International Journal of New Findings in Health and Educational Sciences (IJHES)
ISSN 2980-0641
DOI: 10.63053/ijhes.183

Artificial Intelligence–Based Approaches for the Assessment of Surgical Skills: A Systematic Review

Simin Raissi[1], MohamadTaher Rezanejad[*2]
1. Shahid Behehshti University of Medical Sciences, Tehran, Iran
2.Department of Nursing, School of Nursing and Midwifery, Nursing and Midwifery Care Research Center, Non-Communicable Diseases Research Institute, Shahid Sadougi University of Medical Sciences and Health Services, Yazd, Iran-
*corresbonder author, Department of Nursing, School of Nursing and Midwifery, Nursing and Midwifery Care Research Center, Non-Communicable Diseases Research Institute, Shahid Sadougi University of Medical Sciences and Health Services, Yazd, Iran

ARTICLE INFO

Keywords: *Artificial Intelligence, Surgical Skills Assessment, Machine Learning, Deep Learning, Surgical Education, Simulation-Based Training, Surgical Performance Evaluation, Computer-Assisted Surgery, Surgical Decision Support, Operative Competency*

ABSTRACT

Background: In recent years, artificial intelligence (AI) has emerged as an innovative tool for surgical training and skill assessment. The expansion of simulation-based methods and robotic technologies has enabled the collection of motion, video, and multimodal data, which can be leveraged by AI models to analyze and classify surgeons' skills. However, the existing evidence is scattered and heterogeneous, highlighting the need for a systematic review to evaluate applications, benefits, limitations, and future perspectives. This study aims to provide a systematic review of AI applications in surgical skill assessment, analyze algorithm performance, examine educational feedback, and identify current limitations and future research opportunities in this field.
Methods: A systematic search was conducted across major scientific databases, including PubMed, Scopus, Web of Science, Springer, Nature, JAMA Network, and ScienceDirect. Studies published between 2018 and 2026 that focused on AI applications in surgical skill assessment were included. Data were extracted regarding input data type, algorithm type, performance accuracy, feedback applications, and model validation. Both qualitative and quantitative analyses were performed.
Results: Based on 60 identified studies, AI models demonstrated the capability to analyze motion, video, and multimodal data, achieving surgeon skill classification accuracies ranging from 80% to 95%. The use of personalized and real-time feedback enhanced skill acquisition and facilitated transfer learning. Algorithms were also applied for surgical phase recognition, risk prediction, and clinical decision support. Nevertheless, limitations such as restricted multicenter validation, lack of standardized data collection protocols, and low interpretability of certain models were reported.
Conclusion: AI has the potential to play an effective role in surgical education, skill assessment, and clinical decision support, complementing human instructors in the operating room. To fully exploit AI's potential, the development of standardized frameworks, multicenter validation, increased algorithm transparency, and assessment of long-term educational impacts are essential. Future research should focus on integrating real-world and simulation data and evaluating the transferability of models to practical clinical environments.

