

Research on the Diagnosis and Feedback Paths of Digital Intelligence Teaching Ability among Normal University Students Driven by Multimodal Data

Xiaolan Li, Hong Li *, Qian Xu

Yunnan Normal University, Kunming, Yunnan, 650500, China

* Corresponding author: Hong Li (Email: 651035658@qq.com)

Abstract: Effective diagnosis and precise feedback on normal university students' teaching abilities are key links in ensuring teacher education quality, but traditional methods have long faced challenges such as strong subjectivity and lagging feedback. The rise of multimodal learning analysis technology has provided new possibilities for automated diagnosis and precise feedback in the teaching process; however, existing methods still fall short in terms of data interpretation systematization and interpretability of feedback. Based on an overview of the current research status of teaching ability diagnosis and feedback for normal university students, this study constructs a diagnostic and feedback path model that includes four core components: multimodal data collection, intelligent diagnostic analysis, precise feedback intervention, and ability development tracking. The implementation mechanism is explained from four dimensions: data layer, analysis layer, feedback layer, and tracking layer. The research aims to provide a reference theoretical framework and practical path for diagnosing and providing feedback on intelligent teaching abilities empowered by intelligent technology for normal university students.

Keywords: Multimodal Data; Normal University Students; Digital Intelligence Teaching Ability; Diagnosis and Feedback; Path.

1. Introduction

The digital transformation of education has brought profound changes to the training of normal university students. In 2024, the Ministry of Education announced the second batch of typical cases of "artificial intelligence + higher education" application scenarios, with multiple universities' intelligent training systems for normal university students' teaching abilities selected [1]. East China Normal University uses large models to drive digital human students to build integrated training spaces [2]; South China Normal University has developed the country's first intelligent training system for basic teaching skills for normal university students, using technologies such as IoT sensing and multimodal data analysis to build a "four-dimension 18-item" intelligent training evaluation system. These practices mark the transition from theory to reality in multimodal data-driven teaching diagnostics and feedback.

However, deep-seated difficulties in diagnosing and providing feedback on normal university students' teaching abilities still exist. Normal university students generally face issues such as lack of real classroom scenarios, insufficient internship opportunities, high trial-and-error costs, and a lack of systematic evaluation and personalized intervention [3]. In traditional micro-grid teaching, normal students often fall into a "peer evaluation dilemma"—due to personal consideration and professional limitations, evaluations tend to be superficial. A single mentor often guides dozens of normal university students, making it difficult to observe each trial lecture from a close, multi-angle perspective. Although multimodal learning and analysis technologies have brought new opportunities, existing methods still face issues such as "mechanical and unfocused multimodal data interpretation [4]" and feedback lacking interpretability and feasibility [5].

How to systematically apply multimodal data analysis

technology to the diagnosis and feedback of normal students' digital intelligence teaching abilities under the guidance of educational theory, achieving a complete closed loop from "data collection" to "effective feedback," is an urgent research issue to solve. This study focuses on "how to construct diagnostic and feedback pathways for teacher college students' digital intelligence teaching ability driven by multimodal data," aiming to build a complete path model covering four stages: data collection, intelligent diagnosis, precise feedback, and development tracking, providing an actionable practical solution for cultivating digital intelligence teaching ability among normal university students.

2. Literature Review

2.1. Research Progress on Diagnosis and Feedback on Teaching Ability among Normal University Students

The evolution of methods for diagnosing and providing feedback on normal students' teaching ability has roughly gone through three stages. First, the traditional manual observation and diagnosis stage relies on instructors attending classes or video recordings for comments, which is inefficient and highly subjective. Second, the standardized diagnostic stage based on scales, which uses preset evaluation indicators for structured scoring, improving diagnostic standardization but requiring a long feedback cycle. Third, the data-driven intelligent diagnosis stage, which uses artificial intelligence technology to automatically collect, analyze, and feedback multi-source data from the teaching process.

