

REVIEW

Open Access



Artificial intelligence and machine learning approaches for patient safety in complex surgery: a review

Mohamed Mustaf Ahmed^{1*} , Zhinya Kawa Othman² , Uthman Okikiola Adebayo³ , Omar Kasimieh⁴ ,
Olalekan John Okesanya^{3,5} , Shuaibu Saidu Musa⁶ , Francesco Branda⁷ , Victor C. Cañez Jr.⁸ , Edgar G. Cue⁹
and Don Eliseo Lucero Prisno III^{10,11,12}

Abstract

Artificial intelligence (AI) and machine learning (ML) are increasingly being used in surgical care; however, their real-world impact on patient safety is not well established. This narrative review searched PubMed, Scopus, and Google Scholar for English-language studies published from January 1, 2015, to April 30, 2025, that evaluated AI and ML applications in complex surgery and reported quantitative patient safety outcomes. Eligible included studies were published between 2016 and 2025. In total, 21 studies were synthesized across the preoperative, intraoperative, and postoperative phases of the study. Preoperatively, ML models consistently outperformed traditional risk scores in identifying high-risk patients and anticipating technical difficulties. Intraoperatively, AI-enabled decision support reduced hypotension exposure in a randomized trial, and computer vision systems supported the safety-critical step verification and instrument tracking. Postoperatively, multimodal approaches combining electronic records, imaging, and smartphone wound photographs predicted complications, such as surgical site infection, and facilitated discharge planning. Emerging evidence from ambulatory surgery, imaging-guided triage, and specialty domains, alongside qualitative studies on workforce readiness, highlights implementation opportunities and human factor requirements. Most evidence is retrospective, single-center, or prototype stage with limited external validation and uncertain generalizability across settings, including low- and middle-income countries. Priorities include multicenter prospective trials, standardized outcomes and reporting, continuous bias and model drift monitoring, robust data infrastructure, and equity-focused implementation to translate algorithmic performance into fewer complications, deaths, and costs.

Clinical trial registration

Not applicable.

Keywords Artificial intelligence, Machine learning, Surgery, Patient safety, Complex surgery

*Correspondence:
Mohamed Mustaf Ahmed
momustafahmed@simad.edu.so

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Globally, an estimated 310 million major surgical procedures are performed every year [1]; however, up to 25% of inpatients experience at least one postoperative complication, and the crude perioperative mortality ranges from 0.5% to 5% [2]. These hazards translate into approximately 4.2 million deaths within 30 days of surgery [3] and help drive the >3 million annual deaths attributed to unsafe care, with one in ten patients harmed overall [4]. Historic safety initiatives prove that risk is modifiable: the World Health Organization (WHO) Surgical Safety Checklist reduced inpatient complications from 11% to 7% and mortality from 1.5% to 0.8% across eight countries [5, 6]. Artificial intelligence (AI) and machine learning (ML) are increasingly being integrated into complex surgical care. Recent work frames this as a shift from retrospective audit to proactive, machine-learning-enabled anticipation and prevention of harm across the perioperative continuum [7].

Preoperatively, an extreme-gradient-boosting model trained on 66 846 elective cases achieved an AUROC of 0.95 for predicting in-hospital mortality while providing patient-specific explanations [8], and systematic reviews showed similar gains across bariatric and cardiac cohorts [9, 10]. This trajectory extends beyond inpatient pathways: in ambulatory surgery, a recent review catalogs current AI applications and practical challenges around workflow integration, governance, and evaluation [11]. Intraoperatively, computer-vision platforms quantify surgeon performance with pixel-level fidelity; the latest dual-model framework reached an intersection-over-union of 0.93 and Dice 0.87 for vessel segmentation while distinguishing expert from novice microsurgeons [12], and automated video analytics consistently correlate technical metrics with adverse event rates [13].

Human-factor considerations remain central, and qualitative research with robotic surgical nurses describes evolving roles, trust, and training needs as AI tools enter the operating room [14]. Deep neural networks ingesting high-resolution physiological streams during 56 242 operations lifted discrimination for sepsis (AUROC 0.80 vs. 0.78) and venous thrombo-embolism (0.74 vs. 0.71) over random forests [15]. Postoperatively, a multimodal neural network predicted surgical site infection within 48 h with clinician-level accuracy (AUROC 0.777 vs. 0.762) while cutting staff review time by 80% (51.5 h to 9.1 h) [16], and large-scale early warning systems now report AUROC >0.90 for major adverse events six hours before onset [17]. Specialty-focused overviews (e.g., benign prostate surgery) similarly report emerging utility alongside persistent gaps in validation and standardization that must be closed to translate performance into outcomes [18].

There is a paucity of literature on whether these predominantly retrospective, high-income country models generalize to low- and middle-income settings, remain robust to data drift, preserve equity across age, sex, and comorbidity strata, or translate AUROC gains into fewer deaths, complications, and costs in real-world workflows [19]. Bridging these evidence gaps is essential for aligning AI innovation with the Global Patient Safety Action Plan 2021–2030, enabling regulators to assign appropriate risk classes and guiding hospitals to strategically invest in scarce digital health budgets to maximize safety. This narrative review aims to critically synthesize current evidence on AI-enabled interventions for patient safety across the pre-, intra-, and postoperative continuum in complex surgery, identify persistent methodological and policy deficits that hinder real-world impact, and articulate a translational research agenda capable of turning algorithmic promises into measurable, equitable improvements in surgical outcomes.