1. Introduction

Assessment of surgical technical skills is recognized as a key component in ensuring clinical performance quality and patient safety. A surgeon's technical proficiency not only correlates with patient outcomes but also plays a critical role in competency-based training and the development of future surgeons (1, 2). Traditional assessment methods, such as direct observation, standardized tools like OSATS, and general rating instruments, despite their widespread use, face inherent limitations, including lack of reproducibility, evaluator bias, and high demands on time and human resources (3–5). In recent years, advances in digital technologies and the increasing availability of high-quality surgical video data have created new opportunities for quantitative analysis of surgeon performance. Artificial intelligence (AI), particularly machine learning and deep learning techniques, has gained attention as a novel approach for objective assessment of surgical skills due to its ability to extract complex features from multidimensional data (6–8). Computer vision and motion analysis based on kinematic data of surgical instruments enable precise evaluation of surgeons' behavior and technique. These methods can identify skill-related features such as motion smoothness, targeting accuracy, and tool usage patterns, converting them into measurable metrics (9–11). Early studies have demonstrated that AI models can assess surgical skills with accuracy comparable to human evaluators, and in some cases even outperform them (7, 12–15). Despite these advances, several fundamental challenges hinder the widespread development and application of AI-based methods. A major challenge is the heterogeneity and lack of standardization in study designs, performance metrics, and datasets, which complicates direct comparisons between methods and limits the generalizability of findings (16–18). Additionally, many existing studies rely on small, single-center datasets and lack external validation, increasing the risk of bias and reducing the robustness of the results (7, 19). Some recent studies have focused on AI applications in specific surgical domains, such as laparoscopic surgery, robotic surgery, and simulation. For instance, in laparoscopic surgery, convolutional neural networks (CNNs) have been employed to analyze surgical videos, achieving high accuracy in skill-level classification (20, 21). In robotic surgery, analysis of instrument motion data combined with reinforcement learning algorithms has enabled the identification of complex performance patterns (15, 17, 22, 23). Furthermore, several studies have explored auxiliary applications such as real-time feedback and error prediction, which can play a significant role in enhancing patient safety (24–28). From an educational perspective, AI-based tools not only facilitate rapid and objective skill assessment but also provide quantitative feedback that can enhance training and learning. This capability is particularly valuable in simulation and robotic training environments, where large volumes of data are generated (29–33). However, concerns regarding data privacy, ethical challenges, and legal considerations related to the use of patient and surgeon data must also be carefully addressed (34–36).

Given the rapid and exponential growth of scientific literature on the application of artificial intelligence in surgical skill assessment, a substantial body of research has focused on machine learning, deep learning, computer vision, and motion data analysis. These studies not only demonstrate a wide variety of technological approaches but also exhibit significant differences in study design, evaluation metrics, surgical domains, and data quality, making direct comparisons and definitive conclusions challenging. Therefore, a systematic review that can cohesively and comprehensively synthesize the published evidence is essential. The present systematic review was conducted to provide a structured and scientific framework with several key objectives: first, to analyze and categorize AI methods and algorithms used for surgical skill assessment; second, to identify the advantages and strengths of each approach, including accuracy, reproducibility, and the ability to provide real-time feedback; third, to examine existing limitations and challenges, such as data scarcity, lack of external validation, methodological heterogeneity, and ethical and legal considerations; and finally, to outline future perspectives and recommendations, with a focus on standardization, improved validation, and integration of AI into educational and clinical settings. By emphasizing the synthesis of current evidence, this review aims to serve as a scientific reference and practical guide for researchers, medical educators, and AI technology developers, facilitating the more efficient, accurate, and standardized application of AI in surgical skill assessment. Moreover, this review can help identify research gaps and weaknesses, paving the way for future studies and the development of AI-based frameworks with potential for clinical implementation.

Methods

This study was designed as a systematic review to comprehensively synthesize and analyze the existing evidence on the application of artificial intelligence (AI) for surgical skill assessment. The study design followed the PRISMA 2020 guidelines to ensure transparency, reproducibility, and standardization of the review process. A systematic search was conducted across PubMed, Scopus, Web of Science, IEEE Xplore, and Embase, covering all articles published from January 2018 to February 2026. The search strategy included a combination of keywords related to artificial intelligence, machine learning, deep learning, computer vision, motion analysis, and surgical skills, with language restricted to English. Reference lists of selected studies were also screened to identify potentially missing studies. Inclusion criteria comprised original research studies that utilized AI techniques to assess surgical skills, reported quantitative performance data of the AI models, and were conducted in clinical or surgical simulation settings. Exclusion criteria included narrative reviews, theoretical papers, conference reports without full text, non-surgical studies, and studies lacking AI performance data. Study selection began with independent screening of titles and abstracts by two reviewers, followed by full-text assessment. Discrepancies between reviewers were resolved through discussion, and, if necessary, a third reviewer was involved for final decisions. Data extracted included author(s) and year of publication, country of study, type of surgery and educational context, AI algorithm or model used, data source and sample size, model performance metrics (e.g., accuracy, AUC, F1 score, sensitivity, specificity), comparisons with human assessment when available, and type of validation performed. Data extraction was conducted independently by two researchers, with all data recorded in a standardized form to ensure accuracy. Study quality and risk of bias were assessed according to study design: diagnostic or predictive studies were evaluated using QUADAS-2, non-randomized and observational studies using ROBINS-I, and randomized trials using the Cochrane Risk of Bias Tool. This allowed for assessment of evidence quality and weighting of study results. Data analysis was primarily qualitative and comparative, examining AI models in terms of accuracy, input data type, and surgical domain. Additionally, limitations, research gaps, and opportunities for AI development in surgical skill assessment were identified and reported to provide a scientific and practical framework for future research.

