Past, Present, and Future of AI for Surgical Guidance in Minimally Invasive and Robotic Procedures: A Focused Scoping Review

Remy Miller

Abstract

This focused scoping review examines the development, current applications, and future translational opportunities of artificial intelligence (AI) in minimally invasive surgery (MIS) and robotic-assisted procedures. A literature search was carried out across PubMed, IEEE Xplore, and Google Scholar to evaluate artificial intelligence applications for intraoperative surgical guidance in minimally invasive and robotic procedures. Eligible publications were analyzed across technical architectures, clinical data modalities, evaluation metrics, and translational constraints to characterize the continuum from early manual laparoscopy to modern algorithmic assistance.

Current literature demonstrates a steady transition toward multimodal deep learning architectures, including convolutional neural networks, vision transformers, and recurrent sequence models. Modern platforms achieve high accuracy under controlled conditions in instrument tracking, procedural phase recognition, and anatomical landmark identification. In preclinical settings, supervised task-autonomous systems-such as the Smart Tissue Autonomous Robot-have demonstrated higher consistency in automated suturing metrics relative to manual techniques. However, translation into real-world operating rooms remains constrained by dataset homogeneity, annotation bottlenecks, intraoperative visual disruptions (e.g., surgical smoke and blood fouling), processing latency, and algorithmic opacity. Concurrently, evolving medicolegal and informed-consent frameworks raise unresolved questions regarding institutional accountability and surgeon responsibility when predictive systems fail. Successful clinical translation will likely depend on decentralized federated learning networks, real-time image restoration, structured institutional credentialing, and supervised shared-control systems that assist with standardized procedural steps while keeping clinicians firmly in the decision-making loop.

Introduction

Laparoscopic/Minimally Invasive Surgery (MIS) is a form of surgery that is an alternative to open/traditional surgery. Laparoscopic surgery is a surgical procedure in which a camera is inserted through small incisions in the abdomen. The laparoscope is inserted, and the other instruments are inserted through the small incisions. There are many benefits to laparoscopic surgery: decreased postoperative pain, shorter time in the hospital, faster recovery, lower infection rates due to the small incision size, and reduced tissue disruption. However, relative to open surgery in which the surgeon directly handles anatomy, laparoscopic surgery presents some challenges for surgeons, including diminished direct visualization while looking through the laparoscope, limited tactile feedback, and constrained degrees of freedom for some surgical instruments. Therefore, intraoperative guidance has become essential to counteract the inherent

visualization, orientation, and instrument-control constraints of laparoscopic surgery. Robot-assisted surgery (RAS) extends minimally invasive techniques by providing surgeons with enhanced instrument control, visualization, and dexterity while continuing to maintain the benefits of small-incision surgery. Building on the robotic platforms, Machine Learning (ML) and Deep Learning (DL) address remaining procedural hurdles through real-time multimodal intraoperative analysis. Specifically, computer vision models can segment critical anatomical structures to prevent accidental tissue injury, classify surgical phases for workflow progression, and project augmented reality overlays directly onto the video feed to help compensate for the loss of depth perception (Oh, 2026; Zhang et al., 2026). By assisting surgeons with anatomical identification, tool navigation, and real-time decision support, these intelligent frameworks deliver the potential to improve operative efficiency and clinical outcomes. The objective of this review is to investigate the past developments, current applications, and arising challenges of artificial intelligence in intraoperative guidance across minimally invasive and robotic procedures.

While AI shows great potential for minimally invasive surgery, its implementation is constrained by several central challenges. Regarding real-time technical demands and visual conditions, surgical AI needs to process live video streams with zero or near-zero latency to safely support augmented reality overlays, phase recognition, skill assessment, and closed-loop robotic guidance. If visual input is poor, patient safety is compromised (Dhingra et al., 2025). Moreover, models struggle with visual sharpness in dynamic operating environments; an estimated 40% of minimally invasive surgical time occurs under suboptimal conditions such as smoke, blood, or lens fogging, which contributes to nearly 20% of surgical complications (Dhingra et al., 2025). Unless surgical AI can robustly filter through these artifacts to accurately identify anatomical structures, its actual clinical use will remain heavily constrained. Regarding data shortage as well as generalizability, developing reliable deep learning models requires large volumes of high-quality datasets that are difficult to compile due to strict privacy matters and blockages to data sharing between institutions. Methodological issues, such as splitting training and testing data on a per-frame basis instead of per patient, lead to data leakage that compromises reliability. Consequently, these models regularly fail when exposed to different endoscopic systems, different operating room conditions, or unfamiliar tissue types. For clinical translation, these models require full regulatory approval and safety benchmarks before surgeons can trust and implement them in the operating room. High accuracy in experimental settings does not necessarily guarantee clinical usefulness (Kankanamge et al., 2025). Thorough prospective evaluation and established safety benchmarks are vital before these systems can be reliably integrated into live patient care (Cizmic et al., 2025).

Overall, AI-based intraoperative guidance is potentially extremely beneficial for surgeons, but must possess validation, generalizability, safety, and clinical integration before being applied safely to other patients.

Methods

A focused scoping review was performed across PubMed, IEEE Xplore, and Google Scholar to evaluate artificial intelligence applications in intraoperative surgical guidance for minimally invasive and robotic procedures. The initial search found 63 articles. There was no specific inclusion or exclusion date to find more articles. The main Boolean query used was: (“artificial intelligence” OR “computer vision” OR “deep learning”) AND (“surgical guidance” OR “intraoperative guidance”). To maximize article results, this query was used, but the OR and AND operators were changed to find more articles. Some queries that were used were: ("artificial intelligence" OR "computer vision" OR "deep learning") OR ("surgical guidance" OR "intraoperative guidance") AND ("minimally invasive surgery" OR "robotic surgery"), "minimally invasive surgery AND robotic surgery", and ("artificial intelligence" OR "computer vision" OR "deep learning") AND ("minimally invasive surgery" OR "laparoscopic surgery"). During screening, studies were excluded if they examined open surgical procedures or had little direct relation to guidance and minimally invasive platforms, resulting in 35 peer-reviewed publications. Both included and excluded articles were recorded in a spreadsheet to preserve methodological reproducibility. The eligible articles were evaluated across four criteria: methods, key findings, contributions, and remarks. To differentiate articles, each article was further examined and written in greater detail. For each included study, information was recorded on the surgical task, data/modality, dataset or sample size, and AI method, evaluation metrics, key results, and limitations. From this collection, 16 core studies directly addressing surgical guidance were selected for detailed narrative synthesis, with 10 representative benchmark studies highlighted in the Comparative Table. A limitation of the review is that it is reliant on 16 primary studies, which may leave out commercial robotic systems or other developing intraoperative guidance models.