Methods

We performed a narrative literature review to identify peer-reviewed studies that evaluated artificial intelligence (AI) technologies for patient safety outcomes before, during, and after complex surgical procedures. PubMed, Scopus, and Google Scholar were searched for articles published between January 1, 2015, and April 30, 2025. Search strings combined AI terms (“artificial intelligence,” “machine learning,” “deep learning,” “computer vision,” “natural language processing,” “robotic surgery,” “decision support”) with patient safety outcome terms (“complication,” “adverse event,” “morbidity,” “mortality,” “error,” “infection,” “readmission”) and stage descriptors (“preoperative,” “intraoperative,” “postoperative”) using Boolean operators AND and OR; Medical Subject Headings (MeSH) were applied, where available. Eligible sources were original investigations (randomized trials, cohort, case-control, cross-sectional, and diagnostic test accuracy studies), narrative reviews, and systematic reviews that (i) examined any AI application in a surgical setting, (ii) reported quantitative patient safety outcomes, and (iii) were published in English. Grey literature and non-peer-reviewed abstracts were excluded. Two reviewers (M.M.A. and O. K.) screened the titles, abstracts, and full texts; disagreements were resolved by consensus, and the reference lists of the included papers were hand-searched for additional titles. Data were extracted into a standardized matrix capturing the author, year, study design, surgical stage, specialty, key performance metrics, reported clinical impact, and stated challenges or limitations, and presented in three data extraction tables (preoperative, intraoperative, and postoperative). The findings were narratively synthesized by surgical stage

and AI application to highlight the observed effectiveness and barriers to implementation.

Results

Preoperative AI applications: risk stratification and surgical planning

Twenty-one eligible studies were included. The prespecified search window was from January 1, 2015, to April 30, 2025. No eligible studies were published in 2015. The included studies were published between 2016 and 2025, with majority published after 2020. Most published research is retrospective or cross-sectional, whereas only a small number of studies are randomized or prospective [20–22]. Preoperative studies have mainly focused on risk stratification and complexity forecasting, especially in emergency general surgery, colorectal, and hepatobiliary procedures [23–27]. During surgery, the main aim was hypotension prevention and safety confirmation, as well as instrument tracking, real-time histopathology, and robotic autonomy [21, 28–31]. Postoperative performance is mainly focused on discharge planning, remote wound surveillance, and complication prediction, such as respiratory failure, surgical site infection, and cerebral ischemia [16, 22, 32–34]. These points make it clear that governance frameworks and multicenter trials are necessary for the safe, effective, and widespread application of AI technology.

AI tools applied before incision reliably outperform traditional scores in identifying high-risk patients and anticipating technical difficulties (Table 1). Hashimoto

et al. framed the field, highlighting privacy, bias and workflow challenges while underscoring the surgeon's role [35]. Bertsimas et al. then introduced the gradient-boosted Predictive OpTimal Trees in Emergency Surgery Risk (POTTER) calculator: across 51,457 emergency operations, the model achieved a c-statistic of 0.916 for 30-day mortality, which was well above American Society of Anesthesiologists (ASA) class and American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) [23]. It also generated patient-specific complication probabilities via an intuitive dashboard [23]. External validation in 4,171 emergency laparotomy and general surgery cases confirmed excellent calibration and an area under the receiver operating characteristic curve (AUROC) of 0.93 [20].

Large multi-institution datasets have enabled the development of automated early warning systems. Mahajan et al. deployed a neural network at 20 US hospitals that flagged 13% of 1.25 million surgical patients [24]. By doing so, 52% of ensuing deaths were captured, allowing targeted prophylactic measures and multidisciplinary involvement [24]. Xue et al. fused 87 pre-operative and 94 intra-operative variables from 42,202 mixed procedures and integrated models incorporating both phases achieved AUROCs of 0.83–0.90 for major complications which allowed for 4–7% points better than pre-operative data alone [25]. Procedure-specific planning models are emerging. Yu et al. trained an interpretable XGBoost algorithm on pelvic magnetic resonance imaging (MRI) measurements and routine lab tests to predict technically

Table 1 AI impact on patient safety – Preoperative stage (risk stratification & planning)

Study (Author, year)	Reference	Design	Specialty	Key findings / Outcomes	Challenges / notes
Hashimoto et al. 2018	[35]	Narrative review	General	Outlined AI subfields; emphasized surgeon role in safe AI use	Data privacy, algorithm bias, and integration into clinical workflow identified as major challenges.
Bertsimas et al. 2018	[23]	Retrospective cohort	General (Emergency)	Provided interactive, nonlinear risk assessment for mortality and complications.	Prospective validation in diverse settings needed
El Hechi et al. 2021	[20]	Diagnostic test accuracy study	General (Emergency)	Machine learning model displayed high accuracy in mortality	Emphasized need for integration into clinical workflow (e.g. preop counseling) and caution in different hospital settings due to data variations.
Mahajan et al. 2023	[24]	Diagnostic test accuracy study	All (Mixed surgeries)	Automated preop risk model accurately identified patients at high risk of 30-day mortality complications.	High performance in large dataset; requires ongoing monitoring for bias.
Xue et al. 2021	[25]	Retrospective cohort study	All (Mixed surgeries)	Combined pre/intraoperative data to predict post operative complications	Large single-center study; performance may vary in external settings.
Kapoor et al. 2025	[38]	Diagnostic test accuracy study	General Surgery (HPB)	ML model predicted surgical difficulty with approximately 90% accuracy	Requires broader validation
Yu et al. 2024	[26]	Analytical cross-sectional study	Colorectal (Oncology)	Developed a model to predict complexity of rectal surgery (LaTME).	Pending external validation of model
Choudhury & Asan, 2020	[27]	Systematic Review	Multiple specialties	AI-based decision support systems enhancing patient risk stratification preoperatively	Noted heterogeneity in reporting and lack of standard benchmarks.

Note: AI = Artificial Intelligence; ML = Machine Learning; HPB = Hepatopancreatobiliary; LaTME = Laparoscopic Total Mesorectal Excision; preop = preoperative; intraoperative = during the operation; postoperative = after the operation

difficult laparoscopic total mesolectal excision, yielding an AUROC of 0.86 with 84% accuracy and 92% specificity in the test set [26]. While external validation is pending, the model illustrates how AI can alert surgeons to challenging anatomy and prompt referral or a modified operative strategy. A systematic review of 43 AI-based preoperative decision support tools across eight specialties concluded that most improved discrimination versus legacy tools but suffered from heterogeneous reporting and scarce benchmark datasets, limiting cross-study comparison [27].

