Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 3 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i03.60251

PAPER

A Framework for the Integration of Mobile Technology and Artificial Intelligence with the Aim of Evaluating the Quality of Teaching in Higher Vocational Education

Junxiang Wang(✉)

Shijiazhuang College of Applied Technology, Shijiazhuang, China

2006110054@sjzpt.edu.cn

ABSTRACT

Against the backdrop of global digital transformation in vocational education, the practice-oriented and skill-centered nature of higher vocational education poses challenges for teaching quality evaluation, including fragmented data collection across diverse instructional settings such as theoretical courses, practical training, and internship placements and the lack of effective methods for capturing unstructured operational data. Skill assessments also remain dependent on subjective human judgment, limiting objectivity and quantifiability of performance indicators such as the compliance of electrical wiring tasks. Although mobile technology and artificial intelligence (AI) offer potential solutions, existing approaches lack a systematic evaluation paradigm aligned with vocational education needs. Current research remains limited by insufficient scenario adaptability and misalignment between technological functions and pedagogical requirements. To develop a teaching quality evaluation system for vocational education that integrates mobile technology and AI, the core research questions include the construction of a system framework adaptable to diverse instructional scenarios, the implementation pathways of key technical modules, and the verification of the system's effectiveness. A four-dimensional framework—data acquisition, intelligent analysis, feedback optimization, and management coordination—was established, and a prototype system featuring full-scenario mobile data capture and AI-based skill quantification was implemented. Quasi-experimental studies in three vocational institutions of varying types demonstrate the system's ability to address contextual and technical bottlenecks. The proposed scenario–technology–education alignment paradigm provides a reusable technical solution and empirical basis for quality assurance in vocational education.

KEYWORDS

higher vocational education, teaching quality evaluation, mobile technology, artificial intelligence (AI), skill quantification assessment

Wang, J. (2026). A Framework for the Integration of Mobile Technology and Artificial Intelligence with the Aim of Evaluating the Quality of Teaching in Higher Vocational Education. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(3), pp. 4–17. https://doi.org/10.3991/ijim.v20i03.60251

Article submitted 2025-10-18. Revision uploaded 2025-12-03. Final acceptance 2025-12-14.

1 INTRODUCTION

Global vocational education has entered a stage of high-quality development driven by digital transformation [1, 2]. Compared with general higher education, the practice-oriented and skill-centered attributes of higher vocational education have created distinctive challenges for teaching quality evaluation. First, the fragmentation of instructional scenarios has resulted in discontinuous data acquisition. Quality-related data from theoretical instruction, hands-on training, and workplace practicum are difficult to capture and synchronize in real time, and non-structured data—such as operational processes during practical training or performance behaviors during internships—lack effective and scalable collection mechanisms [3–5]. Second, the subjectivity inherent in skill assessment has hindered quantification efforts. Traditional evaluation approaches rely primarily on visual inspection and experiential judgment by instructors and industry mentors, making it difficult to generate objective and quantitative indicators for tasks such as the compliance of electrical wiring operations or the precision of mechanical machining [6, 7]. The joint advancement of mobile technology and artificial intelligence (AI) has created a critical technological opportunity to address these bottlenecks [8]. The ubiquity of 5G networks, mobile devices, and Internet of Things (IoT) infrastructures enables real-time acquisition and transmission of quality-related data across diverse instructional scenarios, thereby eliminating the temporal and spatial constraints of traditional assessment practices [9]. AI further facilitates the quantification of non-structured data through methods such as computer vision for evaluating operational accuracy during practical training and natural language processing for extracting key information from school–enterprise evaluations [10]. The deep integration of mobile technology and AI thus establishes a technical chain comprising full-scenario data acquisition, intelligent analysis, and instantaneous feedback, providing a viable foundation for reconstructing the teaching quality evaluation system in higher vocational education.

The portability and real-time capabilities of mobile technology have led to its increasingly widespread use in educational evaluation. Existing studies, both domestically and internationally, have primarily concentrated on three areas. First, mobile applications have been employed to facilitate classroom interaction and feedback through functions such as instant response submission and satisfaction rating [11, 12]. Second, learning processes have been tracked by recording behavioral data such as resource access and assignment submissions. Third, off-campus learning has been evaluated by capturing data from internships, fieldwork, and other learning environments [13, 14]. Despite these developments, most existing research remains focused on general instructional scenarios. Limited attention has been given to the scenario adaptability required in higher vocational education, particularly for complex environments such as hands-on training and workplace practicum. As a result, critical challenges—such as real-time acquisition of operational data during training activities and synchronized integration of school–enterprise evaluation data—remain unresolved. AI has exhibited substantial potential in educational evaluation owing to its capacity for intelligent analysis [15, 16]. Current research has concentrated on areas such as learning analytics, automated scoring, and affective computing. However, significant deficiencies in scenario adaptability persist. Much of the existing work is oriented toward generic educational contexts and lacks deep integration with vocational education requirements, including the evaluation of operational compliance in skill-based tasks and coordinated assessment involving both educational institutions and industry partners. Additionally, certain studies have adopted a technology-centric approach that overlooks the fundamental principles of vocational education, particularly its dual emphasis on skill development and industry alignment. This misalignment has led to a disconnect between technological applications and pedagogical needs.

