Vol.7 Iss.5 2026 Communication & Education Review
https://doi.org/10.37420/j.cer.2026.065

Knowledge Extraction and Evaluation Framework for Medical Procedure Videos

Peisong Wang[1], Shengyun Li[1], Hongyang Chen[1], Menghuan Chen[1], Chang Liu[1], Yiping Yu[2]*

1 Clinical Skills Training Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450000, Henan, China

2 Reproductive Medicine Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450000, Henan, China

Corresponding author, E-mail: yyplcyx@163.com

Abstract

This study explores methods for knowledge extraction from medical procedure videos and the development of a comprehensive evaluation framework, with the aim of providing standardized and quantifiable analytical tools for medical education and clinical skills training. Leveraging advances in computer vision and natural language processing, a multimodal knowledge extraction framework is proposed that integrates action recognition, speech semantic analysis, and text annotation to enable the structured extraction of procedural workflows, critical steps, and error patterns. Furthermore, an evaluation system is established by combining hierarchical rating scales, the Behavioral Anchored Rating Scale (BARS) approach, and the Analytic Hierarchy Process (AHP). The framework incorporates weight allocation and reliability and validity testing to ensure objectivity and consistency in assessment outcomes. The proposed multimodal knowledge extraction and quantitative evaluation system can significantly enhance the efficiency and scientific rigor of medical procedure video analysis, providing valuable technical support for standardized skills training and distance medical education.

Keywords

Medical procedure videos; Knowledge extraction; Multimodal analysis; Evaluation framework; Behavioral Anchored Rating Scale (BARS)

Introduction

Medical procedural skills constitute a core competency in clinical practice, and the quality of skills training directly influences patient safety and clinical outcomes. Traditional training approaches primarily rely on face-to-face demonstrations and subjective assessments, often resulting in uneven resource distribution and inconsistent evaluation standards. With the rapid advancement of digital technologies, medical procedure videos have become an increasingly important resource for teaching, learning, and performance assessment. However, their utilization remains largely dependent on manual observation and review, limiting the efficiency of knowledge acquisition and objective evaluation. Knowledge extraction aims to identify and structure

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

procedural workflows, critical steps, and potential errors from unstructured video data, thereby providing evidence-based support for instructional feedback and skills improvement. At the same time, a robust evaluation framework is required to standardize the assessment of procedural quality through quantifiable performance indicators. Nevertheless, medical procedure videos are inherently multimodal, integrating visual actions, spoken explanations, and textual information. Consequently, relying on a single analytical approach is often insufficient to capture the full spectrum of instructional and procedural information. Furthermore, the reliability and validity of evaluation systems are critical determinants of the credibility and effectiveness of skills training outcomes. In this context, developing efficient knowledge extraction methods and scientifically grounded evaluation frameworks for medical procedure videos has become an important research priority in medical education and clinical skills training. Such advancements can facilitate the standardization, objectivity, and scalability of procedural skills assessment while supporting innovative models of competency-based and remote medical education.

1. Research on Knowledge Extraction Methods for Medical Procedure Videos

1.1 Structured Analysis of Video Content

The first step in knowledge extraction is to establish a structured representation of the procedural workflow. Taking laparoscopic cholecystectomy as an example, the entire procedure can be decomposed into eight major stages: preoperative preparation, patient positioning, trocar insertion, abdominal exploration, identification of Calot's triangle, gallbladder dissection, hemostasis verification, and instrument retrieval. This hierarchical workflow decomposition not only captures the sequential relationships among procedural steps but also highlights critical stages that are particularly susceptible to technical errors. The identification of key procedural nodes is based on both standardized clinical guidelines and the potential risks associated with operative mistakes. To construct a clinically meaningful knowledge framework, ten experienced surgeons were invited to annotate 100 surgical videos based on their practical expertise. Through this process, a total of 32 critical procedural nodes were identified, including trocar insertion depth adjustment and electrocautery clipping of the cystic artery. These key nodes possess one or more of the following characteristics: (1) irreversible actions, such as tissue dissection; (2) procedures for which immediate corrective measures may be insufficient to prevent adverse outcomes, such as ongoing hemorrhage; and (3) technically demanding maneuvers that reflect surgical proficiency, such as intracorporeal laparoscopic knot-tying. To quantify the relative importance of each node, a node importance evaluation model was developed, where: Node Importance = Clinical Risk Index × Procedural Difficulty. Based on this model, critical procedural nodes can be ranked according to their significance for surgical safety and technical performance assessment. From a technical perspective, procedure segmentation can be achieved by integrating temporal and spatial feature analysis. First, optical flow algorithms are employed to extract motion vectors between adjacent video frames, generating spatio-temporal maps of instrument manipulation and surgical actions. Second, Hidden Markov Models (HMMs) are used to identify state transitions within the procedural sequence and detect key transition points. Finally, a domain-specific knowledge base is incorporated to infer subsequent procedural steps and refine segmentation boundaries generated by the algorithm through knowledge-guided correction mechanisms [1]. Experimental