Results and Discussion

In this study, a total of 60 studies were evaluated. The findings were qualitatively categorized as follows.

Types of Data and Platforms Used for AI-Based Surgical Skill Assessment

Existing research indicates that artificial intelligence has been applied to assess surgical skills using a variety of data types, including video recordings, kinematic motion data, virtual reality (VR) simulations, and biometric measurements. Among the identified studies, the majority focused on analyzing instrument motion, surgeons' movement trajectories, and surgical video images. The data platforms used for training and validating AI models in surgical skill assessment were categorized into three main types:

a) Motion and Kinematic Data

Several studies utilized instrument motion data and surgeons' kinematic trajectories to train AI algorithms. These data are typically obtained from robotic systems or virtual reality simulations (31, 37–48). For instance, Winkler-Schwartz and colleagues employed datasets extracted from the NeuroVR simulation platform, comprising over 270 performance metrics, which were subsequently used to develop machine learning models (37). Such datasets enable complex and in-depth learning of skill patterns and have been reported in many studies as the primary input for ML/DL models (29, 37–48).

b) Real Surgical Video Images

The second category of data includes recorded videos of actual surgical procedures, which are analyzed using computer vision algorithms (40, 43, 46, 49–54). For example, Pedrett et al., in a systematic review of convolutional neural network (CNN) applications on stored surgical videos, demonstrated that computer vision models can accurately identify skill levels from video streams and even correctly classify surgical phases (38). Similarly, other studies have shown that AI can extract key regions from surgical scenes and use them as the basis for feedback analysis (40, 51).

c) Multimodal (Combined) Data

Some researchers have integrated multiple types of data to enhance model performance, such as combining video with kinematic and biometric data (38, 40, 49). This multimodal approach has improved predictive power, model robustness, and reduced error distribution.

Based on this review, motion and simulation data constitute the most fundamental sources for training AI models in skill assessment. However, for real-world applications and transfer to the operating room, multimodal data integration has demonstrated the best performance.

Accuracy and Performance of AI Algorithms in Surgical Skill Classification

According to the available evidence, AI models have demonstrated remarkable performance in detecting and classifying surgeons' skill levels. Many approaches have employed deep neural networks, classification models, and computer vision systems, reporting high levels of accuracy.

a) Model Accuracy in Simulation

Multiple studies have shown that AI algorithms can accurately identify skill levels in simulated environments. In a randomized trial, Fazlollahi et al. demonstrated that AI models significantly improved Expertise Scores and enabled quantitative analysis (37). Other studies reported that machine learning algorithms achieved accuracies of 85–95% in classifying novices and experts in simulation tasks (39, 42, 55). Therefore, AI models perform very reliably in detecting skill differences within controlled simulation settings.

b) Model Accuracy with Real Surgical Data

Using real surgical data for skill assessment is inherently more complex; however, leading studies have reported positive results. Pedrett et al., in their systematic review, indicated that when models were trained on real-world data, classification accuracies of over 80% were still achieved (38). Furthermore, evidence suggests that deep neural networks can accurately identify surgical phases and even predict intraoperative risk (49).

c) Analysis of Other Performance Metrics

Beyond accuracy, many studies employed more sophisticated statistical metrics such as F1-score, AUC-ROC, precision, recall, specificity, and balanced accuracy to evaluate model performance (38, 43, 49, 51). These metrics indicate that multilayer and deep learning models not only perform well in basic skill classification but also in predicting complex motion phases.