A synthesis of existing studies indicates that although mobile technology and AI have each demonstrated meaningful progress in educational evaluation, their integrated application has not yet produced a systematic solution tailored to the characteristics of higher vocational education. In response to the identified research gaps, three overarching research objectives were established: (a) A teaching quality evaluation framework integrating mobile technology and AI was proposed based on the practice-oriented attributes of higher vocational education, with explicit articulation of system components and their collaborative mechanisms. (b) The core technological modules of the system were developed, and a prototype system was constructed to address key technical bottlenecks, including data acquisition during hands-on training and the quantitative assessment of vocational skills. (c) A multi-case quasi-experimental approach was employed to evaluate system effectiveness, providing empirical evidence to support practical implementation.

The structure of the study is organized below. Section 2 presents a four-dimensional system framework comprising data acquisition, intelligent analysis, feedback optimization, and management coordination, alongside the technological pathways for implementation. Section 3 describes the development of the prototype system and reports the results of both functional and performance testing. Section 4 evaluates system effectiveness through quasi-experimental studies conducted in three higher vocational institutions of differing types. Section 5 discusses the key findings, theoretical contributions, and practical implications and outlines the limitations of the study. Section 6 synthesizes the principal conclusions and outlines directions for future research.

2 SYSTEM FRAMEWORK CONSTRUCTION

To address the limitations of traditional teaching quality evaluation systems in higher vocational education—such as incomplete data coverage, outdated evaluation methods, and insufficient early-warning capabilities—a four-dimensional integrated framework for teaching quality evaluation was proposed through the incorporation of mobile technology and AI. The framework is grounded in a full-scenario data chain and driven by AI-based intelligent analysis, complemented by two additional core modules to form a closed-loop system encompassing data acquisition, analysis, application, and feedback. The four core modules are characterized by strong interdependence. The mobile-enabled full-scenario data acquisition module provides high-quality, multi-source, and heterogeneous data for the entire system. The AI-driven intelligent analysis module processes and mines these data to generate evaluation results and early-warning outputs. The remaining two modules support the application of evaluation results and the optimization of feedback mechanisms. Figure 1 presents the overall framework of the teaching quality evaluation system that integrates mobile technology and AI. The diagram illustrates the structure and interactions among the four core modules, highlights the closed-loop logic of "data acquisition–analysis–application–feedback," and reflects the collaborative support enabled by edge computing and cloud-based clusters.

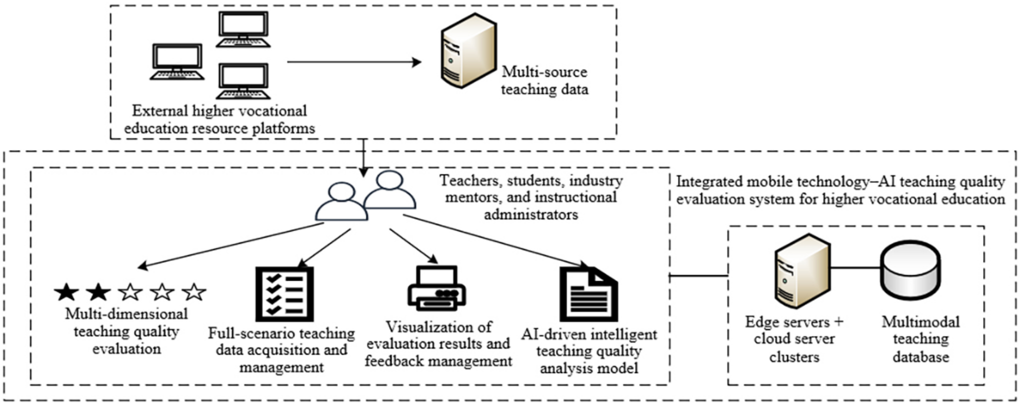


Fig. 1. Overall framework of the teaching quality evaluation system for higher vocational education

2.1 Mobile-enabled full-scenario data acquisition module

Serving as the data foundation of the entire evaluation system, this module employs mobile terminals as the primary interface and integrates IoT sensors with AI-based perception technologies to enable comprehensive, multi-dimensional, and high-precision data acquisition across classroom instruction, hands-on training, workplace practicum, and multi-source evaluation processes. The classroom data acquisition submodule records real-time indicators such as teacher–student interaction frequency, response accuracy rates, and levels of student participation through mobile terminals. AI-based speech recognition is employed to analyze instructional speech rate and coverage of key knowledge points, while AI-driven video analytics are used to detect students' attention levels, thereby enabling the quantification of both subjective and objective classroom data. The hands-on training data acquisition submodule enables students to upload operational videos and images through mobile devices. Equipment operating parameters are synchronously captured using IoT sensors. AI-based computer vision methods are then applied to compare recorded operations with a predefined standard operation library, allowing the system to identify the procedural compliance of tasks such as wiring sequences and operational gestures. The workplace practicum data acquisition submodule uses mobile terminals to facilitate the timely submission of internship logs and supervisor evaluations. AI-based text mining is applied to analyze sentiment tendencies within internship reports, while domain-specific terminology matching is used to assess the attainment of professional skill standards.