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

results demonstrated that the proposed approach accurately identified critical procedural nodes in laparoscopic cholecystectomy videos, achieving an accuracy rate of 92.3%, significantly outperforming the conventional frame-difference method, which achieved an accuracy of 83.6%.

Medical procedure videos primarily contain three information modalities: visual, auditory, and textual. Each modality captures distinct aspects of the procedural environment. The visual modality conveys spatial orientation, instrument movements, and anatomical structures; the auditory modality captures equipment-generated sounds as well as spoken communication among clinicians; and the textual modality includes dialogue transcripts, procedural annotations, and patient-related documentation. Multimodal fusion aims to establish semantic associations across these heterogeneous data sources, enabling a comprehensive "see–hear–read" understanding of procedural activities. For visual–audio fusion, researchers have proposed a dual-stream neural network architecture enhanced with an attention mechanism. In this framework, the visual stream employs a 3D Convolutional Neural Network (3D-CNN) to extract spatiotemporal features from surgical videos, while the auditory stream utilizes a Long Short-Term Memory (LSTM) network to learn temporal patterns from instrument-generated sounds and environmental audio signals. A cross-attention module dynamically adjusts the relative importance of each modality according to procedural context. For example, during tissue dissection stages, greater weight is assigned to visual information, whereas audio signals become more influential during equipment placement and device manipulation phases. Experimental results demonstrated that this multimodal approach achieved an F1 score of 0.89 in surgical phase classification tasks, representing a 21.4% improvement over single-modality methods. For visual–text fusion, maintaining semantic consistency between procedural descriptions and video content is critical. First, pretrained biomedical language models such as BioBERT are employed to extract key medical concepts and surgical terminology, which are then organized into a hierarchical medical vocabulary. Subsequently, object detection and tracking techniques identify corresponding surgical instruments, anatomical structures, and critical regions within video frames, enabling the construction of semantic mappings between textual concepts and visual entities. For instance, the surgical term "cystic artery" can be temporally and spatially linked to the vascular structure appearing between 0.45 and 0.58 seconds in the video sequence. This semantic alignment strategy significantly enhances the accuracy of surgical content interpretation, increasing descriptive accuracy from 67.2% to 91.5%. To achieve comprehensive multimodal integration, a Graph Neural Network (GNN)-based framework can be employed. In this approach, a heterogeneous knowledge graph is constructed, where nodes represent entities such as surgical actions, instruments, and anatomical structures, while edges encode temporal, spatial, causal, and co-occurrence relationships. Through graph-based information propagation, knowledge from different modalities is integrated into a unified representation. In a laparoscopic cholecystectomy case study, visual features, audio signals, and textual annotations were simultaneously identified and linked in real time. Specifically, the system successfully associated the electrocautery hook trajectory, high-frequency electrosurgical sound signals, and the textual concept "gallbladder bed," achieving a multimodal matching accuracy of 84.6%.

1.2 Key Knowledge Element Extraction Techniques

Research on medical procedure videos has increasingly focused on the extraction of critical knowledge elements, which serve as the foundation for intelligent medical knowledge systems. Accurate tracking of surgical instruments, assessment of procedural compliance, and identification of unsafe or inappropriate actions

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

provide essential information for medical training, quality assurance, and clinical decision-making. Recent advances have shifted from single-modality approaches toward integrated multimodal frameworks, enabling multi-level structural analysis that incorporates visual, spatiotemporal, and semantic information.

