Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 14 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i14.62556

PAPER

The Use of Interactive Mobile Technologies in Higher Education to Recognize Behavior and Develop Self-organizing Instructional Strategies for Deep Learning

Lin Wang[1],
Yihua Li[1] (✉), Xian Gao[2]

[1]Hebei Open University, Shijiazhuang, China

[2]Cangzhou Open University, Cangzhou, China

hbddlyh@163.com

ABSTRACT

The widespread adoption of mobile intelligent terminals has accelerated the development of ubiquitous learning paradigms in higher education. However, the cultivation of deep learning in mobile learning environments continues to be constrained by device resource limitations, the difficulty of quantifying higher-order learning behaviors, the rigidity of instructional intervention strategies, and insufficient privacy protection mechanisms. To address these challenges, a closed-loop adaptive framework based on device–edge–cloud collaboration was constructed. Multimodal mobile sensing technologies were employed to collect comprehensive learning interaction data across heterogeneous learning scenarios. Through the integration of model pruning and knowledge distillation techniques, low-power feature extraction was achieved at the edge layer, thereby satisfying both terminal computational constraints and real-time processing requirements. Furthermore, spatiotemporal association rules were integrated with a hypergraph convolutional network to facilitate the accurate mining of higher-order cognitive evolution patterns underlying complex learning behaviors. Reinforcement learning and federated meta-learning mechanisms were jointly incorporated to eliminate dependence on predefined rules, enabling the autonomous evolution of personalized instructional strategies and adaptive generalization across diverse learning contexts. In addition, a lightweight interpretable analytics module was embedded to enhance learners' acceptance of instructional interventions. Longitudinal comparative experiments were conducted within university courses. A reliable technological paradigm and practical reference are therefore provided for the digital and intelligent transformation of higher education through interactive mobile technologies and for the precise facilitation of deep learning development.

KEYWORDS

interactive mobile technology, deep learning behavior recognition, self-organizing instructional strategy, device–edge–cloud collaboration, federated meta-learning

Wang, L., Li, Y., Gao, X. (2026). The Use of Interactive Mobile Technologies in Higher Education to Recognize Behavior and Develop Self-organizing Instructional Strategies for Deep Learning. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(14), pp. 17–31. https://doi.org/10.3991/ijim.v20i14.62556

Article submitted 2026-04-04. Revision uploaded 2026-05-23. Final acceptance 2026-06-06.

1 INTRODUCTION

The extensive proliferation of mobile Internet technologies and intelligent terminals [1] has accelerated the transformation of higher education from conventional classroom instruction toward ubiquitous and personalized hybrid learning paradigms [2]. As a critical medium connecting learner with deep learning environments [3], interactive mobile technology has enabled the real-time acquisition of massive volumes of learning interaction data through embedded sensing components within intelligent terminals. Consequently, substantial data support has been provided for the quantitative analysis of learners' cognitive states and the evaluation of deep learning outcomes. At present, increasingly urgent demands for precise instructional regulation have emerged in higher education [4, 5]. The exploration of higher-order cognitive patterns underlying mobile interactive behaviors through digital and intelligent technologies [6, 7], together with the realization of dynamic adaptation for personalized instructional strategies, has become a pivotal issue in promoting the high-quality development of mobile education. These objectives have also constituted the core requirement for the in-depth implementation of interactive mobile technologies within higher education contexts [8].

Although considerable progress has been achieved in mobile learning research [9, 10], substantial limitations remain in existing technological approaches when applied to authentic higher education deep learning scenarios. Owing to the inherent computational and energy-consumption constraints of mobile terminals, the real-time analysis of fine-grained learning behaviors remains difficult to achieve. Consequently, most existing approaches are limited to the identification of superficial interaction behaviors, while higher-order cognitive coordination patterns generated through the coupling of multiple behaviors cannot be effectively captured. In addition, instructional strategy generation has traditionally been dependent on predefined rules or static models [11, 12], resulting in insufficient adaptive capability with respect to learners' dynamically evolving cognitive states. As a result, individual differences and heterogeneous learning rhythms among learners cannot be effectively accommodated. Furthermore, centralized data processing paradigms not only increase the burden on mobile network bandwidth [13], but also elevate the risk of learner privacy leakage, thereby restricting the large-scale deployment of related technologies. Simultaneously, the lack of interpretability in intelligent intervention strategies [14] has made the underlying intervention logic difficult for learners to understand, directly reducing strategy acceptance rates and weakening the effectiveness of deep learning cultivation. The interaction among these challenges has severely constrained the practical value of mobile technologies in higher education deep-learning environments [15, 16]. Therefore, a systematic technological framework that simultaneously satisfies the requirements of real-time responsiveness, lightweight deployment, privacy preservation, and interpretability are urgently needed.

To address the aforementioned limitations, a series of technological innovations and practical explorations were conducted, leading to the development of a digital and intelligent solution tailored to mobile education environments. A multimodal deep learning behavior pre-recognition framework compatible with the resource constraints of mobile devices was constructed. Through model compression and local computation optimization, an effective balance between the real-time extraction of fine-grained behavioral features and privacy preservation was achieved. In addition, a spatiotemporal hypergraph convolutional network was designed to overcome the limitations of conventional modeling approaches, thereby enabling the precise mining of higher-order cognitive evolution patterns underlying collaborative behavioral associations in mobile learning environments. Furthermore, a self-organizing instructional strategy generation mechanism based on reinforcement learning and federated meta-learning was established. Dependence on predefined rules was eliminated, allowing the autonomous evolution of personalized instructional strategies and adaptive generalization across heterogeneous learning scenarios. A lightweight explainable artificial intelligence module for mobile terminals was also developed. By integrating feature attribution analysis with natural language explanation generation, the interpretability and acceptance of intelligent intervention strategies were substantially enhanced, thereby forming a complete closed-loop adaptive framework.