2.2 AI-driven intelligent analysis module

This module uses the multi-source heterogeneous data obtained from the full-scenario data acquisition module as input and applies AI techniques—such as machine learning and deep learning—to enable intelligent processing, analysis, and mining of the data, thereby providing analytical support for teaching quality evaluation and decision-making. The multi-dimensional evaluation submodule develops an evaluation index system encompassing instructional behaviors, learning outcomes, and school–enterprise collaboration. A random forest algorithm is employed to integrate the normalized data for all indicators. The selection of the random forest method is justified by its strong capability in handling high-dimensional data and its robustness against overfitting. The output results include: (a) a teacher instructional competence index, (b) a course quality rating (classified into five levels, A–E), and (c) a student skill attainment rate (measured by the degree of alignment with professional qualification standards). The teacher instructional competence index is calculated through the weighted integration of eight secondary indicators. The core formula is expressed as:

$$T = \sum_{i=1}^{8} w_i \cdot x_i \tag{1}$$

where, T represents the teacher instructional competence index, w_i denotes the weight of the i-th secondary indicator, and x_i denotes the normalized value of the corresponding indicator.

The personalized diagnostic submodule constructs digital profiles of instructors and students by integrating multi-period data, enabling the identification of instructional weaknesses and learning deficiencies. The risk early-warning submodule employs a long short-term memory (LSTM) algorithm to build a time-series prediction model. The core update equations for the cell state and hidden state are given as:

$$C_t = f_t \odot C_{t-1} + i_t \odot C_t', h_t = o_t \odot \tanh(C_t) \tag{2}$$

where, C_t and C_{t-1} represent the current and previous cell states, respectively; h denotes the current hidden state output; f_t, i_t, and o_t correspond to the forget gate, input gate, and output gate; C_t' represents the candidate cell state; and $\odot$ denotes element-wise multiplication. Historical data from the most recent three academic terms are used as the training set for the model, which is employed to predict potential risks and generate tiered early-warning signals.

The deep mining submodule adopts the Apriori association rule algorithm to explore underlying data relationships. The core evaluation metric is defined as:

$$\text{support}(A \Rightarrow B) = \frac{\text{count}(A \cup B)}{N}, \text{confidence}(A \Rightarrow B) = \frac{\text{count}(A \cup B)}{\text{count}(A)} \tag{3}$$

where, A and B denote itemsets, count($A \cup B$) represents the number of transactions containing both A and B, count(A) represents the number of transactions containing A, and N denotes the total number of transactions. Through this algorithm, associations such as the relationship between equipment utilization and skill attainment rates are identified, providing analytical support for decision-making.

2.3 Mobile-enabled feedback and improvement loop module

As the critical operational component closing the "data acquisition–analysis–application–feedback" loop, this module employs mobile terminals as the primary interaction interface and integrates data visualization with AI-driven recommendation technologies to achieve precise translation of evaluation results into instructional improvement actions. The instantaneous feedback submodule converts outputs from the AI-driven intelligent analysis module into visualized reports—such as radar charts and trend graphs—which are delivered in real time to instructors, students, and administrators via mobile applications/mini-programs. The intelligent recommendation submodule constructs a resource recommendation model based on the digital profiles of instructors and students and their identified deficiencies. Targeted improvement strategies are recommended for instructors, such as exemplary hands-on training cases or demonstration videos, when "insufficient operational guidance" is detected. For students, personalized learning resources—such as specialized exercises or micro-lesson videos for "weak programming logic"—are recommended. The improvement tracking submodule records improvement actions taken by instructors and students through mobile devices and links subsequent evaluation data to generate quantitative analyses of improvement effectiveness. This mechanism enables continuous optimization through the closed-loop cycle of "evaluation–feedback–improvement–re-evaluation."

2.4 Intelligent collaborative management module

This module focuses on enhancing the collaborative efficiency of the entire teaching quality evaluation process and is built on the integration of mobile-office functions with AI-based scheduling to support coordinated management among institutions, instructors, students, and industry partners. The process management submodule establishes an evaluation task management platform on mobile devices, enabling a fully digital workflow covering indicator publication, task allocation, progress tracking, and result approval. Administrators are able to monitor progress in real time, and system-generated notifications are used to remind stakeholders of task milestones, reducing communication overhead. The resource scheduling submodule integrates AI optimization algorithms to construct a resource allocation optimization model based on analytical outputs from the intelligent analysis module. This model supports time-sharing scheduling of hands-on training equipment and cross-class coordination of instructional staff. For example, in response to "imbalanced equipment utilization," the system automatically generates equipment reservation and allocation plans to enhance the efficiency of resource use. The school–enterprise collaboration submodule establishes standardized data interfaces to enable seamless data exchange between institutional evaluation systems and enterprise internship management platforms. Unified internship evaluation indicators and skill attainment standards are implemented, enabling enterprises to upload internship evaluation data in real time and institutions to synchronize updates on students' learning status. This mechanism eliminates information barriers between schools and enterprises and ensures completeness and coherence in evaluation data.