The identification and tracking of surgical instruments constitute a central task in medical video analysis. Deep learning–based object detection models, such as YOLOv8 and Mask R-CNN, have demonstrated remarkable performance in real-time instrument recognition and classification, achieving detection accuracies exceeding 98.2% on the publicly available EndoVis benchmark dataset. In robotic laparoscopic microsurgery environments, researchers have developed a dual-stream instrument-tracking architecture in which the spatial stream utilizes RGB images to learn instrument appearance features, while the temporal stream extracts motion trajectories from optical flow fields. The outputs of both streams are integrated through an attention mechanism, effectively reducing tracking errors. Experimental results indicate that the average tracking deviation of surgical instruments during laparoscopic cholecystectomy can be controlled within 3.2 pixels [2]. In addition to instrument identification, increasing attention has been devoted to extracting the temporal and spatial characteristics of instrument movements. Surgical procedures can be decomposed into a series of spatial coordinates associated with temporal labels, thereby forming a three-dimensional spatiotemporal trajectory model. For example, during total joint replacement surgery, the system can continuously record operational parameters of surgical saws, including cutting depth and angular velocity, and compare them with predefined procedural templates. Such analyses enable objective evaluation of surgical performance and procedural consistency. More recently, Graph Neural Networks (GNNs) have been introduced to model interactions between surgical instruments and anatomical structures. By representing instruments, tissues, and their relationships as graph-based entities, GNNs can capture complex procedural dynamics and identify abnormal operating patterns. In transurethral resection of the prostate (TURP), for instance, GNN-based approaches have achieved an abnormal trajectory detection rate of 87.3%, demonstrating their effectiveness in identifying deviations from standard surgical workflows. The analysis of complication-related events involving multiple instruments represents another important application of intelligent video analytics. Multi-instrument collaborative analysis leverages instrument–action semantic mappings to identify and interpret cooperative interactions among surgical tools. In coronary artery bypass grafting procedures, for example, the system can simultaneously monitor the spatiotemporal relationships among electrosurgical devices, suction instruments, and puncture tools. When the interval between the operation of an electrosurgical instrument and a puncture device exceeds 0.8 seconds, the system automatically generates a standardized alert. Empirical studies have shown that such real-time feedback mechanisms can improve surgical team coordination scores by 21.4% [3].

The development of a standardized knowledge graph is a prerequisite for procedural compliance assessment. By integrating expert annotations with natural language processing (NLP) techniques, surgical guidelines and training materials can be transformed into structured knowledge databases. For example, the standard workflow for laparoscopic appendectomy can be decomposed into 23 procedural steps, 17 instrument usage specifications, and 9 critical surgical safety boundaries. Rule-based verification systems can then compare surgical video content against these predefined standards in real time. Studies have shown that such systems are capable of identifying non-compliant actions in laparoscopic cholecystectomy procedures, with approximately 83.6% of procedural deviations successfully detected. To address variations arising from patient-specific anatomical differences and surgeon-related factors, adaptive compliance assessment mech-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

anisms have been developed. Reinforcement learning–based frameworks dynamically adjust compliance thresholds according to individual anatomical characteristics, enabling more personalized evaluation. During thyroid surgery, this adaptive approach reduced the false identification rate of procedural violations to 6.3%, significantly improving assessment accuracy. Recent advances have further refined compliance evaluation by decomposing surgical videos into discrete action primitives and employing Hidden Markov Models (HMMs) to establish standard procedural sequences. This approach has achieved step-level verification accuracies of up to 92.1% in orthopedic procedures. To enhance robustness and reliability, multimodal assessment methods have also been introduced. These frameworks integrate visual information (e.g., instrument positioning), acoustic signals (e.g., electrocautery sound characteristics), and haptic data (e.g., instrument–tissue contact perception) into a three-dimensional evaluation model. In laparoscopic hernia repair procedures, such multimodal systems achieved a 91.5% detection rate for subtle procedural deviations, such as improper mesh placement angles, outperforming single-modality approaches by 27.8%. Furthermore, digital-twin technology has recently been incorporated into compliance databases, enabling the creation of virtual procedural standards that simulate different surgical environments through physics-based modeling and provide dynamic benchmarks for performance evaluation [4].