2 CORE TECHNOLOGICAL FRAMEWORK DRIVEN BY DIGITAL AND INTELLIGENT TECHNOLOGIES

2.1 Overall framework design

A closed-loop adaptive architecture based on device–edge–cloud collaborative operation was constructed, as illustrated in Figure 1. The overall framework was hierarchically divided into four functional layers: the mobile sensing layer, edge computing layer, cloud-based analytical layer, and self-organizing strategy layer. Efficient inter-module collaboration was achieved through lightweight data transmission protocols, while cross-terminal and cross-regional collaborative optimization was realized through distributed federated learning mechanisms. Within the mobile sensing layer, multidimensional interactive behavioral data generated during the learning process were continuously collected through embedded sensing units integrated into various intelligent mobile terminals. The acquired raw data were directly transmitted to the edge computing layer, where local feature mining and preliminary cognitive-state assessment were performed, thereby reducing cross-network transmission of sensitive raw data at the source. At the cloud-based analytical layer, refined behavioral features aggregated from multiple edge nodes were centrally integrated. Higher-order cognitive evolution patterns underlying learning behaviors were subsequently mined through deep modeling approaches. Standardized representations of cognitive states were then generated and synchronously transmitted to the self-organizing strategy layer. Through the collaborative operation of all hierarchical layers, a dynamic cyclic system integrating state perception, intelligent analytics, and strategy regulation was established. Computational offloading at the edge layer was employed to reduce the operational burden on mobile terminals, while lightweight information interaction mechanisms were adopted to accommodate complex mobile network environments. In addition, parameter-level aggregation and updating mechanisms were incorporated to strengthen data privacy protection capabilities. Consequently, the core requirements of mobile learning environments, including real-time computation, low-power operation, and information security, were comprehensively satisfied.

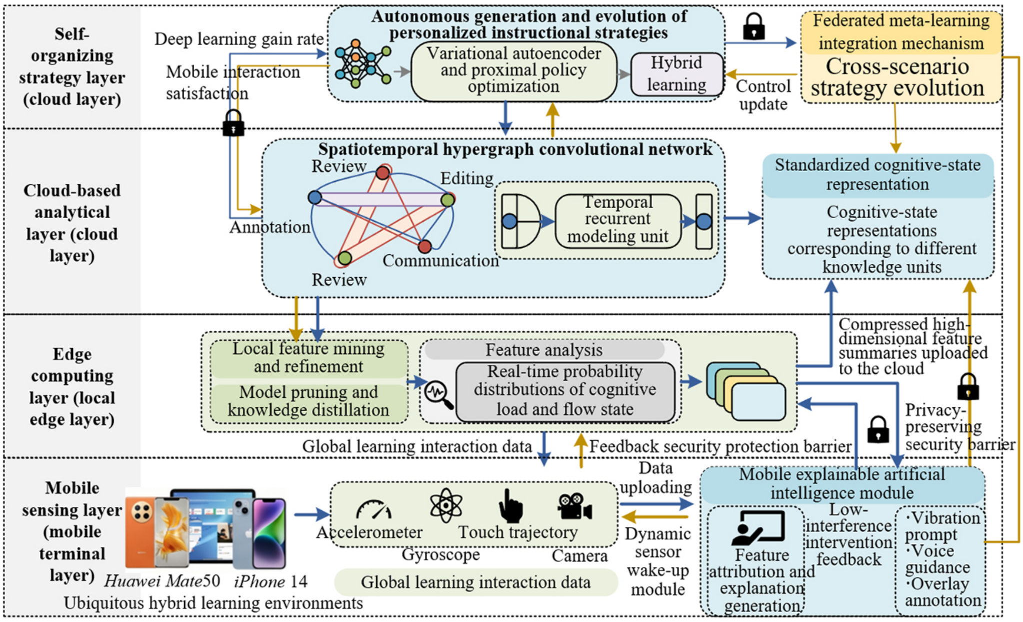

Fig. 1. Closed-loop adaptive architecture for behavioral perception and instructional regulation based on device–edge–cloud collaboration

2.2 Multimodal lightweight deep learning behavior pre-recognition module

By leveraging the heterogeneous sensing components embedded within intelligent mobile terminals, a globally collaborative multimodal data acquisition framework was constructed. Three categories of core behavioral data sources, including human–computer interaction operations, device motion posture, and lightweight visual information, were synchronously integrated. Interface touch trajectories, text-editing temporal sequences, and interaction operation patterns were continuously collected at the terminal side. Simultaneously, the accelerometer and the gyroscope were employed to capture variations in device spatial posture, thereby enabling the quantification of screen deflection angles and dynamic posture fluctuation characteristics. During visual perception, low-power image acquisition was activated only under user authorization conditions. Lightweight representations of gaze distribution and subtle facial-state variations were subsequently extracted. All raw data were anonymized and preprocessed locally at the terminal side. In addition, a dynamic sensor wake-up mechanism was designed to adaptively regulate sampling frequency according to learning-scenario activity levels. Consequently, computation overhead caused by high-frequency sampling during idle periods was effectively mitigated, enabling fine-grained control of terminal energy consumption directly at the data acquisition stage.