Overall, AI algorithms demonstrate higher accuracy in simulation environments, but even when applied to real-world data, advanced performance metrics remain acceptable and promising.

The Role of AI Feedback in Enhancing Learning and Skill Transfer

One of the most consistent findings in the literature is the positive impact of AI-based feedback systems on the acquisition and consolidation of surgical skills.

a) Effects of AI Training in Simulation

The study by Fazlollahi et al. demonstrated that the use of AI feedback in simulation not only directly improved skill scores but also facilitated transfer learning (37). Participants who received AI feedback performed better in more complex tasks compared to control groups and traditional human-led training.

b) Real-Time and Personalized Feedback

Many AI systems are capable of providing real-time, individualized feedback that enhances gradual and targeted learning (37, 39, 42, 49, 52, 55). This type of feedback clearly highlights performance differences between individuals, enabling learners to quickly identify and correct technical errors.

c) Long-Term and Cognitive Effects of AI Feedback
Some studies have reported that AI feedback not only improves technical skills but also enhances learners' cognitive efficiency and confidence (39, 42, 52). These findings suggest that AI can function as an "intelligent coach" throughout the learning process.
Overall, based on current evidence, AI feedback represents one of the most successful applications in surgical skills training, enabling faster and more effective improvement of technical abilities compared to traditional methods.

Applications of AI in Intraoperative Decision Support and Performance
Artificial intelligence in surgery is not limited to training and skill assessment; it also has practical applications in clinical decision support during operations and enhancing patient safety.
a) Surgical Phase Recognition and Procedural Support
Several studies have demonstrated that AI models can recognize surgical phases and even assist surgeons in anticipating the next steps of a procedure (49, 51, 54). This application has the potential to reduce technical errors in real operating rooms and serve as a decision-support system during critical moments of surgery.
b) Prediction of Postoperative Complications and Outcomes
Broader studies have shown that AI can predict the risk of postoperative complications and patient outcomes, thereby aiding preoperative decision-making (39, 49). This capability contributes to improved patient management and reduced postoperative care costs.
Integration of AI with Real and Robotic Surgery
Research by Massimino et al. and other sources has indicated that AI in robotic surgery can assist in analyzing instrument trajectories, classifying surgical phases, and enhancing skill transfer (56). The application of computer vision and deep learning in real operating room settings enables surgical phase recognition and provision of technical feedback. Furthermore, recent studies have shown that AI can assess surgeon performance in real surgical environments with accuracy comparable to simulation, laying the groundwork for the development of intraoperative decision-support systems (57–59).

Challenges and Limitations of AI in Surgical Skill Assessment
The primary advantage highlighted in the reviewed studies is that AI can provide objective, reproducible, and precise feedback, which is often limited in traditional assessment methods (31, 37, 52). However, comprehensive reviews indicate that although AI demonstrates high performance in skill recognition and feedback, fundamental challenges remain regarding validation, generalizability, and clinical applicability. Key limitations include the need for standardized datasets, robust computational infrastructure, ethical and legal considerations, and the generalizability of algorithms to real clinical environments (27, 40, 52). A recent study emphasized that the lack of multicenter and standardized data is one of the greatest barriers to developing AI models in surgery (40).
Additionally, Leon et al. warned that without human oversight and an appropriate regulatory framework, the use of AI in training and assessment may result in bias or erroneous predictions (52).
These challenges can be categorized as follows:
a) Validation and Generalizability
Many studies have been validated only in single-center or simulated environments, with limited multicenter validation (38, 39, 49, 53). This restricts the generalizability of results to broader populations of surgeons.
b) Data Quality and Standardization
The lack of global standards for collecting and annotating surgical training data has made data aggregation and cross-center comparison difficult, representing a major obstacle to developing generalizable AI models (39, 49, 53).
c) Model Explainability
Although many deep learning models demonstrate strong performance, they often cannot provide interpretable reasoning for their judgments in a way that is clinically actionable or understandable to

surgeons (39, 54). This presents practical, ethical, and legal challenges for the widespread adoption of AI in surgical education and assessment.