1.3 Knowledge Graph Construction Methods

With the rapid advancement of intelligent healthcare systems, knowledge graph technology has emerged as a critical approach for medical knowledge management and intelligent clinical applications. Through innovations in ontology modeling, semantic association mining, and dynamic knowledge maintenance, knowledge graphs facilitate the transformation of fragmented information stored across disparate databases into structured, interoperable, and machine-readable knowledge networks. The construction of a medical procedural ontology serves as the foundational framework for high-quality knowledge graphs. Developing a clinically meaningful conceptual system requires the integration of multidisciplinary knowledge, including anatomy, surgery, medical devices, and perioperative management. A hierarchical knowledge architecture is typically established based on four levels: surgical procedure type, operative workflow, instrument utilization, and complication management. This framework not only describes standard procedural steps but also provides guidance on optimal execution strategies, potential complications, and corresponding risk-control measures. Taking laparoscopic cholecystectomy as an example, the procedure can be decomposed into 23 standardized steps, including “insufflation and establishment of pneumoperitoneum,” “dissection of Calot’s triangle,” and “isolation of the cystic duct.” Each procedural step is further characterized by a set of attributes encompassing anatomical information, instrument parameters, risk levels, and other contextual data, resulting in a structured representation supported by more than 12 attribute categories. Such detailed semantic modeling provides a robust foundation for subsequent knowledge retrieval and semantic matching tasks. Recent developments have adopted cross-institutional ontology frameworks based on the Web Ontology Language (OWL) to enhance interoperability across healthcare organizations. Experimental evaluations conducted in five tertiary hospitals demonstrated a terminology-matching accuracy of 92.7%, representing a 38.6% improvement over conventional keyword-based retrieval approaches.

Semantic relationship extraction and association analysis constitute a critical component of knowledge graph construction. Leveraging natural language processing (NLP) and graph computing techniques, this pro-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

cess aims to identify, analyze, and represent latent semantic relationships embedded within medical texts [5]. Pretrained language models such as BERT enable the extraction of complex semantic relationships, including associations such as “laparoscopic surgery–uses–ultrasonic scalpel” and “cholecystitis–increases risk of–conversion to open surgery.” These relationships are subsequently represented as structured knowledge triples consisting of entities, attributes, and relations.

In the construction of a knowledge graph for joint replacement surgery, machine learning techniques were employed to identify hidden associations among clinical variables. The trained system discovered a significant correlation between femoral neck fracture type and prosthesis model selection (correlation coefficient = 0.76, $p < 0.01$). This finding was subsequently incorporated into the latest edition of the Orthopedic Surgery Practice Guidelines. Furthermore, the application of Dynamic Graph Neural Networks (D-GNNs) substantially improved relationship extraction performance, increasing accuracy by 21.4% compared with traditional rule-based approaches and achieving an F1-score of 89.3%. In addition, knowledge embedding techniques were utilized to transform unstructured textual information into vectorized representations, enabling efficient semantic reasoning and predictive modeling. In surgical complication prediction tasks, these approaches achieved a predictive performance score of 0.91, demonstrating their effectiveness in extracting clinically meaningful knowledge from large-scale medical data.

A dynamic knowledge updating mechanism is essential for maintaining the relevance and vitality of a medical knowledge graph. To continuously incorporate emerging evidence from diverse data sources, incremental learning strategies can be employed to facilitate real-time knowledge acquisition and model adaptation. In this study, a three-layer architecture consisting of a Data Lake, Feature Engineering Layer, and Model Iteration Layer was designed to support the integration and processing of heterogeneous data sources, including electronic health records, medical device data, and clinical research datasets.

Within a cardiovascular intervention knowledge graph, the system dynamically analyzed follow-up data from more than 2,000 coronary artery bypass grafting (CABG) cases and identified a significant inverse relationship between the use of new-generation drug-eluting stents and the risk of restenosis ($r = –0.63$). The timely discovery of this association contributed to revisions of relevant clinical guidelines within six months. Reinforcement learning–based update strategies were further implemented to ensure that the Knowledge Freshness Index (KFI) remained consistently above 0.85, representing an improvement of approximately 40 percentage points compared with conventional periodic updating approaches.

Moreover, federated learning techniques were introduced to enable synchronized knowledge graph updates across multiple healthcare institutions while preserving patient privacy and data security. This collaborative framework significantly enhanced rare disease diagnostic performance, increasing diagnostic accuracy from 62% to 89%. These advances highlight the potential of dynamic, continuously evolving knowledge graphs to support evidence-based clinical decision-making and intelligent medical knowledge management.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

2. Construction of an Evaluation System for Medical Procedure Videos

2.1 Principles for Designing the Evaluation Index System

Clinical procedure videos serve as an important medium for medical education, surgical skills training, and professional communication. The quality of these videos directly influences learners' understanding of surgical procedures, as well as the safety and effectiveness of their subsequent clinical practice. Therefore, the development of an objective, scientific, and standardized evaluation system is fundamental to improving the quality of clinical procedure videos and advancing medical education.