Intraoperative AI applications: real-time decision support, computer vision and autonomy

Within the operating room, AI interventions mitigated hemodynamic instability and flagged technical lapses (Table 2). In the HYPE randomized clinical trial ($n = 60$ elective noncardiac surgeries), an ML-derived early warning system reduced the depth and duration of mean arterial pressure < 65 mmHg fourfold compared with standard care (median time-weighted MAP 0.10 vs. 0.44 mmHg · min; $P = 0.001$) [21]. Computer vision (CV) platforms can address two recurring safety hazards. Deol et al. achieved 99.4% precision in real-time detection and counting of 11 laparoscopic instruments which operates as an automated safeguard against retained items [29]. Mascagni et al. trained a deep network to verify attainment of the critical view of safety (CVS) during

laparoscopic cholecystectomy; balanced accuracy was 72%, enabling an on-screen prompt when dissection remained unsafe [36].

Rapid intraoperative pathology was enhanced by Hollon et al. whose convolutional neural network classified brain tumor specimens from stimulated Raman histology within 2.5 min and with 94.6% accuracy, matching frozen section diagnostics [30]. At the frontier of autonomy, the Smart-Tissue Autonomous Robot executed ex vivo porcine bowel anastomoses with suture precision and leak pressures comparable to those of expert surgeons [31]. For higher-level decision support, Shin et al. applied unsupervised clustering to 570 adult spinal deformity cases, uncovering four phenotypes with distinct complication profiles and opening the door to real-time, data-driven intraoperative risk stratification [28]. Nevertheless, a systematic review of AI in robot-assisted surgery found scant prospective evidence that current systems translate to improved patient outcomes, underscoring the need for multicenter trials and human–AI team workflows [37].

Postoperative AI applications: complication prediction, remote monitoring and discharge optimization

After surgery, the AI models accurately predicted adverse events and streamlined recovery pathways (Table 3). Safavi et al. created a ward-based neural network that forecasts next-day discharges (AUROC 0.84), where

Table 2 AI impact on patient safety – Intraoperative stage (decision support, automation & error reduction)

Study (Author, year)	Reference	Design	Specialty	Key findings / outcomes	Challenges / notes
Wijnberge et al. 2020	[21]	Randomized control trial	Anesthesiology	AI early-warning system predicted hypotension before it occurred	This was a single-center preliminary trial. Larger multicenter trials needed to confirm outcome benefits.
Deol et al. 2024	[29]	Analytical Cross-sectional study	Urology (robotic)	Developed a real-time computer vision model to count surgical tools, aiming to prevent retained items.	Study not yet in human surgeries.
Mascagni et al. 2022	[36]	Analytical Cross-sectional study	General Surgery (HPB)	AI model automatically assessed CVS criteria achievement during gallbladder surgery.	Moderate accuracy – unable to catch all nuances
Hollon et al. 2020	[30]	Diagnostic test accuracy study	Neurosurgery	Used a deep neural network with stimulated Raman histology to diagnose brain tumor specimens intraoperatively	Implemented in specialized centers with SRH equipment; generalization to all pathology types needs work
Moglia et al. 2021	[37]	Systematic review	Multiple (Robotic)	Surveyed AI in robotic surgery: autonomous suturing, gesture recognition, and skill assessment algorithms.	Lack clinical validation
Shin et al. 2019	[28]	Case Control Study	Orthopedic (Spine)	Applied models to adult spinal deformity cases to group patients by key features and predict outcomes	Provided a new risk classification beyond conventional surgeon intuition
Shademan et al. 2016	[31]	Diagnostic test accuracy study	General Surgery	STAR successfully performed intestinal anastomoses in ex vivo porcine models with minimal human intervention.	Conducted in laboratory conditions on animal tissue, not clinical use.

Note: AI = Artificial Intelligence; HPB = Hepatopancreatobiliary; CVS = Critical View of Safety; SRH = Stimulated Raman Histology; STAR = Smart Tissue Autonomous Robot

Table 3 AI impact on patient safety – Postoperative stage (monitoring, complications prediction & response)

Study (Author, Year)	Reference	Design	Specialty	Key Findings / Outcomes	Challenges / Notes
Safavi et al. 2019	[22]	Cohort study	General Surgery	A neural network predicted which post-surgical inpatients would be discharged within 24 h	Model was single center; performance may vary elsewhere
Ravenel et al. 2023	[32]	Retrospective study	General Surgery	Applied machine learning to colorectal cancer patients with bilobar liver metastases to identify prognostic factors and develop a nomogram predicting overall survival.	Many models lack external validation and are retrospective. In addition, Heterogeneity in endpoints and sample sizes makes generalizability difficult.
McLean et al. 2025	[16]	Diagnostic test accuracy study	Colorectal (GI surgery)	Combined patient-reported symptoms and wound to predict surgical site infection within 48 h.	Developed in a controlled study; real-world adherence to remote monitoring can affect performance.
Hu et al. 2022	[33]	Diagnostic test accuracy study	Neurosurgery	Applied machine learning to early clinical and imaging data after subarachnoid hemorrhage to predict delayed cerebral ischemia with greater accuracy than traditional approaches.	Highly specific context (neurocritical care); requires continuous multimodal data inputs
Hadaya et al. 2022	[34]	Analytical cross-sectional study	Emergency surgery	Developed a predictive model for post-op respiratory failure in emergency general surgery patients	Focused on only one complication (ARF) using registry data.
Shickel et al. 2023	[15]	Analytical cross-sectional study	All (Mixed inpatient surgery)	Applied a deep neural network on surgeries to predict nine common 30-day complications	Complex model requiring extensive EHR integration

Note: GI = Gastrointestinal; post-op = postoperative; ARF = Acute Respiratory Failure; EHR = Electronic Health Record

prospective use in 605 patients freed 128 bed-days without increasing readmissions [22]. Hadaya et al. mined more than 1 million emergency general surgery cases to build an XGBoost model for predicting acute respiratory failure, with an AUROC of 0.90 [34]. There were patients flagged as high risk who had greater than 10-fold higher mortality and markedly longer stays [33].