Table 1: State of the art and history of artifical intelligence for minimally invasive/laparoscopic surgery

Study	Task	Data/modality	Method/Model	Metric	Key result	Limitation
(Oh, 2026) Artificial intelligence-driven real-time assistance in minimally invasive surgery: a technology-oriented narrative review	Comprehensive intraoperative AI assistance: anatomical segmentation, lesion detection, instrument tracking, phase recognition, and multimodal interaction modeling	Laparoscopic and robotic video feeds paired with multimodal sensor signals (kinematics, force/haptic data)	Technology-oriented narrative review (synthesizing 69 representative studies from 2011–2025) evaluating CNNs, CNN-LSTM/temporal networks, and multimodal DL fusion architectures	Real-time inference latency, lesion detection sensitivity, segmentation boundary precision, and phase classification accuracy	Maps deep learning across intraoperative tasks: CNNs for real-time instrument and anatomical contour segmentation to prevent iatrogenic injury; Deep CNNs for metastasis detection; ML for phase recognition; multimodal DL for force estimation	High latency under intraoperative resource constraints; scarce multi-institutional benchmark datasets; performance degradation across novel camera setups and tissue types; lack of model interpretability for clinical adoption
(Zhang et al., 2026) Artificial Intelligence-Based Analysis of Laparoscopic Imaging for Intraoperative Surgical Decision Support	Real-time spatial perception, tool detection/tracking, pixel-level anatomical segmentation, and procedural phase recognition	Multi-frame endoscopic and laparoscopic video feeds	YOLO family (v2,v8) and Faster R-CNN for detection; symmetric U-Net and Mask YOLACT for segmentation; DETR for spatial detection; TeCNO/MS-TCN for phase recognition; and TripNet/Rendezvous for action triplet recognition	Inference latency (FPS/ms), detection/phase classification accuracy, and spatial boundary precision	Faster R-CNN/YOLOv2 established detection feasibility; YOLOv8 achieved high-accuracy real-time detection with low head complexity; U-Net preserved critical fine-grained anatomical margins via skip connections; temporal models effectively	High latency in instance-level segmentors (Mask YOLACT) limits real-time operating room deployment; severe data leakage risks from per-frame rather than patient-level splitting; class imbalance across brief procedural

					captured surgical workflow transitions	phases; limited multi-center generalizability
(Kankanamge et al., 2025) Artificial intelligence based assessment of minimally invasive surgical skills using standardized objective metrics - A narrative review	Automated objective surgical technical skill assessment and performance evaluation in minimally invasive surgery (MIS)	Video recordings (laparoscopic, robotic, endoscopic, endovascular, ophthalmic) and kinematic sensor data from simulation/benchtop tasks and live procedures	Narrative systematic review following PRISMA guidelines evaluating 24 studies using AI/ML/DL models paired with standardized objective metrics (SOMs); study quality scored via MERSQI	Classification accuracy (63%–100%), correlation coefficients with ground truth, and standardized metric scores (GEARS, OSATS, GOALS)	AI systems reliably automate technical skill assessment across varied surgical specialties; standardized SOMs (GEARS/OSATS in >65% of studies) provide effective benchmarks, demonstrating potential to reduce observer bias and alleviate expert proctor burdens	Wide accuracy disparity (63%–100%) due to methodological and dataset heterogeneity; predominantly single-institution, cross-sectional designs; reliance on simple isolated tasks (suturing/knot tying) rather than full complex procedures; scarce public multi-center benchmark datasets
(Reynolds, 2001) The First Laparoscopic Cholecystectomy	Closed-cavity visceral organ excision (cholecystectomy) through small-port access rather than open laparotomy	Direct optical visualization via rigid side-viewing operative endoscope (no digital video sensors or AI)	CO_2 pneumoperitoneum via Veress needle, custom "galloscope" operative channel, and pistol-grip hemoclip appliers/scissors for ligation and division	Procedural completion without conversion to open laparotomy, reduction in postoperative pain and hospital stay, and effective hemostasis	Performed the first 94 laparoscopic cholecystectomies; established feasibility of closed-cavity gallbladder excision, pioneering the shift toward minimally invasive surgery	Flat 2D optical view; narrow operative channel; absent tactile feedback; challenging visualization leading to eventual adoption of an open-tube approach
(Litynsik, 1999) Mouret, Dubois, and Perissat: The Laparoscopic Breakthrough in Europe (1987-1988)	Multi-port laparoscopic cholecystectomy using shared video display for surgical team coordination	2D electronic videolaparoscopy (monocular endoscope paired with an external micro-video camera transmitting to an analog CRT monitor)	Continuous pneumoperitoneum, multi-trocar port placement, adapted gynecological instruments, and indirect guidance via a shared television screen	Successful complete laparoscopic excision without conversion to open laparotomy, reduced scar size, and shorter hospital stay	Validated reproducible videolaparoscopic cholecystectomy across multiple centers; established the "French technique"; catalyzed the worldwide transition to team-based, video-guided minimally invasive surgery	Loss of direct 3D depth perception; fulcrum effect from rigid instruments; lack of tactile feedback; early optical resolution limits
(Štádler et al., 2021) Review and current update of robotic-assisted laparoscopic vascular surgery	Robotic-assisted laparoscopic vascular surgery (e.g., AAA repair, aortofemoral bypass, aneurysm resection, endoleak management, and revascularization)	Stereoscopic 3D optical video feeds, teleoperation master-slave console kinematic inputs, and procedural clinical records	Comprehensive narrative review synthesizing clinical series, comparative surgical cohorts, and international institution experiences (da Vinci Surgical System)	Vascular anastomosis duration, aortic clamp time, perioperative complication rates, length of hospital stay, and learning curve progression	Robotic assistance resolves rigid laparoscopic constraints through 3D stereoscopic vision, wristed instruments (EndoWrist), and software tremor reduction; significantly accelerates intracorporeal vascular suturing and shortens aortic clamp/ischemia time compared to conventional laparoscopy while matching open repair durability	Complete absence of tactile/haptic feedback forces total reliance on visual cues; substantial capital and maintenance costs ($1.8M–$2M plus disposable instruments); steep learning curve requiring extensive prior laparoscopic foundation; prolonged room setup/docking times; regulatory barriers (off-label vascular use in the US); lack of large multicenter randomized controlled trials
(Pedrett et al., 2023) Technical skill assessment in minimally invasive surgery using artificial intelligence: a systematic review	Automated assessment and classification of technical surgical skill in minimally invasive surgery (MIS)	Surgical video recordings (n=25), robotic or sensor kinematic data (n=22), eye tracking (n=2), and fNIRS neuroimaging (n=2) across 50 reviewed studies	Systematic review (PRISMA framework) evaluating supervised machine learning (n=19) and deep learning (n=34) models mapped to ordinal skill tiers (e.g., novice, intermediate, expert) or interval rubrics (OSATS, GEARS)	Classification accuracy (reported in 70% of studies; range: 0.58–1.00), precision, recall/sensitivity, F1-score, and AUC-ROC	Confirms AI models reliably automate skill grading with moderate to high accuracy, primarily leveraging robotic kinematics and video motion; automates subjective proctor reviews to reduce observer fatigue and peer-review bias	High risk of bias across 84% of included studies; heavy reliance on simulated box trainers (84%) rather than real surgeries (22%); severe lack of external multicenter validation (only 4 studies tested on