A well-designed evaluation framework should provide comprehensive and reliable criteria for assessing the educational and technical value of procedural videos. As the cornerstone of video quality assessment, such a framework must adhere to four essential principles: scientific validity, comprehensiveness, feasibility, and practical applicability [6]. These principles ensure that the evaluation system accurately reflects procedural quality, supports standardized training, and facilitates the effective dissemination of clinical knowledge.

Scientific validity is the foremost principle in the design of a medical procedure video scoring system. It means that the evaluation criteria within the scoring framework must be based on domain-specific medical knowledge and scientific laws, accurately and objectively reflecting the characteristics and quality of content in medical procedure videos. From a medical perspective, this requires that the entire operative process—including all procedural steps and actions—be taken into account. Therefore, when defining evaluation criteria, the scope should comprehensively cover preoperative preparation (such as patient assessment, surgical planning, and instrument preparation), intraoperative techniques (such as incision location, tissue dissection, and hemostasis methods), and postoperative management (such as wound care and complication prevention). These standards should be grounded in foundational medical disciplines, including anatomy, physiology, and pathology, to ensure the scientific validity and correctness of the evaluation indicators.

The principle of systematic integrity emphasizes the establishment of a comprehensive and interrelated evaluation indicator system, in which assessment factors complement and reinforce one another. Medical procedure videos involve multiple dimensions, including surgical techniques, procedural compliance, patient safety, and imaging quality; therefore, the evaluation system must cover all these aspects to avoid bias and partial assessment. For example, when evaluating a clinical examination video, it is insufficient to assess only accuracy or technical proficiency; it is also necessary to consider whether the procedure complies with operational standards, whether patient comfort and safety are ensured, and whether camera angles and image quality clearly present the examination process.

The principle of operability requires that the evaluation scale be practically feasible in application. That is, the assessment criteria should be easy to understand, easy to compute, and easy to apply in real evaluation settings, thereby ensuring their effectiveness. Each indicator should have a clear definition and description to avoid vague or ambiguous concepts. For instance, for the indicator "proficiency," its meaning should be explicitly defined in terms of observable characteristics such as fluency of movement, coordination of posture, and time control, so that evaluators can accurately understand and consistently apply the criterion.

According to the principle of dynamism, evaluation standards for medical procedure videos should be con-

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

tinuously updated in response to the ongoing development of medical technologies and medical education. As surgical techniques evolve, new operative methods, skills, and strategies emerge, leading to corresponding changes in the content of medical procedure videos [7]. Therefore, the evaluation system must possess sufficient flexibility to accommodate such changes and enable timely updates and refinement of assessment criteria. For example, with the emergence of new minimally invasive surgical techniques, corresponding evaluation standards must be developed, including technical requirements for the use of minimally invasive instruments and standardized procedural protocols.

In summary, the formulation of evaluation standards for medical procedure videos should follow four fundamental principles: scientific validity, systematic integrity, operability, and dynamism. Only by adhering to all four principles can a scientific, comprehensive, practical, and flexible evaluation framework be established, thereby supporting high-quality assessment and improvement of medical procedure videos and ultimately contributing to the advancement of medical education and clinical practice.

2.2 Multi-dimensional Evaluation Index Framework

In the evaluation system of medical procedure videos, constructing a multi-dimensional evaluation index framework is of critical importance, as it enables a comprehensive and detailed assessment of the quality and value of medical procedure videos, thereby providing strong evidence for medical education, clinical practice, and the optimization of operational standards. The following sections elaborate on the technical operation dimension, clinical decision-making dimension, and communication and collaboration dimension.

Technical operation is the core of medical procedures and is directly related to patient treatment outcomes and safety. In the evaluation of medical procedure videos, the technical operation dimension includes three key aspects: accuracy, fluency, and aseptic principles.

Accuracy is the lifeline of technical operation. It requires the operator to precisely master every procedural step and every anatomical structure. Taking surgical procedure videos as an example, from incision selection to tissue dissection, and further to the management of critical structures such as blood vessels and nerves, all steps must strictly follow anatomical principles and operational standards. Any minor deviation may lead to serious complications and affect patient prognosis. For example, in cardiac surgery, during coronary artery anastomosis, inaccurate suturing may result in myocardial ischemia or even necrosis. Therefore, the operations demonstrated in the video must ensure a high level of accuracy, which is the primary criterion for evaluating quality.

Fluency reflects the operator's technical proficiency and the continuity of the procedure. A smoothly performed operation allows viewers to clearly understand the entire procedural workflow and perceive the natural and uninterrupted progression of the operation. Fluency indicates that the operator has a solid grasp of technical key points and is able to complete each step without interruption or hesitation.