Remote patient-generated data streams are now actionable. McLean et al. combined smartphone wound photographs with symptom surveys that employed a multimodal deep learning model detected surgical site infection 48 h before clinical diagnosis (AUROC 0.76) while cutting manual surveillance workload by 80% [16]. In neurologic critical care, Hu et al. integrated early computed tomography (CT) and physiological data to forecast delayed cerebral ischemia after subarachnoid hemorrhage (AUROC 0.88), outperforming conventional scores [33]. Oncology applications are also advancing; a scoping review by Ravenel et al. indexed machine learning nomograms that stratify digestive surgery patients' long-term outcomes and guide adjuvant therapy decisions [32]. AI systems have improved preoperative risk stratification, curtailed intraoperative hypotension and technical errors, and accurately predicted diverse postoperative complications (Fig. 1). The strongest evidence is a randomized hemodynamic trial that demonstrated a fourfold reduction in hypotension, whereas large multisite early warning models have already reshaped perioperative workflows. However, most tools remain single-center, retrospective, or in the prototype stage.

Rigorous multicenter trials and standardized reporting are essential before routine clinical use.

Discussion

Artificial intelligence tools are now being designed for every phase of complex surgery; however, most published evidence remains in the early stages and is from single centers. A recurring pattern in the literature is that algorithms achieve impressive performance metrics in silico or in narrowly defined settings but only sporadically demonstrate measurable gains in real-world patient safety. Recent studies have continuously demonstrated that ML models, especially those employing gradient boosting and random forests, are superior to conventional risk scores, such as ASA and NSQIP, in forecasting surgical mortality and complications. For instance, a large experiment using random forest classifiers on more than 49,000 surgical patients produced an AUC of 0.921 for predicting in-hospital mortality, which was higher than the AUCs of both logistic regression and the ASA Physical Status Classification. Integrating preoperative and intraoperative data into deep neural networks further enhanced predictive accuracy, with AUCs reaching 0.924, and demonstrated that patient risk changes over time and can be tracked longitudinally. At a systems level, big-data surgical registries and trauma databases are being coupled with ML to operationalize continuous safety surveillance and decision support at scale,

Characteristic	Preoperative	Intraoperative	Postoperative
 Focus	Risk stratification, surgical planning	Hypotension prevention, safety confirmation	Discharge planning, complication prediction
 AI Aim	Identify high-risk patients, anticipate difficulties	Instrument tracking, real-time histopathology	Remote wound surveillance, predict complications
 Examples	POTTER calculator, early warning systems	Robotic autonomy	Predicting Respiratory failure

Fig. 1 AI applications in surgery. POTTER=Predictive OpTimal Trees in Emergency Surgery Risk

reinforcing the need for standardized data pipelines and governance [39].

These ML-based methods exclusively use structured electronic medical record data, allowing for completely automated risk forecasting without the need for clinician-derived scores and providing more personalized and dynamic risk assessments than traditional methods. However, before widespread implementation can take place, reliable external validation and smooth integration into clinical processes are still necessary [40]. With randomized trials showing a fourfold reduction in hypotension exposure during elective noncardiac surgery [25, 41], AI-derived early warning systems, such as the one evaluated in this review, have demonstrated a significant reduction in both the depth and duration of intraoperative hypotension compared with standard care. Thermoregulatory safety is another intraoperative target: explainable ML has shown promise for predicting hypothermia risk during laparoscopic surgery, broadening the scope of AI-assisted prevention beyond hemodynamics [42].

Similar benefits have been observed in cardiac surgery, where machine learning-based hypotension prediction tools have reduced the severity and duration of hypotension by more than 60% without raising blood pressure or requiring more medication [43]. Additional randomized trials have confirmed that these systems not only lower the occurrence of hypotensive episodes and the amount of time spent in hypotension, but also lower biomarkers of organ injury and oxidative stress, indicating the potential for improved patient outcomes [44, 45]. A Systematic review and meta-analysis further supported that

AI-driven predictive analytics, especially the Hypotension Prediction Index (HPI), consistently outperformed standard monitoring by reducing intraoperative hypotension and demonstrating strong predictive accuracy [46]. Nevertheless, most data originate from single-center or controlled environments, and the long-term effects on patient outcomes and cost-effectiveness remain unclear, underscoring the need for multicenter trials and more extensive validation before widespread clinical use [47].

Multimodal deep learning models that combine wound images with patient-reported symptoms have demonstrated significant potential for the early detection of SSI and for reducing the workload of manual monitoring. Compared to clinician triage, the model, like the one created in this review, identified SSIs up to 48 h prior to clinical diagnosis and had similar diagnostic accuracy while lowering the staff time needed for monitoring by 80% [16]. Broader research, which highlights the necessity of external validation and meticulous integration into clinical workflows, supports these findings by demonstrating that automated, image-based, and multimodal machine learning methods can match or even outperform clinicians in detecting SSI [48]. Additionally, animal model studies have shown that multimodal sensors and machine learning systems can predict SSIs up to 48 h in advance, further corroborating the possibility of early, automated detection [49].

Pathology-focused AI also contributes to safety by standardizing detection and grading, as shown in validated work on prostatectomy specimens [50]. Across surgical subspecialties, focused overviews in spine surgery and benign prostate surgery echo our conclusions