Future Perspectives

Based on these findings, it can be concluded that artificial intelligence holds transformative potential in surgical education and skill assessment, serving as a complementary tool to human instructors and intraoperative decision-support systems. However, the development of standardized frameworks, multicenter validation, and increased algorithm transparency are essential for widespread clinical adoption. Integrating AI with advanced simulations and real-world surgical data will shape the future direction of research and application in surgical training and patient safety improvement (37, 38, 40, 49, 54). Recent studies predict future advancements in two main areas: the development of fundamental multimodal models capable of processing motion, image, and video data, and the integration of AI into real surgical environments for decision support and resident training (22, 33, 40, 52, 57–60). Moreover, a recent review emphasizes that combining AI with traditional training can substantially enhance the quality of learning and skill assessment, establishing AI as an essential complementary component in surgical education (52).

Conclusion

Artificial intelligence, as an innovative tool in surgical skill assessment, simulation-based training, and clinical decision support, has demonstrated the potential to fundamentally transform the quality of education and patient safety. Findings from this systematic review indicate that machine learning and deep learning models are capable of analyzing motion, video, and multimodal data to accurately classify surgeons' skill levels and provide personalized, real-time feedback. These capabilities position AI as an "intelligent coach" and decision-support system within the surgical training and operative workflow. Nevertheless, limitations such as restricted multicenter validation, data quality and standardization issues, and challenges related to model interpretability require careful attention for widespread clinical adoption. To fully leverage AI's potential, the development of standardized data collection frameworks, multicenter validation, and enhanced algorithm transparency are essential. Overall, AI, with its ability to provide quantitative assessment, targeted feedback, and clinical decision support, can complement human instructors and has the potential to transform the future of surgical education and patient safety. Future research should focus on integrating real-world and simulation data, evaluating long-term educational impacts, and assessing the transferability of models to actual clinical settings to ensure that AI can effectively support both surgical training and operative performance.

Acknowledgments

The authors wish to express their sincere gratitude to all individuals who guided and supported the preparation of this manuscript.

Ethical Considerations

This article is a review study and did not involve any human subjects. The authors declare that this report contains no personal information that could lead to the identification of any patient. In reviewing the literature and citing articles included in the reference list, honesty and integrity were maintained, and no unreliable or unverified sources were used. The most appropriate research methods and the latest available techniques were employed.

Ethics Code

Not applicable.

Conflict of Interest

The authors declare no financial or personal conflicts of interest that could have influenced the results or interpretation of this study. None of the authors have affiliations with organizations or entities that might benefit or suffer from the outcomes of this research.

Funding
This work did not receive any financial support from governmental, commercial, or non-profit funding agencies.

Authors' Contributions
Mohammad Taher Rezanejad: Oversight of literature searches and supervision of the manuscript preparation process.
Simin Raeisi: Collaboration in study synthesis and manuscript writing, conducting database and search engine queries, and compiling article records in Excel.
All authors contributed to the drafting and editing of the manuscript and have read and approved the final version.