						independent external datasets); all models remain in early preclinical development
(Moscovici et al., 2026) Robotic versus Laparoscopic Colectomy for Cancer	Comparative evaluation of operative technique, perioperative outcomes, and technology adoption in minimally invasive colectomy and rectal resection	Multicenter randomized controlled trials (e.g., COST, COLOR, ALaCaRT, ACOSOG Z6051, ROLARR, REAL), meta-analyses (n=52,589 left colectomies), national registry data (ACS-NSQIP, n=229,862), and learning curve cohorts	Narrative and clinical review synthesizing comparative evidence across open, conventional laparoscopic, and robotic-assisted surgical platforms	Anastomotic duration, conversion rate to laparotomy, complication/anastomotic leak rates, lymph node harvest, disease-free survival (DFS), overall survival (OS), and cases-to-proficiency learning curves	Robotic assistance significantly accelerates intracorporeal anastomosis (31.6 min vs. 43.0 min laparoscopic), flattens the learning curve (16 vs. 25 cases), lowers left colectomy complications/leaks, and reduces pelvic recurrence in rectal cancer (REAL trial: 1.6% vs. 4.0%); national adoption surged from 4.2% to 14.8% between 2014 and 2020	Higher capital and per-procedure disposable costs; prolonged operating room setup and docking times; persistent lack of tactile/haptic feedback; geographic and institutional access disparities; AI components remain conceptual rather than clinically integrated
(Vitiello et al., 2013) Emerging robotic platforms for minimally invasive surgery	Flexible-access surgical navigation, dexterous tissue manipulation, and intraoperative guidance across extraluminal, intraluminal, transluminal (NOTES), and hybrid procedures	Multimodal engineering review synthesizing mechanical robotic architectures, intraoperative imaging (fluoroscopy, ultrasound, 3T MRI), optical biopsy probes (OCT, pCLE, FLIM), and biometrics (gaze tracking, tactile force)	Comprehensive engineering framework evaluating articulated multi-DoF continuum robots, concentric-tube manipulators, dynamic active constraints (virtual fixtures), and perceptual docking human–robot interfaces	Degrees of freedom (DoFs), outer sheath diameter, tip positioning accuracy (e.g., 0.61–2.24 mm RMS error in 3T MRI concentric-tube testing), structural triangulation, and stability during tissue retraction	Categorizes flexible systems (e.g., Transport, ANUBISCOPE, MASTER, HARP/Flex, i-Snake) that overcome the rigid fulcrum effect and restricted workspaces; establishes that whole-body dynamic active constraints, perceptual docking via eye-gaze tracking, and image-constrained biomechanical modeling are foundational for transitioning from master–slave telemanipulation to shared-control surgical autonomy	High mechanical and computational complexity for multi-DoF joint coordination; lack of real-time tactile/force haptic feedback; non-rigid tissue deformation modeling bottlenecks; bulky external actuation drives; elevated acquisition costs and lack of clinical maturity
(Sone et al., 2023) Evolution of a surgical system using deep learning in minimally invasive surgery (Review)	Automated analysis of minimally invasive surgery video—encomp assing organ/anatomy identification, surgical instrument detection, surgical phase/step recognition, operative time prediction, safe incision line/surface navigation, and evaluating the progress toward autonomous surgical robotics	2D laparoscopic and robot-assisted surgical video recordings, annotated video frames/image datasets (e.g., 156,584 laparoscopic cholecystectomy images; 461 gynecological video frames; 2,854 LC images), multimodal 3D CT preoperative scans, and near-infrared fluorescence imaging	Narrative review synthesizing deep learning pipelines across computer vision and robotics: Mask R-CNN, 2D/3D CNNs, U-Net architectures for semantic segmentation, recurrent/temporal networks (CNN-LSTM, RSDNet), and autonomous robotic control frameworks (e.g., Smart Tissue Autonomous Robot [STAR], RAVEN-II, and da Vinci Research Kit [dVRK])	Classification and step recognition accuracy (e.g., 91.0%–93.2% for procedural phase recognition), segmentation overlap/accuracy (Dice coefficient: 0.84; IoU: 0.80(prostate)–0.91(kidney)), precision/sensitivity, remaining surgery time prediction error (overall average error: ~37%), and robotic task precision (e.g., 2.4×–2.9× improvement in suture spacing/bite consistency over manual execution)	Confirms that deep learning algorithms reliably recognize surgical anatomy, track instruments, and classify discrete procedural steps with high accuracy (>91%) across laparoscopic and robotic procedures; establishes that integrating real-time visual recognition with supervised robotic execution (as demonstrated by STAR for intestinal and vaginal cuff anastomosis) forms the necessary pathway toward task autonomy (Levels 1–3)	Significant drop in algorithm generalization across varied patient anatomies and multi-institution settings; "black-box" model opacity that impairs surgical trust; risk of communication latency and network anomalies during teleoperation; unresolved ethical and legal liability questions regarding autonomous error; autonomous robotic implementations remain restricted to narrow, isolated subtasks (mainly suturing) in preclinical animal/inanimate models rather than

