vol. 12, no. 3, pp. 251–267, 2024. https://doi.org/10.18280/ijcmem.120306

[8] J. Qu, Y. Zhao, and Y. Xie, "Artificial intelligence leads the reform of education models," *Systems Research and Behavioral Science*, vol. 39, no. 3, pp. 581–588, 2022. https://doi.org/10.1002/sres.2864

[9] X. H. Zheng, "Application of distance education combined with artificial intelligence," *Agro Food Industry Hi-Tech*, vol. 28, no. 1, pp. 555–559, 2017.

[10] J. Chen, X. D. Gao, J. Rong, and X. Gao, "A situation awareness assessment method based on fuzzy cognitive maps," *Journal of Systems Engineering and Electronics*, vol. 33, no. 5, pp. 1108–1122, 2022. https://doi.org/10.23919/JSEE.2022.000108

[11] Y. N. K. Chen and C. H. R. Wen, "Taiwanese university students' smartphone use and the privacy paradox," *Comunicar*, vol. 27, no. 60, pp. 61–70, 2019. https://doi.org/10.3916/C60-2019-06

[12] C. Mutimukwe, O. Viberg, C. McGrath, and T. Cerratto-Pargman, "Privacy in online proctoring systems in higher education: Stakeholders' perceptions, awareness and responsibility," *Journal of Computing in Higher Education*, vol. 1, no. 1, pp. 1–30, 2025. https://doi.org/10.1007/s12528-025-09461-5

[13] M. R. Mirsaleh and M. R. Meybodi, "Assignment of cells to switches in cellular mobile network: A learning automata-based memetic algorithm," *Applied Intelligence*, vol. 48, no. 10, pp. 3231–3247, 2018. https://doi.org/10.1007/s10489-018-1136-z

[14] M. Klinkowski, J. Perelló, and D. Careglio, "Machine learning-aided latency prediction in packet-switched Xhaul networks," *IEEE Access*, vol. 14, no. 1, pp. 48072–48088, 2026. https://doi.org/10.1109/ACCESS.2026.3678383

[15] C. Looi, P. Seow, B. Zhang, H. So, W. Chen, and L. Wong, "Leveraging mobile technology for sustainable seamless learning: A research agenda," *British Journal of Educational Technology*, vol. 41, no. 2, pp. 154–169, 2010. https://doi.org/10.1111/j.1467-8535.2008.00912.x

7 AUTHOR

Aihua Gui graduated from Henan University in 2004, works at School of Social Services and Management, Zhumadian Preschool Education College and is engaged in the research of English language teaching and the research of educational and teaching methods and strategies (E-mail: flower7098@163.com).