2.3 Quantitative Methods for Evaluation Criteria

Therefore, in the construction of a medical procedure video evaluation system, it is necessary to pay attention not only to the scientific rigor of the evaluation criteria but also to their feasibility. The application of

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

graded rating scales provides a strong technical foundation for the objective measurement of medical education quality.

During the design process, a graded rating scale can be used to transform difficult-to-evaluate content into measurable indicators, enabling a scientifically grounded assessment of procedural videos. To meet the requirement of "precise measurement," a 5–9 point scale is typically adopted to ensure clear distinctions between levels and to align with human cognitive patterns. For example, in evaluating "plan feasibility," a 7-point scale can be designed ranging from "this plan is highly scientific, closely integrates the clinical case, and demonstrates strong practical applicability, allowing effective implementation of learned knowledge" (highest score 17.5) to "this plan is completely unrelated to the case itself and is entirely imaginative" (lowest score 2.5), with a gradually refined grading process rather than a simple binary classification, thereby using intermediate levels to capture detailed procedural differences.

To ensure completeness and conciseness, attention must be paid to the structure of the scale. For example, the "technical operation dimension" can be further decomposed into hierarchical indicators such as operational accuracy, operational proficiency, and aseptic principles. Specific behavioral criteria can then be defined for third-level evaluation standards. For instance, judgments of aseptic principles may include behaviors such as "whether hand hygiene is performed correctly" and "whether instrument transfer complies with aseptic requirements." The development of such a scale must undergo expert validation and revision after preliminary testing to ensure both exclusiveness and inclusiveness of each level's connotation, avoiding distortion of evaluation data caused by interpretive ambiguity.

The determination of weight distribution should integrate both subjective and objective factors. For example, the Analytic Hierarchy Process (AHP) can be used, in which expert pairwise comparisons are conducted to construct a judgment matrix. Taking the "clinical decision-making dimension" as an example, expert scoring can determine that the weighting ratio between indication selection and risk control is 3:2. For the entropy weighting method, weights are determined based on the degree of data variability: if violations of the "aseptic principle" appear significantly more frequently than other indicators in a hospital's videos, the weight of this indicator will be automatically increased.

3 Conclusion

This study constructs a multimodal knowledge extraction framework and a quantitative evaluation system, enabling efficient analysis and objective assessment of medical procedure videos. Experimental results show that the proposed framework achieves high performance in procedural step recognition accuracy, evaluation consistency, and weight rationality, providing a feasible solution for the standardization and remote implementation of medical skills training. Research on knowledge extraction and evaluation systems for medical procedure videos represents a key component in the digital transformation of medical education. With technological iteration and data accumulation, quantitative evaluation will gradually replace subjective judgment, driving medical training toward greater precision and personalization, ultimately improving the safety and effectiveness of clinical practice.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Funding

This research was supported by the Education and Teaching Reform Research and Practice Project of Zhengzhou University (International Student Education Special Program): Development and Application of an International Perspective–Based Evaluation Standard for Practical Clinical Skills of International Medical Students (Grant No. 2023ZZUJGXM-LXS003).

References

[1]Jiao, S. F., Xu, Y. C., Zhang, X. H., et al. (2025). Application of video feedback teaching in the training of basic surgical skills for medical students. Medical Education Management, 11(1), 57–61.

[2]Xu, Y. (2021). Application of video-recorded operative methods in “Three Basic Skills” training in intensive care medicine. General Practice Nursing, 19(34), 4892–4893.

[3]Duan, X. X., Yuan, Y., & Ying, X., et al. (2025). The impact of operation video combined with workshop teaching on clinical interns’ skill performance. Journal of Bengbu Medical University, 50(5), 694–697.

[4]Xue, L., Liu, B. Y., Xia, D. D., et al. (2025). Creation and teaching exploration of medical professional animation videos under the context of new medical science. Journal of Mudanjiang Medical University, 46(1), 163–166.

[5]Su, R. B., Yang, B., & Jia, Y. W., et al. (2025). Application of PBL teaching method and video-based operation teaching in microsurgery clinical education. Continuing Medical Education, 39(3), 65–68.

[6]Zhang, H. L., Dong, X. B., & Wang, C. L., et al. (2024). Application of micro-lecture videos combined with flipped classroom in undergraduate surgical skills teaching. China Continuing Medical Education, 16(14), 1–7.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)