Refrences
1. Hasan OH, Ayaz A, Khan M, Docherty C, Hashmi P. The need for simulation in surgical education in developing countries. The wind of change. Review article. The Journal of the Pakistan Medical Association. 2019;69(Supl. 1):S62.
2. Hussein N, Van den Eynde J, Callahan C, Guariento A, Gollmann-Tepeköylü C, Elbatarny M, et al. The use of objective assessments in the evaluation of technical skills in cardiothoracic surgery: a systematic review. Interactive CardioVascular and Thoracic Surgery. 2022;35(3):ivac194.
3. Asif H, McInnis C, Dang F, Ajzenberg H, Wang PL, Mosa A, et al. Objective structured assessment of technical skill (OSATS) in the surgical skills and technology elective program (SSTEP): comparison of peer and expert raters. The American Journal of Surgery. 2022;223(2):276-9.
4. Lewandrowski K-U, Bergamashi JP, Telfeian AE, de Carvalho P, Leon J. Training and credentialing standards for minimally invasive spinal surgery techniques: Results of a survey. Pain Physician. 2023;26(1):29-37.
5. Youssef SC, Aydin A, Canning A, Khan N, Ahmed K, Dasgupta P. Learning surgical skills through video-based education: a systematic review. Surgical Innovation. 2023;30(2):220-38.
6. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. Annals of surgery. 2018;268(1):70-6.
7. Lam K, Chen J, Wang Z, Iqbal FM, Darzi A, Lo B, et al. Machine learning for technical skill assessment in surgery: a systematic review. NPJ digital medicine. 2022;5(1):24.
8. Zhou X-Y, Guo Y, Shen M, Yang G-Z. Application of artificial intelligence in surgery. Frontiers of medicine. 2020;14(4):417-30.
9. Nwoye CI. Deep learning methods for the detection and recognition of surgical tools and activities in laparoscopic videos: Université de Strasbourg; 2021.
10. Ahmidi N, Tao L, Sefati S, Gao Y, Lea C, Haro BB, et al. A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery. IEEE Transactions on Biomedical Engineering. 2017;64(9):2025-41.
11. Hashimoto DA, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. Annals of surgery. 2019;270(3):414-21.
12. Caballero D, Sánchez-Margallo JA, Pérez-Salazar MJ, Sánchez-Margallo FM. Applications of artificial intelligence in minimally invasive surgery training: a scoping review. Surgeries. 2025;6(1):7.
13. Kankanamge D, Wijeweera C, Ong Z, Preda T, Carney T, Wilson M, et al. Artificial intelligence based assessment of minimally invasive surgical skills using standardised objective metrics–A narrative review. The American Journal of Surgery. 2025;241:116074.
14. Boal MW, Anastasiou D, Tesfai F, Ghamrawi W, Mazomenos E, Curtis N, et al. Evaluation of objective tools and artificial intelligence in robotic surgery technical skills assessment: a systematic review. British Journal of Surgery. 2024;111(1):znad331.
15. Kutana S, Bitner DP, Addison P, Chung PJ, Talamini MA, Filicori F. Objective assessment of robotic surgical skills: review of literature and future directions. Surgical endoscopy. 2022;36(6):3698-707.
16. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. bmj. 2020;368.
17. Madani A, Liu Y, Pryor A, Altieri M, Hashimoto DA, Feldman L. SAGES surgical data science task force: enhancing surgical innovation, education and quality improvement through data science. Surgical Endoscopy. 2024;38(7):3489-93.
18. Andreatta P, Smith CS, Graybill JC, Bowyer M, Elster E. Challenges and opportunities for artificial intelligence in surgery. The Journal of Defense Modeling and Simulation. 2022;19(2):219-27.
19. Khalid S, Goldenberg M, Grantcharov T, Taati B, Rudzicz F. Evaluation of deep learning models for identifying surgical actions and measuring performance. JAMA network open. 2020;3(3):e201664.
20. Funke I, Mees ST, Weitz J, Speidel S. Video-based surgical skill assessment using 3D convolutional neural networks. International journal of computer assisted radiology and surgery. 2019;14(7):1217-25.
21. Pan M, Wang S, Li J, Li J, Yang X, Liang K. An automated skill assessment framework based on visual motion signals and a