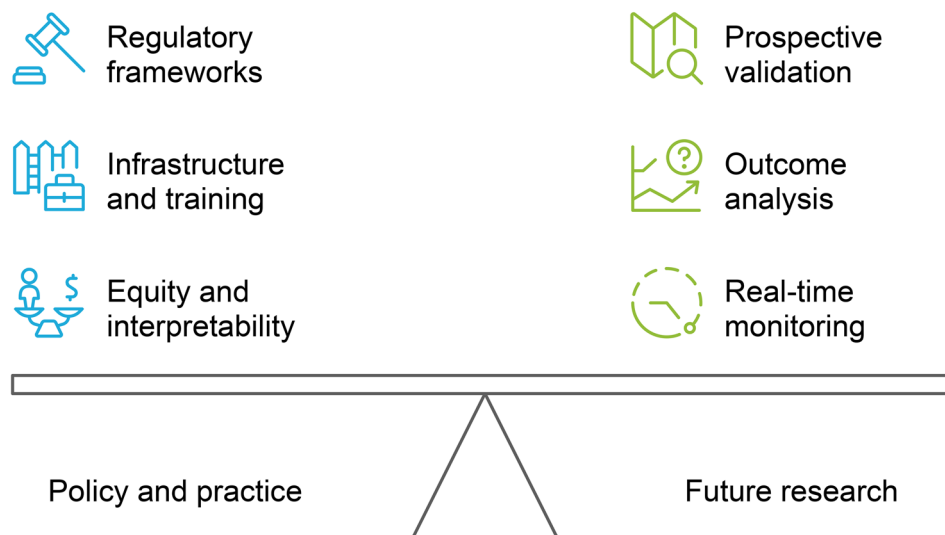


Fig. 2 Balancing policy, practice, and research in AI integration

that growing predictive performance is tempered by heterogeneous reporting, limited external validation, and a need for prospective, outcome-oriented trials before routine use [51]. Preoperatively, non-invasive ML models that integrate ultrasonography are being leveraged to guide selective lymphadenectomy in thoracic esophageal cancer, illustrating how AI can reduce overtreatment and perioperative risk when thoughtfully integrated into decision pathways [52]. Nonetheless, obstacles persist, such as ensuring the capture of high-quality data, keeping patients engaged, and incorporating these technologies into current treatment pathways to optimize their efficacy and scalability for regular practice [48].

Strengths and limitations

One of the major strengths of this review is its thorough coverage of the pre-, intra-, and postoperative continuum, which integrates data from various surgical disciplines and study methodologies. The review includes both technical performance measures and early indicators of real-world impact by incorporating excellent prospective and randomized trials. For example, the incorporation of massive multi-institutional datasets, such as Mahajan et al.'s neural network analysis of more than a million patients, improves the generalizability of the results and facilitates comparisons across various clinical contexts. However, the nature of the available data restricted the review. Most of the included studies were retrospective or single-center and were frequently conducted in high-income environments, which raises questions about external validity and relevance to low- and middle-income nations. Some promising AI tools have been validated only in regulated laboratory settings, whereas others have not been validated externally. Direct comparison between studies is also difficult because of

heterogeneity in outcome definitions, input variables, and performance reporting, as emphasized in systematic reviews.

Implications for policy, practice and future research

AI technologies have the potential to improve patient safety throughout the surgical process; however, turning these gains into regular treatments necessitates specific policies, practical approaches, and research methodologies (Fig. 2).

From a policy standpoint, regulatory frameworks should be changed to tackle the specific challenges of AI, such as transparency, bias reduction, and adaptive approval procedures that enable iterative model improvement while protecting patient safety. To ensure the smooth integration of AI tools into perioperative decision-making without adding to the cognitive load of healthcare professionals, hospitals and healthcare systems should invest in digital infrastructure, interdisciplinary training, and workflow redesigns. Implementation should be guided by an equity-focused design with proactive auditing of model performance across various care settings and patient subgroups. To encourage clinician confidence and continuous adoption, AI outputs must be interpretable, timely, and consistent with the current clinical procedures.

To determine generalizability and cost-effectiveness, future studies should prioritize the prospective multi-center validation of promising tools, especially in diverse geographic and resource contexts. In addition to predictive accuracy, research should also analyze how AI-driven insights impact actual patient outcomes and investigate ways to integrate real-time performance monitoring into deployed systems. By concentrating innovation on these priorities, AI may develop from experimental prototypes

into trustworthy copilots that demonstrably improve surgical safety across the globe.

Conclusion

AI and ML are now involved in every phase of complex surgery, from preoperative risk stratification to intraoperative prevention and postoperative surveillance. Signals of benefit include stronger preoperative discrimination than legacy scores, reduced intraoperative hypotension in randomized evaluations, and earlier detection of postoperative complications with workload reduction. Translation to routine safety gains remains limited by retrospective designs, single-center development, heterogeneous reporting, and scarce external validation. Human factors readiness and governance are equally pivotal, as shown by qualitative work in robotic surgery and the need for explainability, training pathways, and workflow-compatible interfaces. System-level enablers, such as data registries, standardized pipelines, and ongoing performance auditing, are necessary to sustain safe deployment and detect bias and drift. Future work should emphasize multicenter prospective trials, outcome-oriented implementation studies across diverse settings, and cost-effectiveness analyses. With these elements in place, AI and ML can evolve from promising prototypes to equitable and auditable tools that can measurably reduce harm and improve surgical outcomes.

Acknowledgements

The authors acknowledge the use of Paperpal AI (<https://paperpal.com/>) for its "Language Edit" and "Make Academic" features to improve clarity and readability. This assistance was limited to linguistic refinement; all analyses and interpretations are solely those of the authors.

Author contributions

M.M.A and O.K conceptualized and designed the study; Z.K.O, O.J.O and S.S.M performed the literature review and data curation; M.M.A and U.O.A drafted the manuscript; F.B, V.C.G and E.C.G revised the manuscript; D.E.P.III supervised the study; and all authors have read and approved the final manuscript.

Funding

The authors received no specific funding for this work.

Data availability

Not applicable; no new datasets were generated or analyzed.

Declarations

Ethical approval

Not required for this study.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Faculty of Medicine and Health Sciences, SIMAD University, Mogadishu, Somalia

²Department of Pharmacy, Kurdistan Technical Institute, Sulaymaniyah, Kurdistan Region, Iraq

³Department of Medical Laboratory Science, Neuropsychiatric Hospital, Aro, Abeokuta, Nigeria

⁴University of the East Ramon Magsaysay Memorial Medical Center, Quezon City, Philippines

⁵Department of Public Health and Maritime Transport, University of Thessaly, Volos, Greece

⁶School of Global Health, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

⁷Unit of Medical Statistics and Molecular Epidemiology, Campus Bio-Medico University of Rome, Rome, Italy

⁸Office of the University President, Biliran Province State University, Naval, Leyte, Philippines

⁹Office of the University President, Mountain Province State University, Bontoc, Mountain Province, Philippines