						full clinical interventions

Past

Laparoscopic surgeries began on September 12, 1985. The first case of laparoscopic surgery was a cholecystectomy (removal of the gallbladder). Erich Mühe was able to prove that a gallbladder could be successfully removed through laparoscopic surgery using an operative laparoscope and pistol-grip hemoclip appliers/scissors, which was also known as the "Galloskop," predating French and American surgeons. Even with early rejection from the German Surgical Society in 1986, this technique established the basis for modern laparoscopic cholecystectomy, and Mühe later received international recognition from SAGES in 1999. While Mühe proved closed-cavity surgery was possible, moving away from open, direct visualization meant surgeons immediately had to deal with flat 2D laparoscopic optics, zero direct tactile feedback, and a restricted field of view. He later abandoned optics for a 2.5 cm "open tube" due to visibility challenges (Reynolds, 2001). Crucially, these early perceptual barriers shaped the modern research agenda for intelligent surgical systems, ultimately fueling the development of robotic 3D stereoscopic vision, augmented reality depth estimation, and AI-driven anatomical segmentation designed to restore what minimally invasive access took away.

However, this sightline limitation started to change through the work of French surgeons between 1987 and 1988. In March 1987, Philippe Mouret adapted gynecological equipment and introduced electronic laparoscopy. This was done by attaching an external video camera to the scope's eyepiece. Compared to Mühe looking directly through an eyepiece or tube, this was the first instance of video-guided surgery for cholecystectomy, projecting the surgical field onto a television monitor. François Dubois learned about Mouret's work and performed the first laparoscopic cholecystectomy in Paris in April 1988, later publishing his findings to prove that this specific procedure was replicable. Meanwhile, Jacques Périssat in Bordeaux transitioned from using endoscopes for gallstone lithotripsy to performing complete laparoscopic removals after recognizing that removing the entire organ was far more effective than just crushing stones inside it (Litynsik, 1999).

Together, these surgeons popularized what is known as the "French technique," starting multi-trocar port setups and shared monitor visualization. Evolving the operative field to an electronic screen transformed laparoscopy from Mühe's individual craft into a coordinated team procedure. More importantly, streaming the image onto a monitor established the basis for modern surgical guidance, creating the digital surgical video feeds that modern computer vision and robotic platforms use today.

When evaluated together, the different paths taken by Mühe and the French surgeons reveal the central compromise of early minimally invasive surgery: the constant tension between anatomical exposure and visual orientation. Mühe favored his uncompromised direct visibility

approach, choosing not to use true closed-cavity access by switching to an open subcostal tube when 2D optics proved unmanageable in an inflated abdomen (Reynolds, 2001). Conversely, Mouret, Dubois, and Périssat were willing to deal with the tricky parts of looking at a screen instead of looking right into the patient. They had to handle 2D images, an absence of depth perception, and tools moving in the opposite direction of their hands just so they could keep the small-incision method (Litynsik, 1999).

This trade-off significantly changed the trajectory of surgical technology. By converting the operating field into an external video stream, the French technique did not eliminate the visual limitations that caused Mühe to diverge to the open tube; rather, it externalized those limitations onto a monitor. Moving the image to a screen solved the immediate problem of operating room teamwork, but it introduced new perceptual challenges that the human eye alone struggled to navigate. In doing so, early video-laparoscopy created the visual interface and perceptual challenges that present-day computational methods would later address. This historical advancement–spanning direct visualization, early laparoscopic optics, video-guided surgery, and modern digital surgical video–established the direct foundation for computer vision and deep learning models able to deliver real-time AI guidance, ultimately charting the path toward future intelligent and robotic assistance. By separating the surgeon's eyes from direct contact with the patient, early-stage video laparoscopy turned operative guidance into a technological challenge, laying the essential preparation for future intelligence and robotic systems in use today.

Present

Anatomy and Visual Scene Understanding

At the core of intraoperative guidance is the computer vision task of parsing raw video feeds to establish spatial orientation, identify critical structures, and map anatomical safety margins. Recent implementations of computer vision and deep learning in minimally invasive surgery generally attempt to bridge the gap between passive digital video capture and active intraoperative decision support by parsing high-definition video feeds to detect anatomical boundaries and mitigate inadvertent tissue injury in real time (Arakaki et al., 2024) demonstrated that convolutional segmentation networks can reliably localize critical anatomy–such as the cystic duct and recurrent laryngeal nerve–with pixel-level identification accuracies exceeding 90% and evaluate surgical quality metrics with over 93% specificity. Conversely, broader engineering reviews of continuum and articulated systems (Vitiello et al., 2013) emphasized that mechanical flexibility alone cannot prevent operative complications unless instrument kinematics are synchronized with intelligent visual tracking and dynamic safety limits. Together, these outcomes show that converting surgical video into usable spatial data is necessary for guiding complex robotic maneuvers, although real-time vision algorithms cannot yet replace manual clinical judgment because current models have difficulty segmenting structures obscured by blood or adipose tissue and commonly lack external multicenter validation. This indicates that advancing from static boundary recognition toward context-aware, closed-loop AI guidance that

dynamically adapts to operative field degradation is still an essential priority for the future of intelligent robotic assistance.