deep neural network in robot-assisted minimally invasive surgery. Sensors. 2023;23(9):4496.
22. Lavanchy JL, Zindel J, Kirtac K, Twick I, Hosgor E, Candinas D, et al. Automation of surgical skill assessment using a three-stage machine learning algorithm. Scientific reports. 2021;11(1):5197.
23. Lukács E, Levendovics R, Haidegger T. Enhancing autonomous skill assessment of robot-assisted minimally invasive surgery: A comprehensive analysis of global and gesture-level techniques applied on the JIGSAWS dataset. Acta Polytech Hung. 2023;20(8):133-53.
24. Anh NX, Nataraja RM, Chauhan S. Towards near real-time assessment of surgical skills: A comparison of feature extraction techniques. Computer methods and programs in biomedicine. 2020;187:105234.
25. Ben Hmido S, Abder Rahim H, Keller B, Daams F, Schakel M, Goslings JC, et al. Ethical Pitfalls in AI-Based Predictive Models in Surgery. World Journal of Surgery. 2025;49(10):2837-45.
26. Saravi B, Hassel F, Ülkümen S, Zink A, Shavlokhova V, Couillard-Despres S, et al. Artificial intelligence-driven prediction modeling and decision making in spine surgery using hybrid machine learning models. Journal of Personalized Medicine. 2022;12(4):509.
27. Raghavan N, Patel PC, Limon D, Morris MX, Kempenich JW, Rajesh A. Artificial intelligence in surgical education: a 2025 update on adaptive training, feedback, and competency-based education. The American Surgeon™. 2026;92(3):675-86.
28. Zuluaga L, Rich JM, Gupta R, Pedraza A, Ucpinar B, Okhawere KE, et al., editors. AI-powered real-time annotations during urologic surgery: The future of training and quality metrics. Urologic Oncology: Seminars and Original Investigations; 2024: Elsevier.
29. Ogbonnaya CN, Li S, Tang C, Zhang B, Sullivan P, Erden MS, et al., editors. Exploring the role of artificial intelligence (AI)-driven training in laparoscopic suturing: a systematic review of skills mastery, retention, and clinical performance in surgical education. Healthcare; 2025: MDPI.
30. Ledwos N, Mirchi N, Yilmaz R, Winkler-Schwartz A, Sawni A, Fazlollahi AM, et al. Assessment of learning curves on a simulated neurosurgical task using metrics selected by artificial intelligence. Journal of neurosurgery. 2022;137(4):1160-71.
31. Winkler-Schwartz A, Bissonnette V, Mirchi N, Ponnudurai N, Yilmaz R, Ledwos N, et al. Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation. Journal of surgical education. 2019;76(6):1681-90.
32. Bourdillon AT, Garg A, Wang H, Woo YJ, Pavone M, Boyd J. Integration of reinforcement learning in a virtual robotic surgical simulation. Surgical innovation. 2023;30(1):94-102.
33. Guni A, Varma P, Zhang J, Fehervari M, Ashrafian H. Artificial intelligence in surgery: the future is now. European Surgical Research. 2024;65(1):22-39.
34. Arjomandi Rad A, Vardanyan R, Athanasiou T, Maessen J, Sardari Nia P. The ethical considerations of integrating artificial intelligence into surgery: a review. Interdisciplinary Cardiovascular and Thoracic Surgery. 2025;40(3):ivae192.
35. Khan MA. Ethical implications of AI (artificial intelligence) in healthcare mainly focus on surgical procedures: identification of ethical issues of AI in surgical procedures and ranking of ethical issues based on criticality. 2024.
36. Collins JW, Marcus HJ, Ghazi A, Sridhar A, Hashimoto D, Hager G, et al. Ethical implications of AI in robotic surgical training: a Delphi consensus statement. European urology focus. 2022;8(2):613-22.
37. Fazlollahi AM, Bakhaidar M, Alsayegh A, Yilmaz R, Winkler-Schwartz A, Mirchi N, et al. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: a randomized clinical trial. JAMA network open. 2022;5(2):e2149008.
38. Pedrett R, Mascagni P, Beldi G, Padoy N, Lavanchy JL. Technical skill assessment in minimally invasive surgery using artificial intelligence: a systematic review. Surgical endoscopy. 2023;37(10):7412-24.
39. Kenig N, Monton Echeverria J, Muntaner Vives A. Artificial intelligence in surgery: a systematic review of use and validation. Journal of clinical medicine. 2024;13(23):7108.
40. Varghese C, Harrison EM, O'Grady G, Topol EJ. Artificial intelligence in surgery. Nature medicine. 2024;30(5):1257-68.
41. Xiao J, Wolter NE, Davies JC, Propst EJ, Crowson MG, Siu JM. Artificial Intelligence in Surgical Training and Applications to Otolaryngology: A Scoping Review. The Laryngoscope. 2025;135(10):3520-34.
42. Kewalramani D, Roman DS, Lagos SA, Rammsy F, Villagran I, Escalona G, et al. Quality matters: Artificial intelligence-based assessment of feedback quality predicts technical skill improvement. Surgery. 2025;187:109625.
43. Zhang Y, Weng Y, Lund J. Applications of explainable artificial intelligence in diagnosis and surgery. Diagnostics. 2022;12(2):237.
44. Moglia A, Georgiou K, Georgiou E, Satava RM, Cuschieri A. A systematic review on artificial intelligence in robot-assisted surgery. International Journal of Surgery. 2021;95:106151.
45. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022;377:e070904.
46. Funke I, Mees ST, Weitz J, Speidel S. Video-based surgical skill assessment using 3D convolutional neural networks. International Journal of Computer Assisted Radiology and Surgery. 2019;14(7):1217-25.
47. Civaner MM, Uncu Y, Bulut F, Chalil EG, Tatli A. Artificial intelligence in medical education: a cross-sectional needs assessment. BMC Medical Education. 2022;22(1):772.
48. Mascagni P, Vardazaryan A, Alapatt D, Urade T, Emre T, Fiorillo C, et al. Artificial Intelligence for Surgical Safety: Automatic Assessment of the Critical View of Safety in Laparoscopic Cholecystectomy Using Deep Learning. Annals of Surgery. 2022;275(5).
49. Carstens M, Vasisht S, Zhang Z, Barbur I, Reinke A, Maier-Hein L, et al. Artificial intelligence for surgical scene understanding: a systematic review and reporting quality meta-analysis. npj Digital Medicine. 2025;9(1):59.
50. Loftus TJ, Altieri MS, Balch JA, Abbott KL, Choi J, Marwaha JS, et al. Artificial Intelligence–enabled Decision Support in Surgery: State-of-the-art and Future Directions. Annals of Surgery. 2023;278(1).
51. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial Intelligence in Surgery: Promises and Perils. Annals of Surgery. 2018;268(1).
52. Leon S, Lee S, Perez JE, Hashimoto DA. Artificial intelligence and the education of future surgeons. The American Journal of Surgery. 2025;246:116257.