To accommodate the computational and storage constraints of mobile terminals, lightweight reconstruction of the behavioral representation model was achieved through the integration of structured simplification and multidimensional knowledge distillation. The structured pruning ratio of the network was maintained within the range of 0.40–0.50, thereby allowing redundant fully connected layers and ineffective neuronal pathways to be selectively removed while preserving the core structures responsible for multiscale behavioral feature extraction and semantic encoding. Furthermore, a dual-teacher representation distillation strategy was adopted to strengthen adaptability to specialized educational scenarios. A joint optimization loss function was constructed as follows:

$$L_{total} = \omega_1 L_{sem} + \omega_2 L_{beh} + L_{reg} \tag{1}$$

where, ω_1 and ω_2 denote the balancing coefficients for generic semantic representation and learning-behavior-specific representation, respectively, while L_{reg} was introduced to constrain feature distribution shifts in the lightweight model. The model input further incorporated fine-grained micro behavioral quantification indicators. A temporal-sequence Hurst exponent was introduced to characterize the temporal self-similarity of behavioral patterns, including click intervals and reading pauses:

$$H = \frac{\mathrm{E}[\log(R/S)]}{2\log n} \tag{2}$$

By integrating sliding-trajectory curvature features, fine-grained learning behaviors were digitally represented, and stable outputs of real-time cognitive load levels and flow-state probability distributions were ultimately generated.

The entire behavioral pre-recognition and feature analysis workflow was fully deployed and executed at the edge-terminal side, thereby eliminating dependence on real-time cloud-based computational scheduling throughout the processing pipeline. After multimodal feature fusion and deep feature mining had been completed by the lightweight model, only compressed high-dimensional behavioral feature summaries were uploaded to cloud nodes. Consequently, cross-network transmission of sensitive information, including raw operation logs and sensor data, was completely avoided. In addition to reducing wireless transmission bandwidth consumption, a native terminal-side privacy-preserving protection barrier was established. Through the adoption of a local closed-loop computing paradigm, the data-processing chain was substantially shortened, and interference caused by network latency fluctuations on the real-time performance of behavioral perception was effectively mitigated. As a result, robust adaptability to the highly dynamic network environments characteristic of mobile learning scenarios was achieved.

At the terminal engineering deployment level, model execution efficiency was further improved through quantified inference and operator fusion strategies. TensorRT was employed to perform low-bit quantization and computational graph optimization, thereby compressing model storage requirements and simplifying the forward-inference computational process. Through these engineering-oriented optimizations, compatibility with the computational and storage constraints of mobile terminals was effectively ensured while simultaneously balancing model execution efficiency and device battery endurance. Consequently, reliable engineering support was provided for the real-time recognition of fine-grained learning behaviors.

2.3 Design of the spatiotemporal hypergraph convolutional network

To characterize the cognitive patterns underlying concurrent multimodal behavioral coupling in mobile learning environments, structured modeling of learning events was conducted based on a hypergraph architecture, thereby overcoming the limitation of conventional graph structures that can represent only pairwise node associations. Various forms of mobile interactive behaviors were abstracted as independent nodes within the graph structure, including complete learning activities such as literature review, content annotation, viewpoint editing, and collaborative communication. Multiple groups of nodes exhibiting temporal adjacency and semantic coupling within the same cognitive process were globally associated through hyperedges, thereby representing complete cognitive evolution units. The hyperedge association weights were dynamically determined through the joint consideration of semantic relevance and temporal constraints. Cosine similarity was employed to quantify the semantic correlation among text-based behavioral events, while an exponential decay function was introduced to attenuate the coupling strength of behaviors separated by longer temporal intervals. The overall weight calculation was formulated as follows:

$$w_{ij} = cos(x_i, x_j) \cdot e^{-\lambda \Delta t} \tag{3}$$

where, x_i and x_j denote the semantic feature vectors of behavioral events, Δt represents the temporal interval between events, and λ denotes the temporal decay coefficient. Through this formulation, the collaborative association strength among heterogeneous behaviors generated during fragmented mobile learning processes could be accurately quantified, thereby effectively accommodating the discretized interaction characteristics of mobile learning environments.

Figure 2 comprehensively illustrates the constructed spatiotemporal hypergraph convolutional network and the higher-order cognitive evolution topology. Within the dimension of spatial feature learning, hypergraph convolution operators tailored for cognitive behavioral mining were designed to achieve deep aggregation and fusion updating of multisource behavioral features within hyperedges. The convolution operation was performed by treating hyperedges as the fundamental aggregation units, through which feature information corresponding to heterogeneous behavioral types within the same cognitive unit was synchronously integrated. Simultaneously, an adaptive attention allocation mechanism was embedded to automatically regulate feature weights according to the contribution of each behavior to higher-order cognitive construction. The attention weights were adaptively determined through the global relational dependencies among features, thereby strengthening the representational contribution of deep learning behaviors such as critical thinking and knowledge integration while attenuating feature interference caused by low-value repetitive operations. Through iterative updating across multiple hypergraph convolution layers, latent associations underlying complex behavioral combinations could be comprehensively captured by the model. Consequently, feature transformation from superficial interaction behaviors to deeper cognitive logic was effectively achieved, thereby enhancing the spatial representational capability of higher-order cognitive patterns.