¹⁰Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London, UK

¹¹Office for Research, Extension and Innovations, Bukidnon State University, Malaybalay City, Bukidnon, Philippines

¹²Research Office, Palompon Institute of Technology, Palompon, Leyte, Philippines

Received: 18 August 2025 / Accepted: 30 October 2025

Published online: 25 November 2025

References

1. Dobson GP. Trauma of major surgery: A global problem that is not going away. *Int J Surg*. 2020;81:47–54. <https://doi.org/10.1016/j.ijsu.2020.07.017>.
2. Safe surgery. World Health Organization. 2025. <https://www.who.int/teams/integrated-health-services/patient-safety/research/safe-surgery>. (accessed October 29, 2025).
3. Tekalign T, Balta H, Kelbiso L. Magnitude of post-operative mortality and associated factors among patients who underwent surgery in Wolaita Sodo teaching and referral hospital, SNNPR region, Ethiopia. *Afr Health Sci*. 2021;21:1842–8. <https://doi.org/10.4314/ahs.v21i4.42>.
4. Patient safety. World Health Organization. 2023. <https://www.who.int/news-room/fact-sheets/detail/patient-safety>. (accessed October 29, 2025).
5. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat A-HS, Dellinger EP, Herbosa T, Joseph S, Kibatala PL, Lapitan MCM, Merry AF, Moorthy K, Reznick RK, Taylor B, Gawande AA. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med*. 2009;360:491–9. <https://doi.org/10.1056/NEJMsa0810119>.
6. Treadwell JR, Lucas S, Tsou AY. Surgical checklists: a systematic review of impacts and implementation. *BMJ Qual Saf*. 2014;23:299–318. <https://doi.org/10.1136/bmjqs-2012-001797>.
7. Stahel PF, Holland K, Nanz R. Machine learning approaches for improvement of patient safety in surgery. *Patient Saf Surg*. 2024;18:37. <https://doi.org/10.1186/s13037-024-00422-y>.
8. Graebner M, Jungwirth B, Frank E, Schaller SJ, Kochs E, Ulm K, Blobner M, Ulm B, Podtschaske AH, Kagerbauer SM. Enabling personalized perioperative risk prediction by using a machine-learning model based on preoperative data. *Sci Rep*. 2023;13:7128. <https://doi.org/10.1038/s41598-023-33981-8>.
9. Hassan Mukhtar MA, Babiker Ahmed AU, Siddig Mohammed MA, Ibrahim Omer NO, Altom DS, Elnour MAA. The role of artificial intelligence in the prediction of bariatric surgery complications: a systematic review. *Cureus*. 2025. <https://doi.org/10.7759/cureus.82461>.
10. Leivaditis V, Beltsios E, Papatrifiantayllou A, Grapatsas K, Mulita F, Kontodimopoulos N, Baikoussis NG, Tchabashvili L, Tasios K, Maroulis I, Dahm M, Koletsis E. Artificial intelligence in cardiac surgery: transforming outcomes and shaping the future. *Clin Pract*. 2025;15:17. <https://doi.org/10.3390/clinpract1510017>.
11. Liu L, Zhang P, Jia Y, Hou L, Peng D, Li Z, Liang P. Artificial intelligence in ambulatory surgery: current Applications, Challenges, and future directions. *J Med Syst*. 2025;49:146. <https://doi.org/10.1007/s10916-025-02286-w>.
12. Sugiyama T, Tang M, Sugimori H, Sakamoto M, Fujimura M. Artificial intelligence-integrated video analysis of vessel area changes and instrument