2.5 Core technical implementation

Figure 2 presents the domain model of the teaching quality evaluation system for higher vocational education. The diagram defines the core entities—including teaching sessions, system users, evaluation models, and evaluation indicators—and clarifies their attributes as well as the relational logic among them, following the sequence of "teaching session data generation–indicator mapping–model analysis–result output." The core technologies of the system were designed around three objectives: data acquisition accuracy, analytical depth, and collaborative efficiency. Technical breakthroughs and optimizations were implemented for the key stages accordingly. Specifically, the hands-on training operation recognition technology was developed using You Only Look Once (version 8), i.e., YOLOv8, as the base model. A specialized dataset was constructed to cover six categories of representative vocational skills, including electrical wiring and mechanical machining. Mosaic data augmentation was applied to increase sample diversity, and transfer learning strategies were introduced to reduce training costs. An improved loss function was incorporated, with the core Complete Intersection over Union (CIoU) loss expressed as:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{4}$$

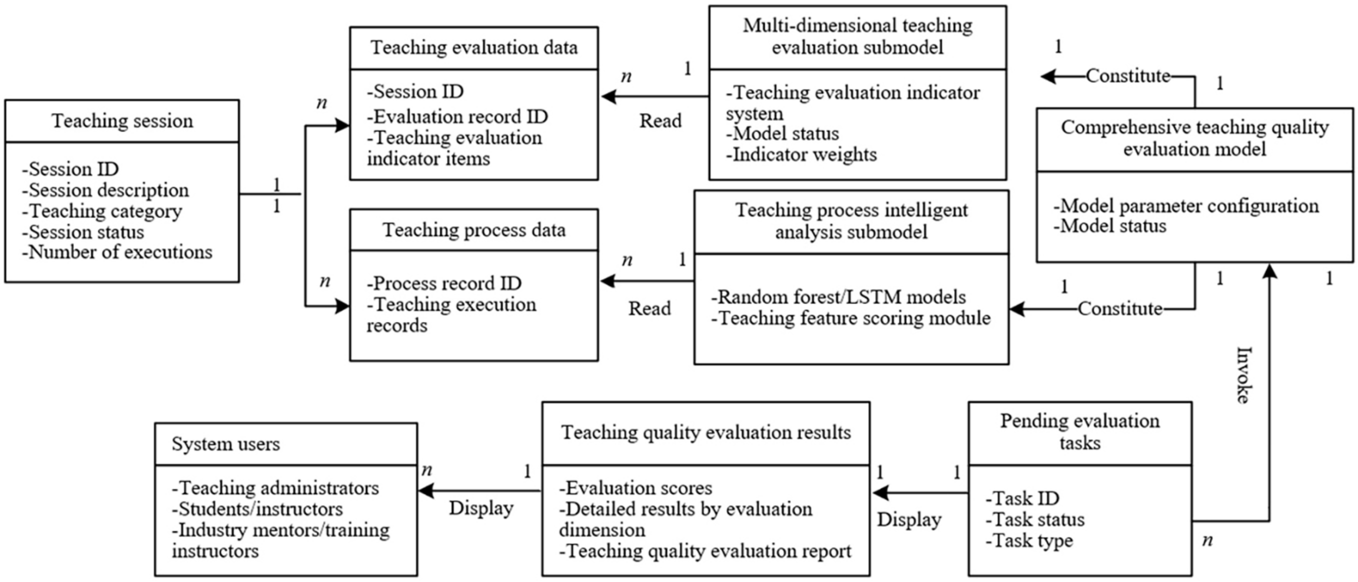


Fig. 2. The domain model of the teaching quality evaluation system for higher vocational education

where IoU denotes the intersection over union between the predicted and ground truth bounding boxes; $\rho^2(b, b^{gt})$ represents the squared Euclidean distance between the centers of the predicted box (b) and ground truth box (b^{gt}); c is the diagonal length of the minimum enclosing box covering both bounding boxes; α is a balancing coefficient; and v measures the consistency of aspect ratios.

The evaluation text analysis technology was developed based on Bidirectional Encoder Representations from Transformers (base version), i.e., BERT-base, using a dual-task learning framework to perform sentiment analysis and keyword extraction simultaneously for evaluation texts originating from instructor evaluations, student evaluations, and school–enterprise mutual assessments. Sentiment analysis is enhanced through a local attention mechanism designed to strengthen the identification of evaluative tendencies. Keyword extraction integrates term frequency-inverse document frequency (TF-IDF) weighting with BERT word embedding vectors, where the core TF-IDF formula is given as:

$$\begin{gathered} TF\_IDF(t,d) = TF(t,d) \times IDF(t), TF(t,d) = \frac{n_{t,d}}{\sum_k n_{k,d}} \\ IDF(t) = \log \frac{|D|}{1 + |d \in D | t \in d|} \end{gathered} \tag{5}$$

where $TF(t, d)$ denotes the term frequency of term t in document d; $n_{t,d}$ represents the number of occurrences of term t in document d; $\sum_k n_{k,d}$ denotes the total number of word occurrences in document d; $IDF(t)$ represents the inverse document frequency of term t; $|D|$ is the total number of documents; and $d \in D | t \in d$ denotes the set of documents in which term t appears.

In the LSTM-based risk prediction model, data preprocessing is conducted using min–max normalization and the 3σ rule for outlier removal. The min–max normalization formula is expressed as:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{6}$$

A mobile technology–AI collaborative technique was adopted to optimize latency performance under 5G environments. An "edge computing + hierarchical data transmission" architecture was implemented. Edge computing nodes were deployed within hands-on training sites so that YOLOv8 inference tasks for operation recognition can be offloaded to the edge layer. For data transmission, H.265 lightweight encoding was employed to compress hands-on training video streams, and a priority-based scheduling mechanism was introduced. Through this approach, end-to-end transmission latency was reduced from 45 ms—observed in standard 5G environments—to 18 ms, meeting the real-time analysis requirements of the system.