Instrument Detection and Tracking

Beyond static anatomical mapping, effective guidance depends on real-time computational algorithms that identify end-effector positions, monitor tool-tissue interactions, and track articulated instruments across constrained operative workspaces. While multi-articulated robotic platforms like the da Vinci system with EndoWrist technology restore mechanical degrees of freedom and eliminate manual tremor during complex procedures like vascular reconstruction (Štádler et al., 2021), the platforms remain cognitively blind without active computer vision tracking. Recent deep learning implementations attempt to bridge this mechanical-visual gap by processing multi-frame laparoscopic video feeds to localize instrument tips and maintain visual boundaries in real time (Oh, 2026; Zhang et al., 2026). Comparative evaluations of detection pipelines demonstrate that anchor-free convolutional architectures and real-time detectors (such as YOLOv8) achieve high-accuracy spatial localization at low computational latency, while multi-object tracking algorithms (e.g., ByteTrack, BoT-SORT, and SurgiTrack) successfully maintain tool identities across consecutive frames even during rapid motion or partial occlusions (Oh, 2026; Zhang et al., 2026). In preclinical and challenge benchmarks, these models reliably predict instrument coordinates and segmentation masks, achieving mean Intersection over Union (mIoU > 0.5) to monitor tool clearance and identify retained foreign items (Zhang et al., 2026; Arakaki et al., 2025). Conversely, broader engineering assessments of continuum and flexible robotic manipulators (Vitiello et al., 2013) show that hardware articulation alone cannot ensure safe navigation through convoluted intraluminal corridors unless paired with continuous tip-pose tracking and dynamic active constraints (virtual fixtures). Together, these findings indicate that automated instrument localization is foundational for transforming passive teleoperated arms into context-aware systems capable of preventing off-target tissue damage; however, current detection pipelines exhibit significant performance degradation when tools are obscured by electrosurgical smoke or mirror-like fluid reflections, and heavy instance segmentation models often exceed acceptable operating-room latency budgets (Oh, 2026; Zhang et al., 2026). Consequently, advancing from isolated bounding-box detection toward low-latency, 3D tool-pose estimation that remains robust across heterogeneous multi-vendor camera systems is a critical priority for closed-loop robotic assistance.

During an endoluminal transluminal intervention utilizing a multi-DoF continuum manipulator (Vitiello et al., 2013), the flexible end-effector traverses a tortuous gastrointestinal corridor outside direct line-of-sight visual exposure. Because continuum systems experience non-linear shaft deflection and lack distal force feedback, real-time tool tracking and dynamic active constraints (virtual fixtures) must constantly register tip pose relative to preoperative imaging; without this synchronized computer-vision tracking, the mechanical reach of the

articulated manipulator risks exerting unrecognized shear forces and causing transmural perforation.

Surgical Phase and Workflow Recognition

Understanding the procedural setting, specifically answering “where is the surgeon right now in the operation?” provides the temporal framework needed to synchronize guidance with ongoing surgical steps. Recent deep learning implementations in minimally invasive surgery increasingly rely on computer vision architectures to decompose multi-step interventions into discrete procedural phases and predict remaining operative duration. A comprehensive review by (Sone et al., 2023) synthesized deep convolutional neural networks (CNNs) across major surgical specialties, highlighting models developed by Kitaguchi et al. that categorized laparoscopic sigmoid colon resections into 11 discrete steps with 91.9% accuracy and classified transanal total mesorectal excision (TaTME) phases with 93.2% accuracy. Building on these segmentation benchmarks, Cheng et al. (also synthesized in Sone et al., 2023) demonstrated a CNN framework achieving 91.0% phase recognition accuracy across 163 multicenter laparoscopic cholecystectomy videos (comprising 156,584 frames), while temporal architectures such as RSDNet developed by Twinanda et al, (Sone et al., 2023) and hybrid CNN-LSTM networks reviewed by (Raptis et al., 2026) successfully model inter-frame sequence dependencies on benchmark datasets like Cholec80 to forecast procedural duration and upcoming operative tasks. Together, these findings confirm that deep temporal models can reliably distinguish sequential operative stages directly from video feeds; however, current workflow architectures remain hindered by performance degradation on out-of-distribution cases caused by non-standardized procedural deviations, unpredictable anatomical variations, and substantial computational latency during multi-frame sequential modeling. Addressing these bottlenecks through lightweight temporal transformers and low-latency edge computing remains a priority to enable seamless intraoperative deployment. Clinically, knowing the active operative phase in real time allows next-generation robotic systems to dynamically update context-aware safety boundaries, automate instrument exchanges, and anticipate critical procedural risks before hazardous surgical transitions occur.

Predictive and Context-Aware Guidance

The ultimate goal of intraoperative guidance is prospective: predicting “what is most likely to happen next, and what assistance should the system provide?” to retain precision, evaluate competence, and forestall technical failure. Although automated skill assessment is frequently viewed as an offline training or credentialing metric, its current value for surgical guidance lies in continuously estimating operator proficiency to provide a foundation for prospective applications such as dynamic safety barriers, adaptive motion scaling, or robotic shared control. Pedrett et al. demonstrated that supervised machine learning and deep neural networks can reliably classify surgical performance along ordinal proficiency tiers and standardized rubrics (OSATS, GEARS) with classification accuracies up to 1.00 using video-extracted tool motions and robotic

kinematic streams across minimally invasive tasks. Similarly, the operator telemetry evaluated by Pérez-Salazar et al. used wearable sensor arrays to monitor ergonomic parameters, showing that automated analytics can capture motor behaviors and distinguish experience levels through models targeted at internal musculoskeletal fatigue rather than outward task execution quality. Together, these studies suggest that intelligent computational models can objectively assess surgical competence across both technical execution and physical operator strain, although their reported outcomes cannot be directly compared because current skill-evaluation frameworks stay confined to preclinical box-trainer simulations, rely on non-standardized evaluation measures, and lack external multi-institutional validation. This indicates that for automated assessment to function as an active intraoperative guidance mechanism rather than a retrospective grading rubric, systems must shift from static post-procedural scoring to low-latency, closed-loop pipelines that dynamically modulate robotic intervention when an operator's technical execution destabilizes.

In a surgical training environment evaluating minimally invasive peg transfer and intracorporeal knot tying (Pedrett et al., 2023), an automated skill assessment pipeline continuously tracks tool tip trajectories and economy of motion. When an operator's kinematic efficiency declines below an expert proficiency threshold, reflecting uncoordinated movement or erratic instrument paths, the underlying network identifies intermediate-to-novice state transitions in real time, demonstrating the potential for closed-loop platforms to dynamically increase motion scaling or enforce protective software boundaries before technical execution errors induce structural trauma.