In terms of practical exploration, universities have made many beneficial attempts. The "Large Model Digital Humans Empowering Teacher Students' Practical Teaching Ability" project developed by East China Normal University generates

multidimensional profiles and classroom diagnostics based on multimodal teaching process data, supporting process diagnosis and feedback on 31 skill points across four major dimensions: basic teaching skills, classroom Q&A skills, instructional design skills, and digital teaching skills [6]. Chengdu Normal University uses a multimodal data fusion analysis system to generate 120,000 digital profiles and 100,000 personalized diagnostic reports, forming a closed-loop mechanism of "data collection—model analysis—precise feedback." The "AI Assessment System for Normal University Teaching Skills" developed by Guangxi Normal University can perform "CT-style" scans of trial lectures, accurately diagnosing issues such as speaking speed, intonation, and alignment between teaching content and objectives, with evaluation results highly consistent with expert scores [7]. South China Normal University's "ShiNengbang" system has built five major evaluation dimensions: emotional support, learning guidance, resource integration, teaching organization, and evaluation improvement, providing assessment services to nearly 300 primary and secondary schools and over 20,000 teachers and normal students [8].

These explorations provide valuable reference for this study, but existing research mostly focuses on the implementation of technology at a single stage, lacking in-depth theoretical construction on the complete diagnostic and feedback pathway—from data collection to diagnostic analysis and then systematic design of feedback intervention and developmental tracking.

2.2. Application of Multimodal Data in Teaching Diagnosis and Feedback

Multimodal Learning Analytics emphasizes simultaneously processing multiple modal data such as video, audio, and text, to provide holistic assessments beyond traditional single-modal evaluations [9]. In teaching diagnostic scenarios, classroom audio and video recordings contain information about verbal and non-verbal behaviors, while lesson plans and blackboard notes contain information about instructional design.

From a technical implementation perspective, teaching behavior diagnosis can be carried out from three dimensions: verbal behavior, nonverbal behavior, and a combination of both. In terms of visual modality, convolutional neural networks can extract information such as teacher and student language expression, gaze distribution, gestures, and facial expressions from teaching videos. In terms of speech modality, speech recognition technology can transcribe classroom speech into text, analyzing characteristics such as speech speed, question types, and teacher-student speech ratios [10]. In multimodal integration, previous studies have established a three-dimensional evaluation system of "behavior—language—emotion." However, existing research mostly focuses on technical optimization and lacks systematic pathway design for embedding multimodal analysis into a complete closed-loop of diagnosis and feedback. This is precisely the entry point of this study.

3. Modeling of Diagnostic and Feedback Pathways

3.1. Theoretical Basis for Path Construction

The construction of this pathway model is supported by three theoretical frameworks.

First, the theory of Gagne's nine major teaching events provides a framework for defining diagnostic content. The nine major teaching events fully cover the entire classroom teaching process and can systematically define key points in the diagnosis of normal university students' teaching behaviors.

Second, formative evaluation theory emphasizes that the purpose of evaluation is not only to identify and identify, but also to promote learning and development, providing value guidance for the design of feedback segments—the ultimate goal of diagnosis is to improve the abilities of normal university students, rather than simply determining their grades.

Third, the human-machine collaboration theory points out that the effective integration of artificial intelligence and human intelligence is a key path to achieving intelligent education, providing a basis for the coordination of automatic system analysis and human intervention in the diagnostic and feedback pathways.

3.2. Model Framework for Diagnosis and Feedback Pathways

Drawing on the closed-loop experience of teaching practice, profiling, diagnosis, and recommendation, as well as the closed-loop mechanisms of data collection, model analysis, and precise feedback, this study constructs a "four-ring progressive" path model for diagnosing and providing feedback on teacher-trained students' digital intelligence teaching abilities driven by multimodal data (see Figure 1).