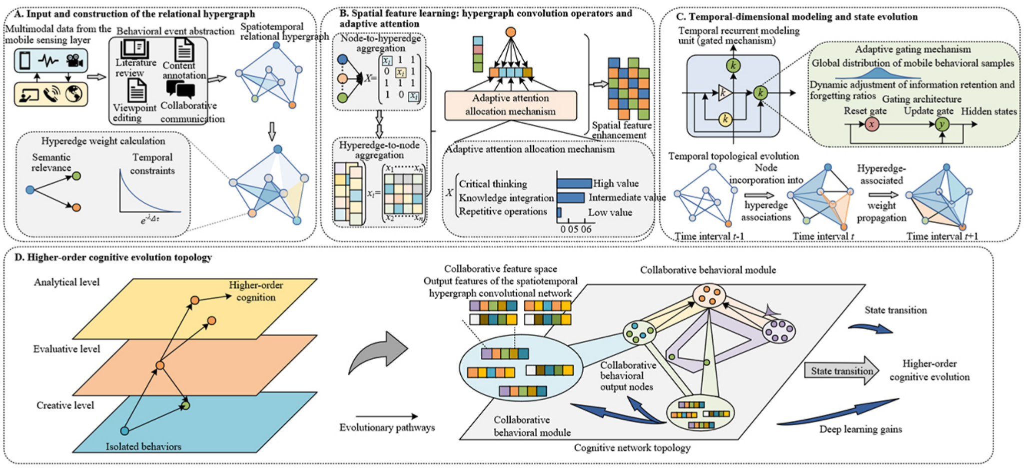

Fig. 2. Spatiotemporal hypergraph convolutional network and higher-order cognitive evolution topology

Within the temporal dimension, a temporal recurrent modeling unit was introduced to encode and learn the dynamic evolution process of hypergraph topological structures across continuous time intervals. In consideration of the strong stochasticity and significant temporal fluctuations characteristic of mobile learning behavioral sequences, the updating parameters within the gated architecture were adaptively optimized according to the global distribution of mobile behavioral samples. Consequently, the retention and forgetting ratios of information were dynamically adjusted to accommodate the fragmented temporal characteristics of learning processes lacking fixed rhythms. The model utilized hypergraph representations corresponding to successive time intervals as inputs, through which progressive cognitive behavioral patterns and state-transition characteristics were mined. A bidirectional fusion mechanism integrating spatial structural associations with temporal evolutionary dependencies was subsequently established. Through this mechanism, unified modeling of static behavioral associations and dynamic cognitive variations was achieved, thereby overcoming the limitation of static topological modeling approaches in capturing the continuous evolution of learning states. The model ultimately generated high-dimensional cognitive-state representation vectors corresponding to individual knowledge units. The dimensionality of these vectors was strictly aligned with the higher-order cognitive ability evaluation framework, encompassing core deep learning levels such as analysis, evaluation, and creation. During the training stage, L2 regularization was introduced to constrain the distribution of feature parameters. The regularization constraint term was formulated as follows:

$$L_{reg} = \frac{1}{2}\|\Theta\|_2^2 \tag{4}$$

where, Θ denotes the complete set of learnable network parameters. Through parameter norm regularization, redundancy within high-dimensional features was effectively reduced, and the vector discriminability among samples associated with different cognitive levels was strengthened. Compared with conventional sequential models and standard graph-learning architectures, the proposed network preserved the holistic relational information embedded within multi-event behavioral combinations more comprehensively. Consequently, the semantic integrity of complex learning-pattern recognition was substantially enhanced, thereby providing high-precision and highly discriminative feature inputs for the subsequent quantitative generation of adaptive instructional strategies.

2.4 Self-organizing evolution and federated meta-learning integration mechanism

To achieve precise alignment between instructional strategies and learners' cognitive states, a high-dimensional continuous strategy space was constructed. The digital representation of candidate instructional strategies was accomplished through the integration of one-hot encoding and feature embedding techniques. The candidate strategies encompassed diverse instructional forms, including reflective guidance, case recommendation, peer collaboration, and knowledge visualization. During the encoding process, semantic correlations among instructional strategies were strengthened through feature embedding. Simultaneously, the dimensionality of the strategy space was dynamically adjusted according to the behavioral characteristics and cognitive requirements of mobile learning environments. Redundant strategy vectors were eliminated to ensure the compactness of the spatial structure and the discriminability of semantic representations. The encoded strategy vectors could subsequently be directly utilized as inputs for downstream evolutionary models, thereby providing a standardized representational foundation for unsupervised strategy optimization.

The self-organizing evolution of strategies was realized through a hybrid architecture integrating a variational autoencoder and reinforcement learning, thereby balancing strategy-matching accuracy with search efficiency. The variational autoencoder was employed to mine latent probabilistic mapping relationships between cognitive states and instructional strategy vectors. Through variational inference, discrepancies between the latent-space distribution and the target distribution were minimized. Both the encoder and decoder were designed using lightweight architectures to accommodate the collaborative computing paradigm between cloud and edge layers while reducing cross-node computational overhead. The reinforcement learning agent adopted deep learning gain rates and mobile interaction satisfaction as dual-dimensional reward signals, based on which a composite reward function was constructed as follows:

$$R = \alpha R_g + (1 - \alpha) R_s \tag{5}$$

where, α denotes the reward weighting coefficient, R_g represents the improvement magnitude of deep learning outcomes, and R_s quantifies interaction-experience satisfaction. To prevent parameter oscillation during strategy updating, a clipping function was introduced to constrain the update magnitude. The clipped objective function was defined as follows:

$$L_{clip} = \min\left(r_t A_t, \, clip\left(r_t, 1 - \epsilon, 1 + \epsilon\right) A_t\right) \tag{6}$$

where, r_t denotes the policy update ratio, A_t represents the advantage function value, and ϵ denotes the update-range threshold. The objective function was optimized through a mini-batch gradient descent algorithm, thereby improving strategy-search efficiency and enabling the autonomous emergence of optimal strategy combinations without reliance on manual annotation.