3 PROTOTYPE SYSTEM DEVELOPMENT AND TESTING

To ensure the core performance requirements of "full-scenario adaptability, high-concurrency support, and efficient intelligent collaboration," a design approach integrating hierarchical decoupling and edge–cloud collaboration was adopted. On this basis, a four-layer technical architecture—comprising the perception layer, transmission layer, intelligence layer, and application layer—was constructed. The perception layer serves as the system's data acquisition entry point and integrates multiple types of terminals and sensing devices. Mobile terminals incorporate a lightweight data acquisition software development kit (SDK) capable of capturing multimodal data such as text, video, and interaction trajectories. IoT sensors collect equipment usage states and resource linkage data. High-definition cameras simultaneously capture operational processes in hands-on training environments, providing raw data for subsequent AI-based analysis. The transmission layer adopts a hybrid transmission scheme that integrates "5G as primary, WiFi as supplementary + edge computing." The 5G network employs network slicing to allocate a dedicated data transmission channel, ensuring low-latency transfer of high-definition video and large instructional files. WiFi is used in fixed indoor classroom scenarios to reduce energy consumption. Edge computing nodes are deployed to perform data preprocessing and caching, including data cleaning, format conversion, and high-frequency resource pre-storage, thereby reducing transmission pressure on the cloud. End-to-end transmission latency is maintained within 0.5s. The intelligence layer provides the core computational support for the platform and adopts a cloud-collaborative architecture. AI algorithm engines are deployed in the cloud and integrate models for resource value assessment, personalized recommendation, and supply–demand early warning. The distributed database cluster consists of relational and non-relational databases, complemented by a Redis caching cluster to improve the efficiency of high-frequency data access. The application layer follows a multi-terminal adaptation design, enabling the development of a student–instructor mobile application, an administrative web system, and an inter-institutional collaboration mini-program. These components support functions such as resource learning, platform management, and cross-institution sharing, with data synchronized across all interfaces in real time.

The prototype system development follows the principles of "technological maturity and scalability," with explicit specification of the development environment, core functional implementation pathways, and data storage scheme to ensure reproducibility and iterative extensibility. Regarding the development environment, the mobile application was developed using the cross-platform framework Flutter 3.7 to ensure compatibility with both iOS and Android systems, supported by Android Studio Hedgehog and Xcode 14.3. The backend was constructed using Python 3.9 and the Django 4.2 framework to provide RESTful API services for front–backend data interaction. The AI algorithm engine was developed on TensorFlow 2.10, with model training conducted on an NVIDIA RTX 3090 GPU under the Ubuntu 20.04 operating system. A master–slave replication architecture was adopted for the database to ensure data reliability and fault tolerance. Core functional implementation focuses on three key modules. The data acquisition module enables real-time data upload across multiple instructional scenarios. A "resource QR-code association" feature is embedded within the mobile application, allowing users to bind instructional materials or equipment by scanning their QR codes, with all associated access trajectories automatically recorded. The intelligent recommendation module provides a "personalized resource showcase" on the home interface of the student–instructor mobile application. This interface is dynamically updated based on user profiles and supports interactive functions such as "one-click bookmarking" and "online inquiry." The inter-institutional collaboration module enables cross-institutional resource access through a unified API interface. The administrative web system provides resource review and permission configuration functions and supports the batch import of high-quality external instructional resources.

4 EXPERIMENTAL RESULTS AND ANALYSIS

Table 1. System functionality test results

Test Module	Representative Functional Point	Number of Test Cases	Function Implementation Rate (%)	Data Accuracy (%)	Exception Handling Pass Rate (%)
Mobile-enabled full-scenario data acquisition module	Classroom answer data acquisition	30	100	99.2	96.7
	Hands-on training operation video upload	25	100	98.8	93.3
	Internship report text upload	20	100	99.5	95.0
	IoT equipment data synchronization	25	100	98.5	92.0
AI-driven intelligent analysis module	Multi-dimensional classroom teaching quality evaluation	35	100	92.5	90.0
	Hands-on training operation compliance recognition	40	100	90.3	88.0
	Academic risk prediction	30	100	89.6	91.7
Mobile-enabled feedback and improvement loop module	Visualization and push delivery of evaluation reports	20	100	99.0	95.0
	Personalized improvement resource recommendation	25	100	92.0	92.0
Intelligent collaborative management module	Hands-on training equipment scheduling and reservation	20	100	98.0	90.0
	School–enterprise data interconnection	25	100	97.5	92.0