Step One: Multimodal Data Acquisition. Through IoT sensing devices such as cameras and microphones in smart classrooms, multimodal data from normal university students' teaching and training processes is collected in a supportive manner, including video, audio, and text data. Data collection emphasizes "accompaniment" and "insensitivity" to avoid artificial interference with the natural teaching state of normal university students.

Second Step: Intelligent Diagnostic Analysis. Automated processing and analysis of the collected multimodal data includes three sub-steps: First, profile generation, which generates a comprehensive profile of normal university students' teaching abilities based on multidimensional data; Second, problem identification: by comparing with preset evaluation indicators, it accurately identifies the strengths and weaknesses of normal university students in various teaching stages; Third, attribution analysis, which deeply examines possible causes of problems to provide a basis for accurate feedback.

Third Step: Precise feedback intervention. Based on diagnostic analysis results, three types of feedback are generated: first, visual diagnostic reports, which intuitively present various dimensions in the form of radar charts, bar charts, and other formats; Second, personalized improvement suggestions, providing specific and actionable directions for improvement based on identified problems; Third, targeted resource recommendations, intelligently pushing adapted learning resources and training tasks based on the students' weaknesses.

Fourth Link: Capability Development Tracking. Conduct longitudinal tracking of normal university students, record multiple diagnoses and feedback results, and form personal growth files. By comparing changes in ability profiles at different time points, the actual effects of feedback interventions are evaluated, and subsequent diagnostic

priorities and feedback strategies are dynamically adjusted. The four rings form a continuously iterative closed loop,

driving a spiral increase in normal students' teaching abilities.

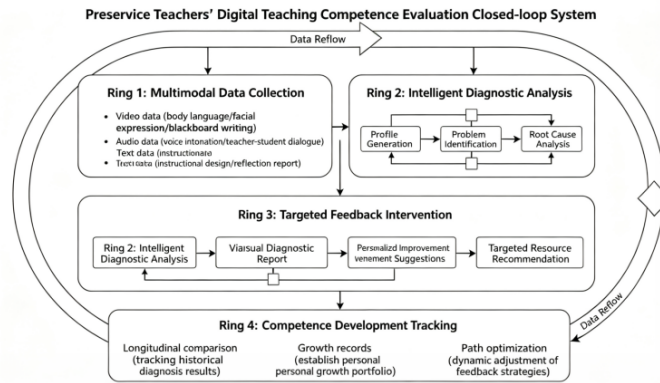


Figure 1. Four-Ring Progressive Path Model for Diagnosis and Feedback of Teacher Normal University Students' Digital Intelligence Teaching Ability

The diagnostic indicator system is the core basis for the operation of the path model. To ensure the scientific and systematic nature of diagnosis, this study refers to South China Normal University's "Four-Dimension 18 Indicators" framework and quantifiable observation indicators for basic teaching skills, and constructs a diagnostic index system for

normal students' digital and intelligent teaching abilities (see Table 1). The system includes 3 primary indicators, 9 secondary indicators, and 27 evaluation observation points, each marked with its corresponding data modal type, providing a clear mapping relationship for subsequent multimodal data collection and analysis.

Table 1. Diagnostic indicator system for digital and intelligent teaching ability among normal university students

Primary indicators	Secondary indicators	Evaluate observation points	Data modalities
Speech and behavior	Standardization of teaching language	Pronunciation Speaking quickly Spoken words	Voice
	Classroom questions and feedback	Classroom questions Feedback language Language of instruction	Voice text
Nonverbal acts	Body language and teaching posture	posture Shoulder and leg condition Posture and voice coordination	Vision
	Facial expressions and emotional expression	Facial expressions Polarity in expression Positive emoji usage	
	Blackboard writing and media usage	Fonts and handwriting Write the layout on the board Multimedia tools	
Combining verbal and nonverbal language	Classroom introduction effect	Introduction Stage	Visual Voice
	Classroom rhythm and control	Time allocation	
	Overall teaching appeal	Overall expressiveness	Omnimodal

In the indicator system, the "speech behavior" dimension focuses on highly quantifiable language features, the "nonverbal behavior" dimension emphasizes visual analysis of body language and facial expressions, and the "combination of speech and nonverbal language" dimension emphasizes comprehensive evaluation of multimodal collaboration [11]. The indicator system was constructed with reference to South China Normal University's "Four-Dimension 18 Indicators" micro-skills observation framework and quantifiable observation indicators for basic teaching skills.