53. Winkler-Schwartz A, Yilmaz R, Mirchi N, Bissonnette V, Ledwos N, Siyar S, et al. Machine Learning Identification of Surgical and Operative Factors Associated With Surgical Expertise in Virtual Reality Simulation. JAMA Network Open. 2019;2(8):e198363-e.
54. Soleymani A, Li X, Tavakoli M. Chapter 7 - Artificial intelligence in robot-assisted surgery: Applications to surgical skills assessment and transfer. In: Boubaker O, editor. Medical and Healthcare Robotics: Academic Press; 2023. p. 183-200.
55. Park JJ, Tiefenbach J, Demetriades AK. The role of artificial intelligence in surgical simulation. Front Med Technol. 2022;4:1076755.
56. Soleymani A, Li X, Tavakoli M. Artificial intelligence in robot-assisted surgery: Applications to surgical skills assessment and transfer. Medical and Healthcare Robotics. 2023:183-200.
57. Loftus TJ, Altieri MS, Balch JA, Abbott KL, Choi J, Marwaha JS, et al. Artificial intelligence–enabled decision support in surgery: state-of-the-art and future directions. Annals of Surgery. 2023;278(1):51-8.
58. Khalifa A, Tahhan O, Albazooni M, Saeed M, Hamdi R, Stanners M, et al. Automated and Artificial Intelligence (AI)-Derived Performance Assessment in Surgical Simulation: A Systematic Review. Cureus. 2025;17(12).
59. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, et al. Artificial intelligence and surgical decision-making. JAMA surgery. 2020;155(2):148-58.
60. Wang Z, Majewicz Fey A. Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. International Journal of Computer Assisted Radiology and Surgery. 2018;13(12):1959-70.