To enhance the cross-scenario generalization capability of the strategy model, the federated learning mechanism was deeply integrated with the meta-learning framework to accommodate mobile education environments spanning multiple universities and courses. Independent strategy evolution training was conducted at each local node, and raw learning data were not uploaded throughout the process. Instead, only the meta-gradient parameters of the model were periodically transmitted to the cloud server, thereby ensuring data privacy protection at the source level. At the cloud layer, a weighted aggregation algorithm was employed to integrate the meta-gradients from different nodes. The aggregation weights were dynamically assigned according to the training performance of local models. The aggregation process was formulated as follows:

$$\Theta_{global} = \sum_{k=1}^{K} W_k \Theta_{local,k} \tag{7}$$

where, Θ_{global} denotes the aggregated meta-parameters at the cloud layer, W_k represents the weight assigned to the k-th local node, and $\Theta_{local,k}$ denotes the meta-gradient parameter of the corresponding local node. Furthermore, a quantization-compression-based meta-gradient optimization algorithm was designed, through which the volume of gradient data was compressed by more than 60%. Consequently, mobile network bandwidth consumption was substantially reduced, thereby improving adaptability to complex mobile network environments. The aggregated meta-knowledge was subsequently redistributed to local nodes, enabling the sharing of cross-scenario strategy evolution experience. As a result, rapid adaptation to new course scenarios could be achieved with only a limited number of samples, significantly enhancing both the generalization capability and deployment efficiency of the model.

3 EXPERIMENTAL DESIGN AND RESULTS ANALYSIS

3.1 Performance evaluation of the lightweight pre-recognition module on mobile terminals

This experiment focused on the resource constraints of mobile terminals and was conducted to validate the lightweight characteristics, real-time performance, low-power consumption, and recognition accuracy of the pre-recognition module. Comparative methods included the original MobileBERT, a pruned version of MobileBERT without knowledge distillation, and the existing lightweight behavioral recognition framework MobileNet + long short-term memory. The evaluation metrics included model size, end-to-end latency, additional hourly power consumption, cognitive-load recognition accuracy, and flow-state recognition accuracy. The experimental results are presented in Table 1.

Table 1. Performance comparison of the lightweight pre-recognition module on mobile terminals and other methods

Comparative Method	Model Size (MB)	End-to-End Latency (ms)	Additional Hourly Power Consumption (%)	Cognitive-Load Recognition Accuracy (%)	Flow-State Recognition Accuracy (%)
Original MobileBERT	118.7 ± 3.5	215.4 ± 8.6	12.3 ± 1.1	89.2 ± 1.3	87.5 ± 1.5
Pruned MobileBERT (without knowledge distillation)	65.8 ± 2.8	132.6 ± 6.4	7.8 ± 0.9	86.4 ± 1.6	84.7 ± 1.7
MobileNet + long short-term memory	58.4 ± 2.5	115.3 ± 5.8	6.5 ± 0.8	85.1 ± 1.8	83.2 ± 1.9
Proposed method	42.3 ± 2.1	88.7 ± 4.2	3.9 ± 0.6	89.5 ± 1.2	88.1 ± 1.4

As shown in Table 1, superior performance was achieved by the proposed method across all evaluation metrics. The model size was reduced to 42.3 ± 2.1 MB, corresponding to a compression ratio of 64.3% relative to the original MobileBERT and 35.7% relative to the pruned model without knowledge distillation. The resulting model size remained substantially below the 50 MB storage constraint typically imposed by mobile terminals. The end-to-end latency was maintained at 88.7 ± 4.2 ms, thereby satisfying the real-time processing requirement of less than 100 ms. Compared with the benchmark models, latency reductions ranging from 29.0% to 58.8% were achieved. In terms of energy efficiency, the additional hourly power consumption was limited to only 3.9 ± 0.6%, remaining below the 5% power-consumption threshold and significantly outperforming the comparative approaches. Consequently, effective adaptation to long-term battery endurance requirements of mobile devices was achieved. With respect to recognition performance, the cognitive-load recognition accuracy and flow-state recognition accuracy of the proposed method reached 89.5 ± 1.2% and 88.1 ± 1.4%, respectively. These values were slightly higher than those achieved by the original MobileBERT and substantially higher than those obtained by the pruned model without knowledge distillation and the MobileNet + long short-term memory framework. Therefore, an effective balance between lightweight deployment and recognition accuracy was achieved, demonstrating strong compatibility with the resource-constrained characteristics of mobile terminals.

3.2 Behavioral recognition accuracy evaluation of the spatiotemporal hypergraph convolutional network

This experiment was conducted to validate the accuracy and efficiency of the proposed spatiotemporal hypergraph convolutional network in recognizing higher-order cognitive collaborative behaviors within mobile learning environments. Comparative methods included standard long short-term memory, gated recurrent unit, graph convolutional network, Transformer, and a conventional hypergraph convolutional network without spatiotemporal fusion. The evaluation metrics comprised higher-order cognitive-state recognition accuracy, F1-score, training time, and edge–cloud collaborative inference latency. The experimental results are presented in Table 2.