3.3. Core Features of the Path Model

This path model has three core features. First, closed-loop nature, with four links connected and cyclical rather than a one-time end-to-end evaluation. Second, data-driven, based on automated collection and analysis of multimodal data,

promoting teacher cultivation from relying on fuzzy experience to refining evidence-based standards. Third, personalization: diagnosis and feedback are based on actual performance data of normal university students, achieving "targeted improvement."

4. Mechanisms for Implementing Diagnostic and Feedback Pathways

4.1. Data Layer: Accompanied Acquisition of Multimodal Data

The data layer is the foundation for path implementation. On the hardware side, relying on devices such as cameras and microphone arrays deployed in smart classrooms, accompanying data collection of normal university students' teaching and training processes is realized. The types of collected data include: classroom teaching audio and video

data, teaching design documents and courseware text data, and self-evaluation data such as teaching reflection reports. At the preprocessing level, the collected raw data is subjected to denoise, segmentation, annotation, and other processing to extract audio and video data separately into audio streams and video frame sequences to prepare for subsequent analysis.

4.2. Analysis Layer: Intelligent Diagnosis of Multimodal Data

The analysis layer is the technical core of the path. This study adopts a four-layer architecture of "data layer — analysis layer — decision layer — presentation layer" to design an intelligent analysis system (see Figure 2), where the analysis layer contains three parallel analysis modules.

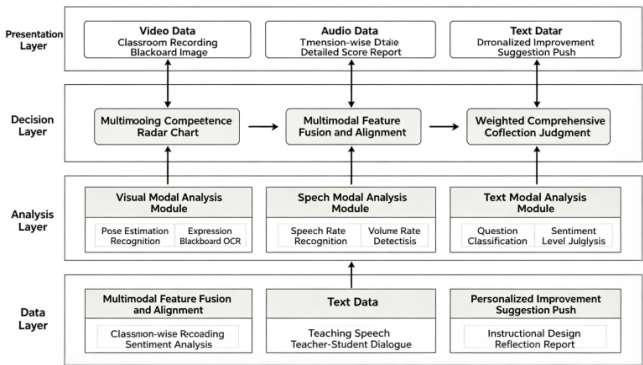


Figure 2. Four-layer architecture of the multimodal teaching behavior intelligent analysis system

Caption: The data layer is responsible for collecting and

preprocessing multi-source data; The analysis layer handles automated extraction of features from each modality; The decision-making layer completes multimodal fusion and comprehensive scoring; The display layer visualizes the analysis results. The four-layer architecture achieves a complete closed loop from "data collection" to "effective feedback."

Visual modal analysis module. Using computer vision technology to analyze the body movements, facial expressions, and blackboard standardization of Fan Sheng [12]. Posture estimation algorithms are used to extract features such as gesture frequency, body orientation, and walking trajectory; Using a deep learning-based facial expression recognition model, the richness of facial expressions and emotional expression during classroom lectures were analyzed [13].

Speech modal analysis module. Using speech recognition technology to transcribe classroom speech into text. At the speech level, analyze features such as speech speed, volume changes, and pause distribution; At the textual level, analyze the standardization of teaching language, the distribution of question types, and the ratio of teacher-student verbal interaction.

Multimodal feature fusion module. Quantitative features extracted from each mode are aligned and fused, and weighted integration is performed according to a preset evaluation indicator system to generate quantitative scores and qualitative descriptions for each dimension.

The technical implementation path for intelligent diagnostic analysis of multimodal data is shown in Figure 3.