Parallel to technical proficiency metrics, predictive guidance extends directly to the console operator's physical state. Recent AI approaches to intraoperative ergonomic monitoring and physical strain prediction generally attempt to overcome the unmonitored musculoskeletal stress and fatigue of robotic teleoperation to protect operator endurance and maintain surgical accuracy during lengthy procedures. Pérez-Salazar et al. demonstrated that machine learning algorithms, notably Multiple Linear Regression and Multilayer Perceptrons, can reliably predict joint-specific postural risk and muscle load ($R^2 > 0.85$, RMSE < 0.05) using wearable inertial motion tracking, electromyography, and standardized RULA scoring during simulated robotic tasks. Conversely, broader ergonomic evaluations by Štádler et al. (2021) found that while the seated robotic console eliminates the static shoulder strain inherent to standard laparoscopic standing positions, it shifts physical tension to the surgeon's neck, forearms, and wrists rather than fully eliminating procedural fatigue. Together, these studies suggest that sensor-driven predictive systems can reliably identify ergonomic failure points before physical exhaustion impairs operative control, although their findings cannot be directly equated because one evaluates algorithmic predictions on standardized, short-duration simulator drills while the other assesses macroscopic physical strain across real-world, multi-hour clinical procedures. This suggests that translating unobtrusive biomechanical telemetry from benchtop simulations toward real-time, closed-loop alerts within the operating room console is still a major opportunity for intraoperative guidance.

Synthesizing these four functional domains: visual scene analysis, instrument tracking, workflow recognition, and predictive context awareness shows that although robotic platforms and deep learning models have successfully addressed the physical rigidity and perceptual limits of early laparoscopy, the modern surgical field remains fragmented between isolated technical innovations. Static anatomical segmentation, rigid and flexible tool tracking, phase classification, and ergonomic or skill-based telemetry each demonstrate high dependability in restricted, regulated environments; however, they consistently meet barriers in real-time integration, tactile feedback, and multicenter clinical generalizability. Connecting this critical divide requires transitioning from passive, retrospective analysis to active, multimodal assistance-synchronizing robotic kinematics, operator strain tracking, and computer vision within closed-loop systems to realize truly collaborative, intelligent surgical guidance.

During a multi-hour robotic-assisted aortobifemoral bypass requiring continuous intracorporeal running vascular anastomosis (Štádler et al., 2021) the console surgeon operates under strict warm ischemia clamp deadlines without distal haptic feedback. While wrist articulation and physiological tremor filtration facilitate rapid, fine-pitch suture placement, prolonged optical engagement at the console shifts physical strain to the cervical spine and wrist flexors; in the absence of tactile resistance, unchecked neuromuscular fatigue risks unintentional suture overtensioning or tearing of calcified, friable aortic tissue.

Future and Limitations

Future Directions: Autonomous Robotics and Shared-Control Systems

While deep learning models have demonstrated high precision in segmenting intraoperative instruments and defining safe anatomical dissection planes, translating these visual outputs into direct mechanical assistance remains an unresolved frontier. However, contemporary clinical systems, such as the da Vinci platform, remain confined to teleoperated master-slave motion replication without independent decision-making, and existing visual models degrade when they encounter obscured or anatomically variable organs. Furthermore, current robotic platforms operate primarily as passive visual monitors rather than closed-loop actors, leaving surgeons vulnerable to manual execution errors during repetitive reconstructive tasks. An emerging trajectory to handle these constraints is the development of semi-autonomous and supervisory shared-control platforms (Levels 2-3 task autonomy) that can independently execute standardized, high-precision procedural steps under human monitoring (Sone et al., 2023). In preclinical evaluations, the Smart Tissue Autonomous Robot (STAR) integrated 3D visual tracking with automated suture planning, achieving 2.4-2.9-fold greater consistency in suture bite size and spacing, superior anastomotic leak pressures, and a 44% reduction in operator workload compared to manual laparoscopy. Similar platforms, such as RAVEN-II, have demonstrated automated residual tissue ablation. Nonetheless, substantial translational uncertainty persists regarding whether autonomous execution can expand beyond structured suturing to dynamic tasks like sharp dissection for emergency hemostasis. Additionally, deep learning “black box” opacity obscures algorithmic reasoning, leaving legal

liability and regulatory responsibility unresolved among surgeons, hospitals, and software developers if intraoperative automated systems malfunction.

Future Directions: Medicolegal Oversight, Credentialing, and Patient-Focused Safety Frameworks

Contemporary minimally invasive and robotic procedures are increasingly augmented by computational models designed to map tissue boundaries, predict complications, and assess technical capability. However, rapid technological deployment has outpaced institutional policy and legal jurisprudence, creating significant liability ambiguities regarding surgeon responsibility, algorithm accountability, and informed consent. Furthermore, modern common law negligence standards-specifically the transition from peer-standard defined models (the Bolam test with Bolitho refinement) toward patient-centered disclosure doctrines (the Montgomery and Modified Montgomery rulings) -leave hospitals and clinicians exposed to medical malpractice claims when opaque algorithmic tools fail or cause undetected harm, such as occult electrosurgical burns outside the laparoscope's field of view. An emerging trajectory involves standardizing clinical credentialing through simulation-based benchmarks and establishing multidisciplinary institutional oversight mechanisms, such as New Technology Committees (NTCs), to evaluate software-guided interventions. Standardized programs developed by professional societies-including the Fundamentals of Laparoscopic Surgery (FLS) and Fundamentals of Robotic Surgery (FRS)-demonstrate that objective competency testing and uniform protocols mitigate procedural complications and lower institutional liability spikes. (Jeganathan et al., 2025) Similarly, institutional bodies provide organized frameworks for tracing early clinical outcomes and managing procedural introduction. Nonetheless, significant translational uncertainty persists regarding how to formally license and regulate autonomous surgical software within disparate global legal systems. It remains unresolved how courts will divide legal liability between commercial software developers and operating surgeons when opaque artificial intelligence systems provide flawed real-time recommendations that result in adverse surgical events.

Future Directions: Optical Restoration, Core Models, and Federated Data Architectures

Contemporary intraoperative guidance systems depend heavily on deep neural networks-such as U-Net variants, YOLO detectors, and transformer models to process live laparoscopic video and robotic kinematic streams for anatomical segmentation and phase recognition. However, these models degrade sharply when confronted with dynamic intraoperative visual artifacts like electrosurgical smoke, hemorrhage, and lens fogging, while model applicability is crippled by small single-center datasets, hardware heterogeneity, and the prohibitive manual labor required for pixel-level expert annotation. To overcome these optical and data barriers, emerging research priorities include real-time optical restoration transformers, self-supervised surgical core models, and decentralized federated learning networks that do not share protected health data. Specifically, dedicated visual restoration networks like LGUformer

(Local-Global U-shared Transformer) successfully eliminate dense surgical smoke to restore clear operative fields and enhance downstream instrument tracking, whereas synthetic 3D generation tools like Vision Blender and decentralized federated networks expand dataset variety across multi-center environments (Raptis et al., 2026). Furthermore, self-supervised core models pretrained on massive uncurated surgical archives greatly reduce the manual annotation bottleneck. Nonetheless, substantial translational uncertainty persists regarding whether multistage visual dehazing and high-capacity transformer architectures can execute concurrently within strict, sub-30-millisecond latency budgets without inducing perceptual lag. In addition, deep models developed on synthetic or uncurated datasets risk exhibiting distribution changes and visual hallucinations when deployed on atypical in vivo anatomy or during sudden intraoperative hemorrhages.