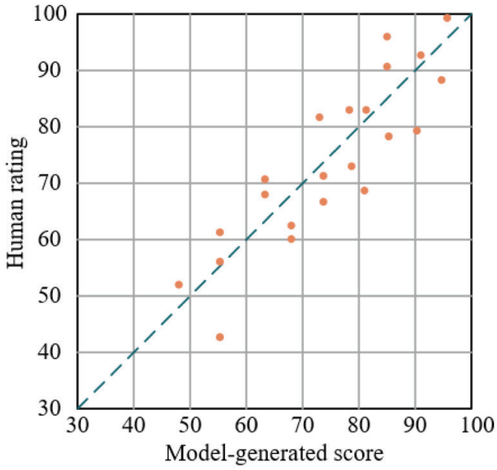
To verify the completeness of each core module, the accuracy of data processing, and the system's ability to handle exceptional scenarios—and to ensure that the platform meets the operational requirements of the full teaching quality evaluation process in higher vocational education—black-box testing was conducted across eleven representative functional points in the four core modules. As shown in Table 1, all functional points achieved a 100% implementation rate, confirming that the module design reflects no functional omissions and can fully support the closed-loop process of "data acquisition–analysis–feedback–collaboration." Overall data accuracy remained above 92.0%. Structured or semi-structured data types—such as classroom answer records and internship reports—achieved accuracy levels exceeding 99%, while evaluation and prediction accuracy within the AI analysis module exceeded 89.6%, demonstrating the system's precision in data handling. Exception handling pass rates exceeded 88.0% across all functional points, indicating that common abnormalities—such as video upload interruptions or device connection failures—can be effectively managed. Collectively, the results of the functionality testing demonstrate that the system design aligns with the practical needs of teaching quality evaluation in higher vocational education and exhibits high completeness and reliability.

Table 2. System performance test results

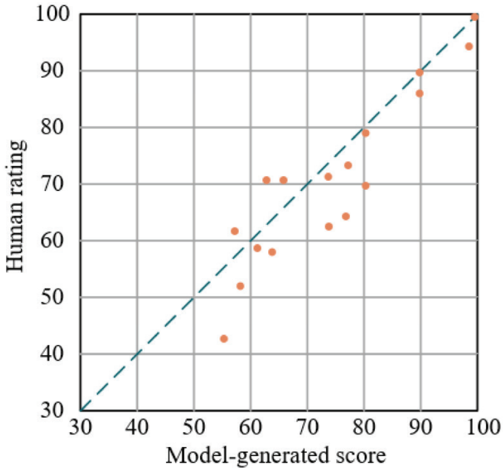
Test Type	Concurrent Users (persons)	Test Task	Average Data Transmission Latency (ms)	Average AI Analysis Response Time (s)	Server CPU Utilization (%)	Server Memory Utilization (%)
Concurrent performance test	200	Concurrent classroom answering + data upload	42	1.2	38	45
	500	Concurrent training video upload + AI analysis + report viewing	82	2.1	65	62
	800	Concurrent multi-scenario data acquisition + analysis + collaborative scheduling	98	2.8	82	78
Stability test	500	Continuous mixed tasks for 12 hours	85 ± 3	2.3 ± 0.2	68 ± 5	65 ± 4
	500	Continuous mixed tasks for 24 hours	88 ± 4	2.5 ± 0.3	72 ± 6	69 ± 5
Stress test	1000	High-load tasks (1080p video upload + complex AI analysis)	125	3.5	95	89

To evaluate the system's performance stability under multi-user concurrent access, long-duration operation, and high-load conditions typical of higher vocational education campuses and to ensure its applicability to large-scale teaching evaluation scenarios, performance testing was conducted across concurrency, stability, and stress scenarios. As shown in Table 2, under the core benchmark of 500 concurrent users, average data transmission latency reached 82 ms, average AI analysis response time reached 2.1s, and server CPU and memory utilization reached 65% and 62%, respectively—all within acceptable operational thresholds. Even under 800 concurrent users, all performance indicators remained within expected limits, demonstrating strong concurrent processing capability. In stability testing, when 500 concurrent users executed mixed tasks continuously for 24 hours, fluctuations in transmission latency and AI analysis response time remained within ±4 ms and ±0.3s, respectively. Server resource utilization remained stable with no crashes or abrupt performance degradation, indicating robust long-term operational reliability. In the stress test scenario involving 1000 concurrent users, performance thresholds were not met; however, this load exceeds the maximum realistic user volume expected in any single higher vocational campus evaluation scenario and does not affect system applicability. Collectively, the performance testing results indicate that the system satisfies the performance requirements of large-scale teaching quality evaluation in higher vocational education and demonstrates strong concurrent processing capability and long-term operational stability.

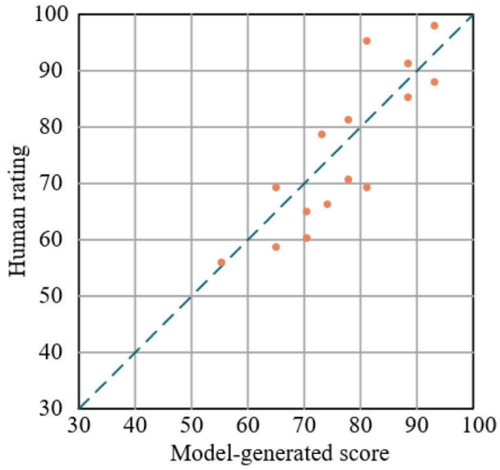
To evaluate the reliability and technical advantages of the AI-based evaluation models across multiple instructional scenarios, a consistency analysis was conducted between model-generated scores and expert human ratings for three core evaluation contexts: classroom teaching, hands-on training operations, and internship report assessment. Quantitative metrics and statistical tests were applied to assess consistency. As shown in Figure 3, the scatter distributions for all three models exhibit strong linear trends. Calculated coefficients of determination for the classroom teaching evaluation model, hands-on training operation recognition model, and internship report analysis model were 0.89, 0.86, and 0.84, respectively. The corresponding root-mean-square errors were 3.2, 4.1, and 4.5. Further analysis using paired t-tests revealed that the p-values for all three scenarios exceeded 0.05, indicating no statistically significant differences between model-generated scores and expert ratings.