- motion for microsurgical skill assessment. *Sci Rep.* 2025;15:27898. <https://doi.org/10.1038/s41598-025-13522-1>.
13. Hamza H, Shabir D, Aboumarzouk O, Al-Ansari A, Shaban K, Navkar NV. Automated skills assessment in open surgery: A scoping review. *Eng Appl Artif Intell.* 2025;153:110893. <https://doi.org/10.1016/j.engappai.2025.110893>.
 14. Alptekin HM, Kara Ö, Seyhan Ak E, Şabihat F. Robotic surgical nurses' perceptions of artificial intelligence: a qualitative study. *J Robot Surg.* 2025;19:717. <https://doi.org/10.1007/s11701-025-02924-y>.
 15. Shickel B, Loftus TJ, Ruppert M, Upchurch GR, Ozrazgat-Baslanti T, Rashidi P, Bihorac A. Dynamic predictions of postoperative complications from explainable, uncertainty-aware, and multi-task deep neural networks. *Sci Rep.* 2023;13:1224. <https://doi.org/10.1038/s41598-023-27418-5>.
 16. McLean KA, Sgrò A, Brown LR, Buijs LF, Mountain KE, Shaw CA, Drake TM, Pius R, Knight SR, Fairfield CJ, Skipworth RJE, Tsaftaris SA, Wigmore SJ, Potter MA, Bouamrane M-M, Harrison EM, Baweja K, Cambridge WA, Chauhan V, Czyzykowska K, Edirisooriya M, Forsyth A, Fox B, Fretwell J, Gent C, Gherman A, Green L, Grewar J, Heelan S, Henshall D, Iiuoma C, Jayasangaran S, Johnston C, Kennedy E, Kremel D, Kung J, Kwong J, Leavy C, Liu J, Mackay S, MacNamara A, Mowitt S, Musenga E, Ng N, Ng ZH, O'Neill S, Ramage M, Reed J, Riad A, Scott C, Sehgal V, Sgrò A, Steven L, Stutchfield B, Tominey S, Wilson W, Wojtowicz M, Yang J. Multimodal machine learning to predict surgical site infection with healthcare workload impact assessment. *NPJ Digit Med.* 2025;8:121. <https://doi.org/10.1038/s41746-024-01419-8>.
 17. Sim T, Cho EY, Kim J, Lee KH, Kim KJ, Hahn S, Ha EY, Yun E, Kim I-C, Park SH, Cho C-H, Yu GI, Ahn BE, Jeong Y, Won J-Y, Cho H, Lee K-B. Prospective external validation of a deep-learning-based early-warning system for major adverse events in general wards in South Korea. *Acute Crit Care.* 2025;40:197–208. <https://doi.org/10.4266/acc.000525>.
 18. Nedbal C, Tramanzoli P, Somani BK, De Stefano V, Stramucci S, Pirola GM, Lim EJ, Yuen SKK, Galosi AB, Panthier F, Einikev D, Hermann T, Tiong HY, Gauhar V. Current utility of artificial intelligence in benign prostate surgery: a review of literature on behalf of the EAU section of endourology. *World J Urol.* 2025;43:631. <https://doi.org/10.1007/s00345-025-05524-5>.
 19. Global patient safety report 2024. World Health Organization 2024. <https://www.who.int/publications/i/item/9789240095458>. (accessed October 29, 2025).
 20. El Hechi MW, Maurer LR, Levine J, Zhuo D, El Moheb M, Velmahos GC, Dunn J, Bertsimas D, Kaafarani HMA. Validation of the artificial Intelligence-Based predictive optimal trees in emergency surgery risk (POTTER) calculator in emergency general surgery and emergency laparotomy patients. *J Am Coll Surg.* 2021;232:912–e9191. <https://doi.org/10.1016/j.jamcollsurg.2021.02.009>.
 21. Wijnberge M, Geerts BF, Hol L, Lemmers N, Mulder MP, Berge P, Schenk J, Terwindt LE, Hollmann MW, Vlaar AP, Veelo DP. Effect of a machine Learning-Derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery. *JAMA.* 2020;323:1052. <https://doi.org/10.1001/jama.2020.0592>.
 22. Safavi KC, Khaniyev T, Copenhaver M, Seelen M, Zenteno Langle AC, Zanger J, Daily B, Levi R, Dunn P. Development and validation of a machine learning model to aid discharge processes for inpatient surgical care. *JAMA Netw Open.* 2019;2:e1917221. <https://doi.org/10.1001/jamanetworkopen.2019.17221>.
 23. Bertsimas D, Dunn J, Velmahos GC, Kaafarani HMA. Surgical risk is not linear: derivation and validation of a Novel, User-friendly, and Machine-learning-based predictive optimal trees in emergency surgery risk (POTTER) calculator. *Ann Surg.* 2018;268:574–83. <https://doi.org/10.1097/SLA.0000000000002956>.
 24. Mahajan A, Esper S, Oo TH, McKibben J, Garver M, Artman J, Klahre C, Ryan J, Sadhasivam S, Holder-Murray J, Marroquin OC. Development and validation of a machine learning model to identify patients before surgery at high risk for postoperative adverse events. *JAMA Netw Open.* 2023;6:e2322285. <https://doi.org/10.1001/jamanetworkopen.2023.22285>.
 25. Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, Kannampallil T, Abraham J. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Netw Open.* 2021;4:e212240. <https://doi.org/10.1001/jamanetworkopen.2021.2240>.
 26. Yu M, Yuan Z, Li R, Shi B, Wan D, Dong X. Interpretable machine learning model to predict surgical difficulty in laparoscopic resection for rectal cancer. *Front Oncol.* 2024;14. <https://doi.org/10.3389/fonc.2024.1337219>.
 27. Choudhury A, Asan O. Role of artificial intelligence in patient safety outcomes: systematic literature review. *JMIR Med Inf.* 2020;8:e18599. <https://doi.org/10.2196/18599>.
 28. Shin J, Choi JY, Kim YW, Chang JS, Yoon SY. Quantification of risk factors for cervical ossification of the posterior longitudinal ligament in Korean populations. *Spine (Phila Pa 1976).* 2019;44:E957–64. <https://doi.org/10.1097/BRS.0000000000003027>.
 29. Deol ES, Henning G, Basourakos S, Vasdev RMS, Sharma V, Kavoussi NL, Karnes RJ, Leibovich BC, Boorjian SA, Khanna A. Artificial intelligence model for automated surgical instrument detection and counting: an experimental proof-of-concept study. *Patient Saf Surg.* 2024;18:24. <https://doi.org/10.1186/s13037-024-00406-y>.
 30. Hollon TC, Pandian B, Adapa AR, Urias E, Save AV, Khalsa SSS, Eichberg DG, D'Amico RS, Farooq ZU, Lewis S, Petridis PD, Marie T, Shah AH, Garton HJL, Maher CO, Heth JA, McKean EL, Sullivan SE, Hervey-Jumper SL, Patil PG, Thompson BG, Sagher O, McKhann GM, Komotar RJ, Ivan ME, Snuderl M, Otten ML, Johnson TD, Sisti MB, Bruce JN, Muraszko KM, Trautman J, Freudiger CW, Canoll P, Lee H, Camelo-Piragua S, Orringer DA. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat Med.* 2020;26:52–8. <https://doi.org/10.1038/s41591-019-0715-9>.
 31. Shademan A, Decker RS, Opfermann JD, Leonard S, Krieger A, Kim PCW. Supervised autonomous robotic soft tissue surgery. *Sci Transl Med.* 2016;8. <https://doi.org/10.1126/scitranslmed.aad9398>.
 32. Ravenel M, Joliat G-R, Demartines N, Uldry E, Melloul E, Labgaa I. Machine learning to predict postoperative complications after digestive surgery: a scoping review. *Br J Surg.* 2023;110:1646–9. <https://doi.org/10.1093/bjs/znad229>.
 33. Hu P, Li Y, Liu Y, Guo G, Gao X, Su Z, Wang L, Deng G, Yang S, Qi Y, Xu Y, Ye L, Sun Q, Nie X, Sun Y, Li M, Zhang H, Chen Q. Comparison of conventional logistic regression and machine learning methods for predicting delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage: a multicentric observational cohort study. *Front Aging Neurosci.* 2022;14. <https://doi.org/10.3389/fnagi.2022.857521>.
 34. Hadaya J, Verma A, Sanaiha Y, Ramezani R, Qadir N, Benharash P. Machine learning-based modeling of acute respiratory failure following emergency general surgery operations. *PLoS ONE.* 2022;17:e0267733. <https://doi.org/10.1371/journal.pone.0267733>.
 35. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg.* 2018;268:70–6. <https://doi.org/10.1097/SLA.0000000000002693>.
 36. Mascagni P, Vardazaryan A, Alapatt D, Urade T, Emre T, Fiorillo C, Pessaux P, Mutter D, Marescaux J, Costamagna G, Dallemagne B, Padoy N. Artificial intelligence for surgical safety. *Ann Surg.* 2022;275:955–61. <https://doi.org/10.1097/SLA.0000000000004351>.
 37. Moglia A, Georgiou K, Georgiou E, Satava RM, Cuschieri A. A systematic review on artificial intelligence in robot-assisted surgery. *Int J Surg.* 2021;95:106151. <https://doi.org/10.1016/j.ijsu.2021.106151>.
 38. Kapoor A, Sidhu BS, Singh J. Artificial intelligence-based detection of pre-operative surgical difficulty in laparoscopic cholecystectomy patients using gradient based model. *Int Surg J.* 2025;12:555–61. <https://doi.org/10.18203/2349-2902.isj20250814>.
 39. Pape H-C, Starr AJ, Gueorguiev B, Wanner GA. The role of big data management, data registries, and machine learning algorithms for optimizing safe definitive surgery in trauma: a review. *Patient Saf Surg.* 2024;18:22. <https://doi.org/10.1186/s13037-024-00404-0>.
 40. Hill B, Brown R, Gabel E, Lee C, Cannesson M, Loohuis LO, Johnson R, Jew B, Maoz U, Mahajan A, Sankararaman S, Hofer I, Halperin E. Preoperative predictions of in-hospital mortality using electronic medical record data 2018. <https://doi.org/10.1101/329813>.
 41. Angus DC. Randomized clinical trials of artificial intelligence. *JAMA.* 2020;323:1043. <https://doi.org/10.1001/jama.2020.1039>.
 42. Su W, Wang X, Jia H, Chang W, Jiang S, Ge H, Dong S, Yu J, Ma G, Meng Y. Explainable prediction of hypothermia risk in laparoscopic surgery: a retrospective cross-sectional study using machine learning. *BMC Surg.* 2025;25:433. <https://doi.org/10.1186/s12893-025-03071-9>.
 43. Schuurmans J, Rellum SR, Schenk J, van der Ster BJP, van der Ven WH, Geerts BF, Hollmann MW, Cherpanath TGV, Lagrand WK, Wynandts PR, Paulus F, Driessen AHG, Terwindt LE, Eberl S, Hermanns H, Veelo DP, Vlaar APJ. Effect of a machine Learning-Derived early warning tool with treatment protocol on hypotension during cardiac surgery and ICU stay: the hypotension prediction 2 (HYPE-2) randomized clinical trial. *Crit Care Med.* 2025;53:e328–40. <https://doi.org/10.1097/CCM.00000000000006518>.
 44. Murabito P, Astuto M, Sanfilippo F, Via L, La, Basile F, Cappellani A, Longhitano L, Distefano A, Volti GL, Vasile F. Performance of hypotension prediction index