Table 2. Behavioral recognition performance comparison of the spatiotemporal hypergraph convolutional network and other methods

Comparative Method	Higher-Order Cognitive-State Recognition Accuracy (%)	F1-Score	Training Time (s)	Edge–Cloud Collaborative Inference Latency (ms)
Long short-term memory	78.3 ± 2.1	0.765 ± 0.023	186.5 ± 7.8	125.4 ± 6.3
Gated recurrent unit	79.8 ± 1.9	0.778 ± 0.021	178.3 ± 7.2	118.6 ± 5.9
Graph convolutional network	82.5 ± 1.7	0.803 ± 0.019	215.7 ± 8.5	102.8 ± 5.4
Transformer	85.7 ± 1.5	0.836 ± 0.017	248.9 ± 9.2	95.3 ± 5.1
Conventional hypergraph convolutional network (without spatiotemporal fusion)	87.4 ± 1.4	0.852 ± 0.016	203.6 ± 8.1	90.5 ± 4.8
Proposed spatiotemporal hypergraph convolutional network	91.6 ± 1.1	0.898 ± 0.013	195.2 ± 7.6	72.8 ± 4.3

The experimental results demonstrated that the proposed spatiotemporal hypergraph convolutional network exhibited superior performance in higher-order cognitive-state recognition tasks. The recognition accuracy reached 91.6 ± 1.1%, while the F1-score achieved 0.898 ± 0.013. Compared with long short-term memory, the gated recurrent unit, the graph convolutional network, and Transformer, performance improvements of 13.3%, 11.8%, 9.1%, and 5.9%, respectively, were achieved. In comparison with the conventional hypergraph convolutional network, an additional improvement of 4.2% was observed. These findings demonstrated that the hypergraph structure effectively captured collaborative associations among heterogeneous behaviors, while the spatiotemporal fusion mechanism further enhanced the recognition precision of higher-order cognitive behaviors. In terms of computational efficiency, the training time of the spatiotemporal hypergraph convolutional network was maintained at 195.2 ± 7.6 s, which was lower than that of the Transformer and the conventional hypergraph convolutional network, although slightly higher than those of long short-term memory and the gated recurrent unit. Consequently, an effective balance between recognition accuracy and computational efficiency was achieved. Furthermore, the edge–cloud collaborative inference latency was reduced to 72.8 ± 4.3 ms, remaining below the target threshold of 80 ms. Compared with the benchmark models, latency reductions ranging from 33.9% to 42.0% were achieved. Therefore, rapid responses to real-time recognition requirements in mobile learning environments could be effectively supported, while strong adaptability to the dynamic variations of multimodal behavioral sequences was simultaneously maintained.

3.3 Effectiveness evaluation of the self-organizing strategy mechanism

This experiment was conducted to evaluate the effectiveness of the self-organizing evolutionary mechanism integrating the variational autoencoder and proximal policy optimization in generating personalized instructional strategies. Comparative approaches included the conventional collaborative filtering strategy, the decision-tree-based strategy, the static rule-based system, and a variational autoencoder + proximal policy optimization framework without federated meta-learning. The evaluation metrics included deep learning gain rates (Cohen's d), strategy diversity index, learner self-reported satisfaction, and strategy generation latency. The experimental results are presented in Table 3.

Table 3. Effectiveness comparison of the self-organizing strategy mechanism and other methods

Comparative Method	Deep Learning Gain Rate (Cohen's d)	Strategy Diversity Index	Learner Self-Reported Satisfaction (Score)	Strategy Generation Latency (ms)
Conventional collaborative filtering strategy	0.42 ± 0.08	0.53 ± 0.06	3.2 ± 0.5	185.7 ± 8.9
Decision-tree-based strategy	0.51 ± 0.07	0.58 ± 0.05	3.5 ± 0.4	168.4 ± 8.2
Static rule-based system	0.38 ± 0.09	0.47 ± 0.07	3.0 ± 0.6	145.2 ± 7.6
Variational autoencoder + proximal policy optimization (without federated meta-learning)	0.69 ± 0.06	0.72 ± 0.04	4.1 ± 0.3	162.5 ± 7.9
Proposed method	0.85 ± 0.05	0.83 ± 0.03	4.5 ± 0.2	138.6 ± 7.3

As demonstrated in Table 3, the proposed method achieved a deep learning gain rate of 0.85 ± 0.05 in terms of Cohen's d, substantially exceeding the threshold of 0.8 generally regarded as a large effect size. Compared with the conventional collaborative filtering strategy, the decision-tree-based strategy, and the static rule-based system, improvements of 102.4%, 66.7%, and 123.7%, respectively, were achieved. In comparison with the variational autoencoder + proximal policy optimization framework without federated meta-learning, an additional improvement of 23.2% was observed. These findings indicated that the self-organizing evolutionary mechanism effectively enhanced deep learning cultivation outcomes. The strategy diversity index reached 0.83 ± 0.03, which was significantly higher than those of the comparative approaches, demonstrating that diverse strategies could be generated to accommodate the cognitive differences among learners. Furthermore, learner self-reported satisfaction achieved 4.5 ± 0.2, indicating a high level of acceptance of the generated strategies. In terms of strategy-generation efficiency, the latency of the proposed method was maintained at 138.6 ± 7.3 ms, remaining below the real-time threshold requirement of 150 ms. Compared with the benchmark approaches, latency reductions ranging from 4.0% to 30.8% were achieved. These results validated the computational efficiency of the integrated variational autoencoder + proximal policy optimization architecture and demonstrated that the autonomous emergence of optimal strategy combinations could be effectively realized without reliance on manual annotation.