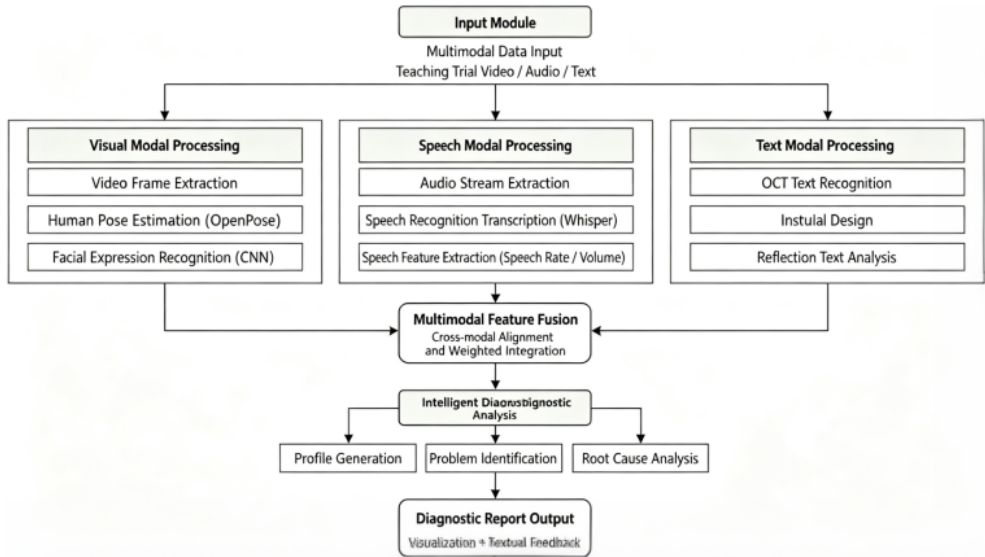


Figure 3. Implementation path of intelligent diagnostic analysis technology for multimodal data

Caption: Visual modality uses OpenPose for pose estimation and CNN for expression recognition; Speech modality is performed using Whisper for speech recognition and transcription; Text modalities are parsed using OCR and NLP technologies. After fusing each modal feature, a comprehensive diagnostic report is generated.

4.3. Feedback Layer: Precisely Generates Feedback Content

The feedback layer is the value embodiment of the path. Based on diagnostic analysis results, generate three types of feedback content.

Visualized diagnostic reports.Using radar charts, bar charts, line charts, and other forms, it visually presents the performance of normal students across various evaluation dimensions, making abstract teaching abilities perceptible and comparable.

Personalized improvement suggestions. Generate targeted improvement recommendations for specific issues identified during diagnosis. For example, it is recommended to conduct rhythm control training for the issue of "too fast speech," and to adjust the teaching design for the issue of "deviation from teaching objectives."

Targeted resource recommendations. Based on the abilities

and weaknesses of normal university students, it intelligently pushes adapted learning resources, training tasks, and reference cases from the resource library.

Feedback pushes emphasize timeliness and comprehensibility—feedback is generated immediately after diagnosis, presented in language that normal university students can understand, avoiding excessive technical expression.

4.4. Tracking Layer: Dynamic Tracking of Capability Development

The tracking layer is a continuous guarantee of the path. Establish personal growth files for normal university students, recording detailed results and feedback from each diagnosis. By vertically comparing ability profiles at different time points, the book presents the development trajectory and changing trends of normal university students' teaching abilities. Dynamically evaluate the actual effect of feedback interventions based on tracking data—if a dimension shows no significant improvement after multiple feedbacks, adjust diagnostic refinement or feedback strategies; If there is significant improvement, subsequent diagnostic focus should be adjusted accordingly. The tracking layer enables the entire path to adapt and continuously optimize.