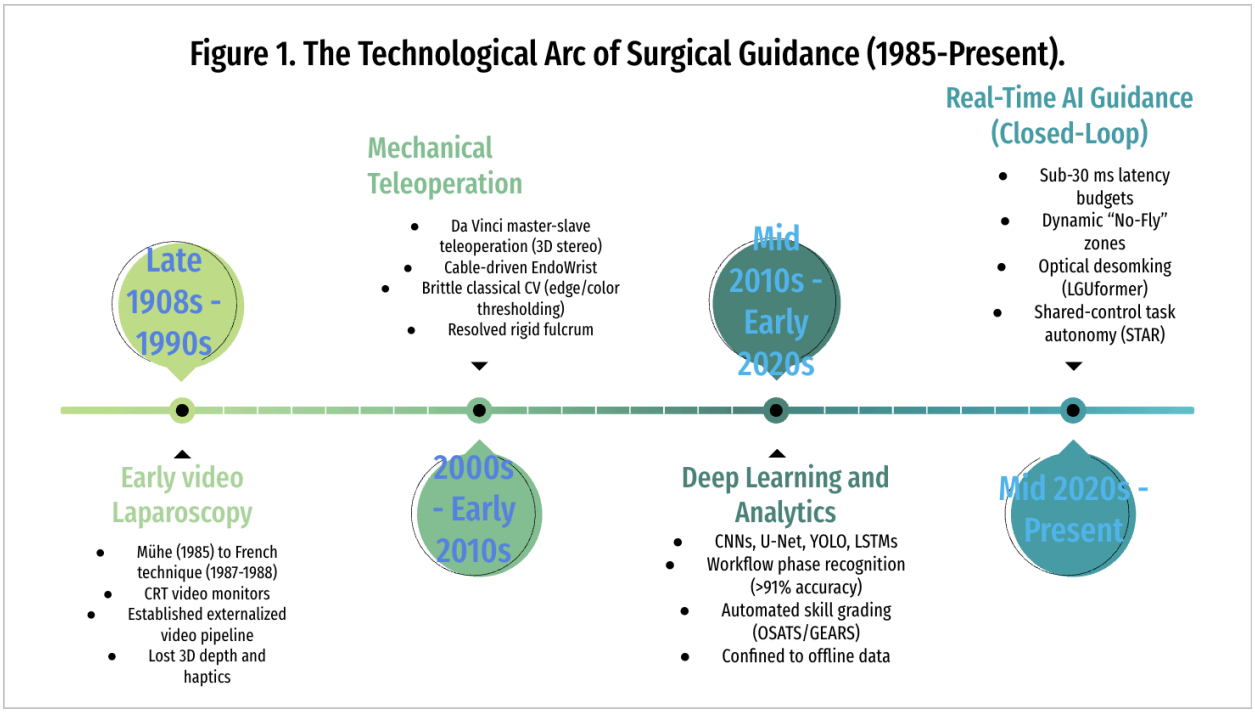


Figure 1: The Technological Arc of Surgical Guidance (1985-Present)

Limitations of Present Evidence and Methodological Constraints

Limitations of This Review

This review has several inherent methodological boundaries. First, it was conducted as a focused narrative scoping review rather than an exhaustive PRISMA-compliant systematic review of quantitative meta-analysis. The search was restricted to three electronic databases (PubMed, IEEE Xplore, and Google Scholar) using targeted keyword combinations, which may have left out relevant engineering proceedings or proprietary industry trials. In addition, the study selection used purposive sampling to trace the direct technological progression from manual laparoscopy to shared-control robotics, rather than relying on dual-blinded screening.

There was also no cut-off for publication dates, so more foundational surgical milestones, even from the mid-1980s to modern deep learning frameworks, were deliberately captured, regardless of the fact that these were both very different eras and study designs. Finally, because our synthesis centers on a curated set of 16 core papers, with 10 in the comparative table, direct head-to-head statistical comparison across these varied studies remains limited.

Despite the rapid proliferation of artificial intelligence in minimally invasive surgery, critical methodological and clinical constraints obstruct translating current guidance systems from experimental sandboxes to routine operating theaters. The foremost limitation identified across the reviewed literature is an acute dependence on retrospective, single-center datasets derived predominantly from specialized, high-volume academic institutions (Jeganathan et al., 2025; Raptis et al., 2026). This narrow sampling frame induces profound selection and spectrum biases, as algorithms trained on optimized operative fields and standardized patient cohorts consistently fail to maintain diagnostic or navigational fidelity when confronted with real-world anatomical variations, atypical pathologies, or hardware discrepancies inherent to multi-vendor laparoscopic towers (Raptis et al., 2026). Furthermore, the pervasive dependence on manual pixel-level annotation creates a severe reproducibility bottleneck, introducing interobserver variability that compromises ground-truth reliability over the surgical benchmarks (Raptis et al., 2026).

Beyond data architecture, fundamental technical and operational bottlenecks constrain real-time clinical deployment. The intraoperative visual environment is inherently non-rigid and turbulent; common surgical occurrences–including smoke dissemination from electrosurgical dissections, sudden hemorrhage, optical glare, and mechanical camera lens soiling–frequently degrade the perceptual accuracy of segmentation networks, leading to tracking dropouts or erroneous edge delineation (Sone et al., 2023; Raptis et al., 2026). Concurrently, the computing overhead required to execute multi-stage deep learning pipelines- such as concatenating convolutional feature extraction, temporal sequence modeling, and optical restoration filters–threatens the ultra-low latency threshold (sub-30 milliseconds) necessary to avert perceptual delay during robot manipulation (Raptis et al., 2026). Finally, the abiding “black box” architecture of deep neural networks poses an unaddressed patient-safety hazard. The inability to inspect algorithmic reasoning in real time obscures failure modes, rendering it impossible for surgeons to differentiate between valid navigational cues and erroneous hallucinations through critical dissection steps, thereby compounding legal vulnerabilities and delaying institutional clearance (Sone et al., 2023; Jeganathan et al., 2025).