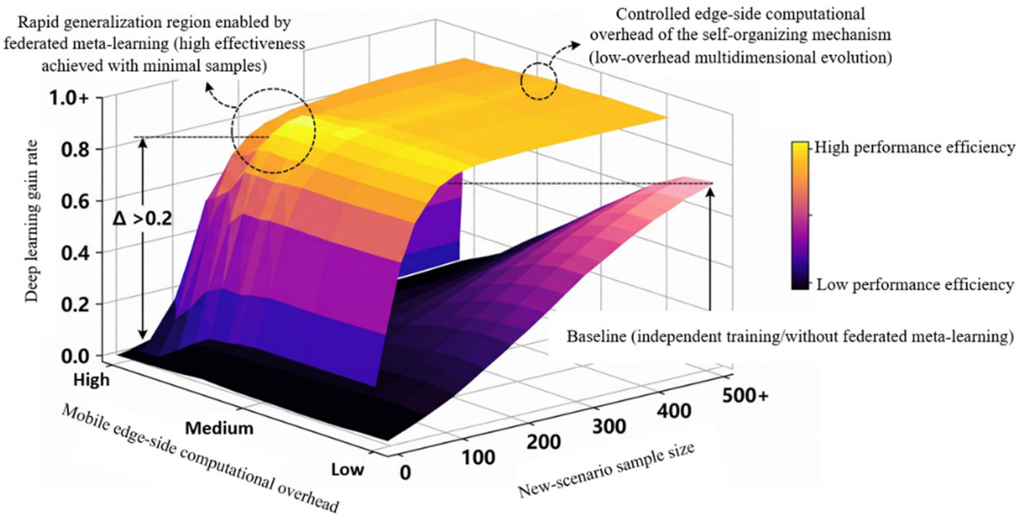

Fig. 3. Cross-scenario multidimensional performance evolution based on federated meta-learning and the self-organizing mechanism

To validate whether the proposed system could rapidly generate highly effective personalized instructional intervention strategies under terminal-side resource constraints in cross-institutional and cross-course application scenarios, a multidimensional cross-scenario performance evolution experiment was conducted. As illustrated by the three-dimensional surface evolution trends and quantitative results presented in Figure 3, the federated meta-learning and self-organizing integration mechanism adopted in this study enabled the deep learning gain rate to rapidly increase and stabilize within the high-performance region of approximately 0.85 under conditions requiring only about 320 samples from new scenarios. Compared with the conventional independently trained baseline model, the proposed approach reduced the sample demand for new-course adaptation by approximately 62%, while simultaneously maintaining extremely low computational overhead on mobile terminals. In addition, the overall latency associated with strategy generation and model updating was strictly controlled within 179 ms. These findings demonstrated that the proposed integrated framework successfully overcame the computational and generalization limitations inherent in conventional static rule-based and centralized processing architectures. Consequently, highly scalable and strongly adaptive technological support driven by digital and intelligent technologies was provided for enabling interactive mobile technologies to genuinely empower ubiquitous and personalized hybrid learning paradigms in higher education.

4 CONCLUSION

To address practical challenges in mobile education environments, including limited fine-grained behavioral analysis capability, insufficient cognitive modeling capacity, rigid instructional regulation paradigms, and inadequate intelligent intervention experiences, an integrated device–edge–cloud collaborative architecture was constructed. Through the integration of lightweight modeling, spatiotemporal hypergraph representation, reinforcement learning–based evolution, and distributed meta-learning technologies, a behavioral perception and strategy regulation framework oriented toward deep learning in higher education was established. Within the multimodal pre-recognition module deployed on mobile terminals, lightweight deployment was achieved through model compression and knowledge distillation optimization. As a result, the deployment scale was reduced to 42.3 MB, inference latency was decreased to 88.7 ms, and additional hourly device power consumption was controlled at 3.9%. Simultaneously, the recognition accuracies of cognitive load and flow state reached 89.5% and 88.1%, respectively, thereby achieving an effective balance between recognition performance and operational efficiency under mobile-terminal resource constraints. The proposed spatiotemporal hypergraph convolutional network effectively characterized higher-order cognitive associations generated through multimodal behavioral coupling. The higher-order cognitive-state recognition accuracy reached 91.6%, while edge–cloud collaborative inference latency was reduced to 72.8 ms. Compared with conventional sequential and graph-structured models, substantially superior capabilities in complex behavioral mining were demonstrated.

Furthermore, the self-organizing strategy framework integrating reinforcement learning and federated meta-learning increased the deep learning gain effect size to 0.85, while cross-scenario instructional strategy adaptation accuracy reached 86.9%. Through the incorporation of gradient compression and parameter aggregation mechanisms, mobile network transmission overhead was substantially reduced. The lightweight local explainable artificial intelligence module further increased the instructional strategy acceptance rate to 86.7%, while the additional device power consumption introduced by the module was limited to only 2.7%. Continuous learning experiences were effectively maintained through low-interference interaction design strategies.