a) Consistency of Classroom Teaching Evaluation Model Scores



b) Consistency of Hands-on Training Operation Recognition Model Scores



c) Consistency of Internship Report Analysis Model Scores

Fig. 3. Scatter plots of consistency between model-generated scores and human ratings in higher vocational teaching quality evaluation

These findings confirm, from a statistical perspective, that the AI-based models embedded in the system provide scoring accuracy comparable to that of domain experts across multiple instructional scenarios. Moreover, the results highlight the system's technical breakthroughs in addressing long-standing challenges in teaching quality evaluation within higher vocational education. Unlike traditional human evaluations—which are inherently subjective and labor-intensive—the integration of mobile full-scenario data acquisition with AI-driven intelligent analysis enables an automated, quantitative, and objective evaluation process. At the same time, the subtle performance differences among models across different scenarios also suggest that further optimization is needed to enhance model adaptability in highly complex hands-on operational environments.

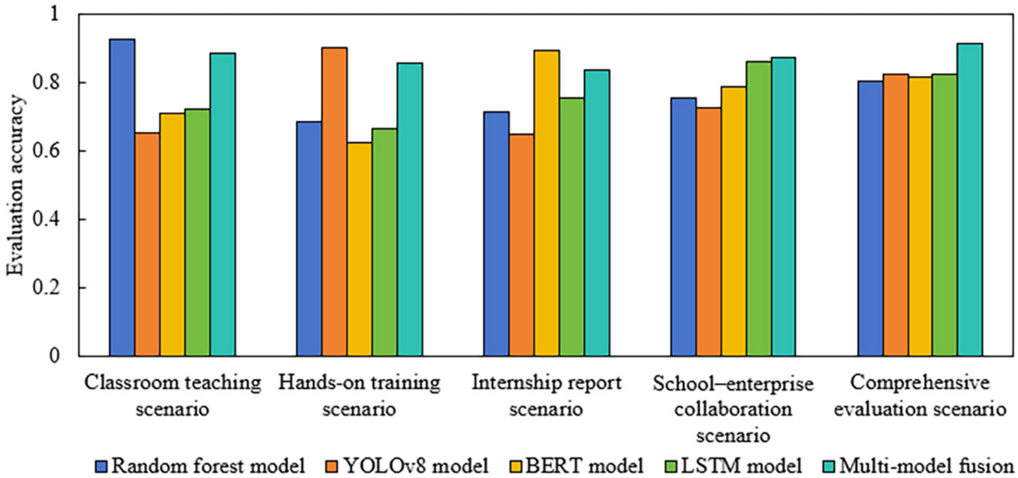


Fig. 4. Multi-scenario performance comparison of teaching quality evaluation models in higher vocational education

To assess the scenario adaptability of the AI models tailored for core teaching quality evaluation contexts in higher vocational education and to examine the synergistic enhancement achieved through multi-model fusion, a multi-model, multi-scenario performance comparison experiment was conducted. This experiment further quantifies the feasibility of the system's technical pathway of "scenario-specific precision evaluation and global fusion optimization." As shown in Figure 4, each scenario-specific model demonstrates a clear performance advantage within its corresponding context. The random forest model, designed for structured classroom teaching evaluation data, achieved an accuracy of 92.5%, representing a 23.8-percentage-point improvement over its accuracy of 68.7% in the hands-on training scenario. The YOLOv8 model, optimized for image-based hands-on training operations, reached an accuracy of 90.3%, substantially higher than its 64.8% accuracy in the internship report text scenario. The BERT model, adapted for internship report analysis, achieved an accuracy of 89.2%, while the LSTM model, tailored for school–enterprise collaborative time-series data, achieved an accuracy of 86.1%. All models performed considerably better in their adapted scenarios than in non-adapted ones, confirming the scientific validity of designing models in alignment with the characteristics of scenario-specific data. This addresses the long-standing limitation of traditional single-model systems. More importantly, the multi-model fusion strategy achieved "near-optimal to optimal" performance across all scenarios. In the comprehensive evaluation scenario, the fused model achieved an accuracy of 91.5%, improving upon the best-performing single models by 1.2% to 5.1%.

In non-adapted scenarios, the fused model also exceeded the accuracy of all single models except for the designated scenario-specific models by at least 19.2 percentage points. These results fully reflect the complementary and integrative value of the fusion strategy across diverse evaluation contexts.

5 CONCLUSION

This study addressed long-standing challenges in teaching quality evaluation in higher vocational education, including the subjectivity of manual assessment, insufficient adaptability across heterogeneous instructional scenarios, and low efficiency in data coordination. A four-dimensional integrated evaluation system incorporating mobile technology and AI was designed and implemented to overcome these limitations. The findings demonstrate substantial value at both the technical and application levels. From a technical perspective, a full-process evaluation framework was established to support classroom instruction, hands-on training, internship activities, and so on. Core technological advances were achieved through the integration of YOLOv8, BERT, and LSTM models with 5G-enabled edge computing, enabling breakthroughs in hands-on operation recognition, evaluation text analysis, risk prediction, and mobile technology–AI collaborative processing. From an application perspective, multi-dimensional experimental verification—including function, performance, usability, and model accuracy—demonstrated that the system exhibits complete and reliable functionality, meets large-scale concurrency requirements, and achieves excellent usability ratings. The high consistency between AI-generated scores and expert ratings further validates the system's feasibility and effectiveness in supporting intelligent teaching quality evaluation. Collectively, the proposed system offers a reusable technical paradigm for advancing the digital transformation of quality assurance in higher vocational education.