- for prevention of hypotension and results on biomarkers of organ injury and oxidative stress during elective non-cardiac surgery: a randomized controlled trial 2021. <https://doi.org/10.21203/rs.3.rs-851209/v1>
45. Murabito P, Astuto M, Sanfilippo F, La Via L, Vasile F, Basile F, Cappellani A, Longhitano L, Distefano A, Li Volti G. Proactive management of intraoperative hypotension reduces biomarkers of organ injury and oxidative stress during elective Non-Cardiac surgery: A pilot randomized controlled trial. *J Clin Med*. 2022;11:392. <https://doi.org/10.3390/jcm11020392>.
46. Mohammadi I, Firouzabadi SR, Hosseinpour M, Akhlaghpasand M, Hajikarimloo B, Tavanaei R, Izadi A, Zeraatian-Nejad S, Eghbali F. Predictive ability of hypotension prediction index and machine learning methods in intraoperative hypotension: a systematic review and meta-analysis. *J Transl Med*. 2024;22:725. <https://doi.org/10.1186/s12967-024-05481-4>.
47. van der Ven WH, Veelo DP, Wijnberge M, van der Ster BJP, Vlaar APJ, Geerts BF. One of the first validations of an artificial intelligence algorithm for clinical use: the impact on intraoperative hypotension prediction and clinical decision-making. *Surgery*. 2021;169:1300–3. <https://doi.org/10.1016/j.surg.2020.09.041>.
48. Tabja Bortesi JP, Ranisau J, Di S, McGillion M, Rosella L, Johnson A, Devereaux P, Petch J. Machine learning approaches for the Image-Based identification of surgical wound infections: scoping review. *J Med Internet Res*. 2024;26:e52880. <https://doi.org/10.2196/52880>.
49. Hughes CML, Jeffers A, Sethuraman A, Klum M, Tan M, Tan V. The detection and prediction of surgical site infections using multi-modal sensors and machine learning: results in an animal model. *Front Med Technol* 2023;5. <https://doi.org/10.3389/fmedt.2023.1111859>
50. Kudo MS, Gomes de Souza VM, Estivallet CLN, de Amorim HA, Kim FJ, Leite KRM, Moraes MC. The value of artificial intelligence for detection and grading of prostate cancer in human prostatectomy specimens: a validation study. *Patient Saf Surg*. 2022;16:36. <https://doi.org/10.1186/s13037-022-00345-6>.
51. Arjmandnia F, Alimohammadi E. The value of machine learning technology and artificial intelligence to enhance patient safety in spine surgery: a review. *Patient Saf Surg*. 2024;18:11. <https://doi.org/10.1186/s13037-024-00393-0>.
52. Li Y, Gao Y, Li B, Zhang Y, Zhao Y, Deng C, Fan F, Pan Y, Huang Q, Fu F, Yu Y, Ma Y, Sun Y, Zhang Y, Xiang J, Chen H, Zhang Y. A non-invasive approach based on ultrasonography and machine learning for selective cervical lymphadenectomy in thoracic esophageal cancer. *Int J Surg*. 2025. <https://doi.org/10.1097/JS9.0000000000003803>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.