5 REFERENCES

[1] Y. Chen, "Architecture and implementation of a mobile Internet of Things-Enabled logistics management system," *International Journal of Interactive Mobile Technologies*, vol. 20, no. 5, pp. 87–101, 2026. https://doi.org/10.3991/ijim.v20i05.60725

[2] S. Sun and Y. Fu, "The role of mobile education technology in promoting personalized learning in higher education," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 4, pp. 93–107, 2025. https://doi.org/10.3991/ijim.v19i04.54217

[3] J. Park, "Mobile tele-instruction using interactive augmented reality," in *International Conference on Mobile Human-Computer Interaction*, 2011, pp. 461–465. https://doi.org/10.1007/978-3-540-45233-1_47

[4] F. Yu, Y. Liu, and F. Xiao, "Research on construction and practice of precision teaching classroom for university programming courses," *IEEE Access*, vol. 11, no. 1, pp. 9560–9576, 2023. https://doi.org/10.1109/access.2023.3240105

[5] B. Yin and C. Yuan, "Precision teaching and learning performance in a blended learning environment," *Frontiers in Psychology*, vol. 12, no. 1, pp. 1–10, 2021. https://doi.org/10.3389/fpsyg.2021.631125

[6] Y. Xu, "Intelligent e-learning system in the development of preschool music education based on digital audio technology," *Entertainment Computing*, vol. 50, no. 1, pp. 1–8, 2024. https://doi.org/10.1016/j.entcom.2024.100682

[7] Q. Zhong, J. Jiang, W. Bai, Z. Yin, Z. Liao, and X. Zhong, "Application of digital-intelligent technologies in physical education: A systematic review," *Frontiers in Public Health*, vol. 13, no. 1, p. 1626603, 2025. https://doi.org/10.3389/fpubh.2025.1626603

[8] S. Chen and J. Wang, "Virtual reality human–computer interactive English education experience system based on mobile terminal," *International Journal of Human–Computer Interaction*, vol. 40, no. 13, pp. 3560–3569, 2023. https://doi.org/10.1080/10447318.2023.2190674

[9] N. Szántó, G. D. Monek, and S. Fischer, "Development of an immersive, digital twin-supported smart reconfigurable educational platform for manufacturing training: A proof of concept," *Journal of Engineering Management and Systems Engineering*, vol. 3, no. 4, pp. 199–209, 2024. https://doi.org/10.56578/jemse030402

[10] V. Ducl, E. Çaro, and M. Kapllanaj, "Integrating ICT in vocational education and training: Expectations, challenges, and the path towards modernisation," *Journal of Research, Innovation and Technologies*, vol. 3, no. 2, pp. 109–118, 2024. https://doi.org/10.57017/jorit.v3.2(6).02

[11] R. S. Codding, A. VanDerHeyden, and R. Chehayeb, "Using data to intensify math instruction: An evaluation of the instructional hierarchy," *Remedial and Special Education*, vol. 45, no. 3, pp. 173–188, 2023. https://doi.org/10.1177/07419325231194354

[12] J. H. Ryoo, R. H. Tai, and A. D. Skeeles-Worley, "Examination of longitudinal invariance on a framework for observing and categorizing instructional strategies," *Research in Science Education*, vol. 50, no. 2, pp. 489–504, 2018. https://doi.org/10.1007/s11165-018-9698-7

[13] S. Park, C. Gonzalez, and P. Puranam, "Decision centralization and learning from experience in groups: Separating context from aggregation effects," *Management Science*, vol. 1, no. 1, pp. 1–15, 2024. https://doi.org/10.1287/mnsc.2022.01507

[14] W. Lenhard, H. Baier, and J. Hoffmann, "Computer assisted reading comprehension instruction: Effects of automatically generated content feedback," *Zeitschrift für Pädagogische Psychologie*, vol. 26, no. 2, pp. 135–148, 2012.

[15] S. Hu, K. Laxman, and K. Lee, "Exploring factors affecting academics' adoption of emerging mobile technologies-an extended UTAUT perspective," *Education and Information Technologies*, vol. 25, no. 5, pp. 4615–4635, 2020. https://doi.org/10.1007/s10639-020-10171-x

[16] J. Zhang and S. Yu, "Assessing the innovation of mobile pedagogy from the teacher's perspective," *Sustainability*, vol. 14, no. 23, p. 15676, 2022. https://doi.org/10.3390/su142315676

6 AUTHORS

Lin Wang graduated from Hebei Normal University with a Master of Laws degree. Currently, she works as a Lecturer at College of Liberal Arts and Law Education, Hebei Open University, Shijiazhuang 050000, China, Associate Professor, with her research focus on adult higher education. She has presided over and completed 7 provincial-level research projects and 6 department-level research projects. Additionally, she has published 10 papers, 2 monographs and edited 2 textbooks of Law (E-mail: Wl479603043@163.com).

Yihua Li graduated from Ordnance Engineering College with a postgraduate degree in Computer Science and Technology. Currently, she works as a Lecturer at College of Science and Engineering, Hebei Open University, Shijiazhuang 050000, China, with her research focus on adult higher education. She has presided over 1 department-level educational reform project, and 3 projects that have been completed. Additionally, she has published 4 papers and 2 monographs (E-mail: hbddlyh@163.com).

Xian Gao graduated from Hebei Normal University with a bachelor's degree in software engineering. Currently, she is a Lecturer at Ministry of Open Education, Cangzhou Open University, Cangzhou 061001, China, specializing in computer science and technology. She has led one research project on social science development in Cangzhou City, published three papers, and co-authored one book titled "Research on Modern Computer Information Security Technology" (E-mail: 1991501267@qq.com).