Conclusion

This scoping review traces the technological arc of surgical guidance in minimally invasive, laparoscopic, and robotic procedures, from passive visualization aids to context-aware, multimodal deep learning systems. Contemporary platforms show outstanding proficiencies in instrument localization, workflow phase classification, and anatomical landmark recognition under controlled conditions (Sone et al., 2023; Raptis et al., 2026). However, the field currently

sits at a key translational impasse: present models largely remain retrospective observers confined to scientific settings and constrained by hardware heterogeneity, optical degradations, and acute legal uncertainties related to algorithmic liability and patient disclosure standards (Jeganathan et al., 2025; Raptis et al., 2026).

Overcoming these translational hurdles requires a coordinated shift across engineering, clinical validation, and regulatory policy. Technologically, the field must transition from static, single-center architectures to federated learning networks, self-supervised surgical core models, and real-time optical restoration frameworks that preserve visual clarity without violating strict latency budgets (Raptis et al., 2026). Concurrently, institutions must establish standardized credentialing protocols and adaptive medical device evaluation committees to thoroughly audit algorithmic reliability, protect patient safety, and resolve complex liability divisions between software developers and clinical operators (Sone et al., 2023; Jeganathan et al., 2025).

Current evidence suggests that over the next several years, the trajectory of surgical guidance is likely to focus on the gradual realization of supervised shared-control robotics (Levels 2-3 autonomy) (Sone et al., 2023). Future clinical translation is unlikely to center on the displacement of human surgical judgment by autonomous systems, but may increasingly emphasize the deployment of context-aware digital co-pilots that automate high-precision, repetitive tasks–such as intracorporeal suturing–while actively monitoring operative fields for out-of-sight complications (Sone et al., 2023; Jeganathan et al., 2025). Through aligning computational innovation together with strong legal governance and multicenter prospective validation, AI-augmented guidance has the potential to transform minimally invasive surgery into a standardized, safer, and inherently more reproducible discipline.

Works Cited

Arakaki, Shintaro et al. “Artificial Intelligence in Minimally Invasive Surgery: Current State and Future Challenges.” JMA journal vol. 8,1 (2025): 86-90. doi:10.31662/jmaj.2024-0175

Cizmic, Amila et al. “Telestration with augmented reality in minimally invasive and robotic-assisted surgery: a scoping review.” Surgical endoscopy vol. 39,12 (2025): 8000-8013. doi:10.1007/s00464-025-12380-2

Dhingra, Juslyn et al. “Clear Vision, Clear Savings: Enhancing Efficiency in Minimally Invasive Surgery.” JSLS : Journal of the Society of Laparoendoscopic Surgeons vol. 29,3 (2025): e2025.00023. doi:10.4293/JSLS.2025.00023

Jeganathan, J Ravichandran et al. “Minimally invasive surgery: a historical and legal perspective on technological transformation.” Journal of robotic surgery vol. 19,1 408. 21 Jul. 2025, doi:10.1007/s11701-025-02589-7

Kankanamge, D et al. “Artificial intelligence based assessment of minimally invasive surgical skills using standardised objective metrics - A narrative review.” American journal of surgery vol. 241 (2025): 116074. doi:10.1016/j.amjsurg.2024.116074

Litynski, G S. “Profiles in laparoscopy: Mouret, Dubois, and Perissat: the laparoscopic breakthrough in Europe (1987-1988).” JSLS : Journal of the Society of Laparoendoscopic Surgeons vol. 3,2 (1999): 163-7.

Moscovici, Avihai et al. “Robotic versus Laparoscopic Colectomy for Cancer.” Advances in surgery vol. 60,1 (2026): 1-14. doi:10.1016/j.yasu.2026.03.001

Oh, Jieun. “Artificial intelligence-driven real-time assistance in minimally invasive surgery: a technology-oriented narrative review.” Journal of minimally invasive surgery vol. 29,2 (2026): 70-81. doi:10.7602/jmis.2026.29.2.70

Pedrett, Romina et al. “Technical skill assessment in minimally invasive surgery using artificial intelligence: a systematic review.” Surgical endoscopy vol. 37,10 (2023): 7412-7424. doi:10.1007/s00464-023-10335-z

Pérez-Salazar, Manuel J et al. “Correlation Study and Predictive Modelling of Ergonomic Parameters in Robotic-Assisted Laparoscopic Surgery.” Sensors (Basel, Switzerland) vol. 24,23 7721. 3 Dec. 2024, doi:10.3390/s24237721

Raptis, S. P., et al. "Artificial Intelligence Analysis of Minimally Invasive Surgery Data." Journal of Robotic Surgery, vol. 20, article 186, 2026, https://doi.org/10.1007/s11701-026-03155-5.

Reynolds, W Jr. “The first laparoscopic cholecystectomy.” JSLS : Journal of the Society of Laparoendoscopic Surgeons vol. 5,1 (2001): 89-94.

Sone, K., Tanimoto, S., Toyohara, Y., Taguchi, A., Miyamoto, Y., Mori, M., Iriyama, T., Wada-Hiraike, O., Osuga, Y."Evolution of a surgical system using deep learning in minimally invasive surgery (Review)". Biomedical Reports 19.1 (2023): 45.

Štádler, Petr et al. “Review and current update of robotic-assisted laparoscopic vascular surgery.” Seminars in vascular surgery vol. 34,4 (2021): 225-232. doi:10.1053/j.semvascsurg.2021.10.007

Vitiello, Valentina et al. “Emerging robotic platforms for minimally invasive surgery.” IEEE reviews in biomedical engineering vol. 6 (2013): 111-26. doi:10.1109/RBME.2012.2236311

Zhang, Zheyuan et al. “Artificial Intelligence-Based Analysis of Laparoscopic Imaging for Intraoperative Surgical Decision Support.” Annual review of biomedical engineering vol. 28,1 (2026): 135-162. doi:10.1146/annurev-bioeng-110824-114236