6 REFERENCES

[1] B. Muchlas, P. Budiastuti, M. Khairudin, B. Santosa, and B. Rahmatullah, "The use of personal learning environment to support an online collaborative strategy in vocational education pedagogy course," *International Journal of Interactive Mobile Technologies*, vol. 17, no. 2, pp. 24–41, 2023. https://doi.org/10.3991/ijim.v17i02.34565

[2] Y. Wang and L. Feng, "Vocational education in the era of big data: Course design and optimization strategy based on educational technology," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 22, pp. 143–158, 2024. https://doi.org/10.3991/ijim.v18i22.52447

[3] M. Podaril and R. Hrmo, "Introduction of a quality management system for vocational education and training in Slovakia," *International Journal of Engineering Pedagogy*, vol. 3, no. 3, pp. 18–23, 2013. https://doi.org/10.3991/ijep.v3i3.2733

[4] 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

[5] E. Van Rensburg, "Evaluating work-based learning: Insights from an illuminative evaluation study of work-based learning in a vocational qualification," *Industry and Higher Education*, vol. 22, no. 4, pp. 223–232, 2008. https://doi.org/10.5367/000000008785201739

[6] M. Gill and G. Sharma, "Evaluation of vocational training program from the trainees' perspective: An empirical study," *Pacific Business Review International*, vol. 6, no. 5, pp. 35–43, 2011.
[7] Sukardi, Wildan, and A. Subhani, "Experiential learning in entrepreneurship teaching: An evaluation based on importance performance analysis," *International Journal of Instruction*, vol. 15, no. 4, pp. 453–472, 2022. https://doi.org/10.29333/iji.2022.15425a
[8] B. Alsanousi, A. S. Albesher, H. Do, and S. Ludi, "Investigating the user experience and evaluating usability issues in AI-enabled learning mobile apps: An analysis of user reviews," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, pp. 18–29, 2023. https://doi.org/10.14569/IJACSA.2023.0140602
[9] G. Kim and M. Rim, "Internet of things in the 5G mobile communication system: The optimal number of channels in channel hopping," *International Journal of Networked and Distributed Computing*, vol. 6, no. 2, pp. 108–117, 2018. https://doi.org/10.2991/ijndc.2018.6.2.6
[10] D. Steingard and K. Rodenburg, "Assessing the societal impact of academic research with artificial intelligence (AI): A scoping review of business school scholarship as a 'force for good'," *Learned Publishing*, vol. 38, no. 3, p. e2010, 2025. https://doi.org/10.1002/leap.2010
[11] S. Salimi and F. Mohammadzad, "Transformative teaching practices: Real-time feedback for dynamic classroom engagement," *Creative Nursing*, vol. 31, no. 4, pp. 473–479, 2025. https://doi.org/10.1177/10784535251340117
[12] J. S. Kaushik, K. Raghuraman, S. Verma, V. Arya, V. K. Gehlawat, and J. S. Kaushik, "Online interactive flipped classroom teaching in pediatrics for medical undergraduates," *Cureus*, vol. 15, no. 4, p. e37603, 2023. https://doi.org/10.7759/cureus.37603
[13] R. Smit, K. Hess, A. Taras, P. Bachmann, and H. Dober, "The role of interactive dialogue in students' learning of mathematical reasoning: A quantitative multi-method analysis of feedback episodes," *Learning and Instruction*, vol. 86, p. 101777, 2023. https://doi.org/10.1016/j.learninstruc.2023.101777
[14] M. Stephan, L. Servadei, J. Arjona-Medina, A. Santra, R. Wille, and G. Fischer, "Scene-adaptive radar tracking with deep reinforcement learning," *Machine Learning with Applications*, vol. 8, p. 100284, 2022. https://doi.org/10.1016/j.mlwa.2022.100284
[15] M. Podoliak, O. Zagranovska, V. Posmitna, N. Golovchak, and O. Kushnirchuk, "Evaluating the impact of AI-based tools on language proficiency and motivation: Experimental evidence from philology students in Ukraine," *Journal of Research, Innovation and Technologies*, vol. 4, no. 3, pp. 271–282, 2025. https://doi.org/10.56578/jorit040303
[16] J. Lee and A. G. Esposito, "ChatGPT or human mentors? Student perceptions of technology acceptance and use and the future of mentorship in higher education," *Education Sciences*, vol. 15, no. 6, p. 746, 2025. https://doi.org/10.3390/educsci15060746

7 AUTHOR

Junxiang Wang graduated with a Master of Engineering from Tiangong University and currently serves as a Lecturer at the School of Marxism, Shijiazhuang College of Applied Technology, China. Her research focuses on Computer Application Technology, the Application of Modern Educational Technology, and Online Education (E-mail: 2006110054@sjzpt.edu.cn).