5. Conclusion and Outlook

This study focuses on the diagnostic and feedback pathways for teacher-university students' digital intelligence teaching abilities driven by multimodal data, mainly carrying out four aspects: reviewing domestic and international research and practical progress in this field, pointing out that existing research lacks systematic diagnostic feedback path design; Drawing on the closed-loop mechanism of multi-university practical training, a four-stage path model was constructed for "multimodal data collection, intelligent diagnosis and analysis, precise feedback intervention, and capability development tracking." Clarify the path implementation mechanism from four layers: data, analysis, feedback, and tracking; The feasibility of the model was preliminarily verified based on cases from multiple universities. This study still has limitations: it lacks unified empirical data support, and systematic comparisons have not been conducted on practical cases from various schools; Some value judgment diagnostic dimensions still require manual participation, and automation levels need improvement; The applicability of the model to normal university students across different disciplines and stages remains to be tested. In the future, the scope of samples and practice will be expanded, large models will be introduced to tackle technical challenges in emotional computing and semantic understanding, human-machine collaboration models will be explored, and systems will be organically combined with personalized teacher guidance.

Acknowledgments

This article is a research achievement of the Yunnan

Normal University Graduate Research and Innovation Fund project "Construction of a Digital Intelligent Teaching Ability Assessment System for Normal Students and Implementation of an Intelligent Support System" (Project No.: YJSJJ26-B078).

References

- [1] Cai, J. Q. (2025). Research on enhancing quality education for targeted normal students. *Journal of Huangshan University*, 27(4), 135–140.
- [2] Liu, C. N., & Liu, W. (2025). Practical cases and reflections on artificial intelligence empowering normal university students. *Teacher Education Forum*, 38(3), 42–50.
- [3] Wu, Z. M., Liu, M., Su, Y. F., et al. (2025). Empirical study on educational data narrative empowering teacher students' teaching reflection. *Modern Educational Technology*, 35(8), 57–66.
- [4] Wu, Z. M., Liu, M., Su, Y. F., et al. (2025). Empirical study on educational data narrative empowering teacher students' teaching reflection. *Modern Educational Technology*, 35(8), 57–66.
- [5] Hu, X. Y., Xu, K. X., & Zhang, Y. B. (2024). Precise assessment and visualization of ability targeting teacher profiling. *China Educational Technology*, (1), 104–110.
- [6] Huang, Q. S., Zhang, S. Q., & Du, Y. Y. (2026). Immersion and empowerment: A qualitative study on the formation mechanism of normal university students' digital teaching abilities. *Teacher Education Research*, 38(1), 78–84.
- [7] Jiang, J. F., Huang, Q. J., & Huang, X. Y. (2025, November 12). AI as a mentor: Solving the problem of teacher evaluation for normal university students. *Guangxi Daily*, p. 009. <https://doi.org/10.28292/n.cnki.ngxrb.2025.011920>.
- [8] Yang, X. Y. (2026). Research on artificial intelligence empowering teacher ethics in universities [Doctoral dissertation]. Guangxi Normal University. <https://doi.org/10.27036/d.cnki.ggxsu.2026.001875>.
- [9] Wu, Z. M., Liu, M., Su, Y. F., et al. (2025). Empirical study on educational data narrative empowering teacher students' teaching reflection. *Modern Educational Technology*, 35(8), 57–66.
- [10] Cui, J. J., Kang, L., & Ma, M. (2024). A review of research on classroom teacher-student interactive intelligent analysis technology. *Computer Science*, 51(10), 40–49.
- [11] Liu, M. J. (2024). Research on multimodal behavior patterns of teachers in classroom interaction events [Doctoral dissertation]. Central China Normal University. <https://doi.org/10.27159/d.cnki.ghzsu.2024.002641>.
- [12] Liu, C. N., & Liu, W. (2025). Practical cases and reflections on artificial intelligence empowering normal university students. *Teacher Education Forum*, 38(3), 42–50.
- [13] Feng, J. (2023). Research on student expression recognition algorithms in classroom scenarios [Doctoral dissertation]. Xi'an Polytechnic University. <https://doi.org/10.27390/d.cnki.gxbfc.2023.000140>.