Kasimieh *et al. Art Int Surg.* 2025;5:557-71
DOI: 10.20517/ais.2025.92

Artificial Intelligence Surgery

Systematic Review

Open Access

Deep learning in real-time image-guided surgery: a systematic review of applications, methodologies, and clinical relevance

Omar Kasimieh[1], Mohamed Mustaf Ahmed[2,3], Zhinya Kawa Othman[4], Ifrah Ali[3], Mulki Mukhtar Hassan[3], Philine Muriel Maulion[5], Adetola Emmanuel Babalola[6], Olalekan John Okesanya[7,8], Bonaventure Michael Ukoaka[9], Francesco Branda[10], Don Eliseo Lucero-Prisno III[11,12,13]

[1]Department of Medicine, University of the East Ramon Magsaysay Memorial Medical Center, Quezon City 1113, Philippines.
[2]SIMAD Institute for Global Health, SIMAD University, Mogadishu 2526, Somalia.
[3]Faculty of Medicine and Health Sciences, SIMAD University, Mogadishu 2526, Somalia.
[4]Department of Pharmacy, Kurdistan Technical Institute, Sulaymaniyah 46001, Iraq.
[5]Faculty of Medicine, Manila Central University, Caloocan City 1400, Philippines.
[6]Kornberg School of Dentistry, Temple University, Philadelphia, PA 19140, USA.
[7]Department of Public Health and Maritime Transport, University of Thessaly, Volos 3822, Greece.
[8]Department of Medical Laboratory Science, Neuropsychiatric Hospital, Abeokuta 110101, Nigeria.
[9]Community and Clinical Research Division, First On-Call Initiative, Port Harcourt 500001, Nigeria.
[10]Unit of Medical Statistics and Molecular Epidemiology, Campus Bio-Medico University of Rome, Rome 00128, Italy.
[11]Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.
[12]Center for Research and Development, Cebu Normal University, Cebu 6000, Philippines.
[13]Research Unit, Bukidnon State University, Malaybalay City 8700, Philippines.

Correspondence to: Dr. Mohamed Mustaf Ahmed, SIMAD Institute for Global Health, SIMAD University, Mogadishu 2526, Somalia. E-mail: momustafahmed@simad.edu.so

How to cite this article: Kasimieh O, Ahmed MM, Othman ZK, Ali I, Hassan MM, Maulion PM, Babalola AE, Okesanya OJ, Ukoaka BM, Branda F, Lucero-Prisno III DE. Deep learning in real-time image-guided surgery: a systematic review of applications, methodologies, and clinical relevance. *Art Int Surg.* 2025;5:557-71. https://dx.doi.org/10.20517/ais.2025.92

Received: 8 Sep 2025 **First Decision:** 30 Oct 2025 **Revised:** 31 Oct 2025 **Accepted:** 12 Nov 2025 **Published:** 16 Dec 2025

Academic Editor: Takeaki Ishizawa **Copy Editor:** Xing-Yue Zhang **Production Editor:** Xing-Yue Zhang

Abstract

Aim: Real-time image guidance using deep learning is being increasingly used in surgery. This systematic review aims to characterize intraoperative systems, mapping applications, performance and latency, validation practices, and the reported effects on workflow and patient-relevant outcomes.

Methods: A systematic review was conducted on PubMed, Embase, Scopus, ScienceDirect, IEEE Xplore, Google Scholar, and Directory of Open Access Journals from December 31, 2024. Eligible English-language, peer-reviewed diagnostic accuracy, cohort, quasi-experimental, or randomized studies (2017-2024) evaluated the learning for real-time intraoperative guidance. Two reviewers screened, applied the Joanna Briggs Institute checklists, and extracted the design, modality, architecture, training, validation, performance, and latency. Heterogeneity precluded the meta-analysis.

Results: Twenty-seven studies spanning laparoscopic, neurosurgical, breast, colorectal, cardiac, and other workflows met the criteria. The modalities included red-green-blue laparoscopy or endoscopy, ultrasound, optical coherence tomography, cone-beam computed tomography, and stimulated Raman histology. The architectures were mainly convolutional neural networks with frequent transfer learning. Reported performance was high, with classification accuracy commonly 90%-97% and segmentation Dice or intersection over union up to 0.95 at operating-room-compatible speeds of about 20-300 frames per second or sub-second per-frame latency; volumetric pipelines sometimes required up to 1 min. Several systems demonstrated intraoperative feasibility and high surgeon acceptance, yet fewer than one quarter reported external validation and only a small subset linked outputs to patient-important outcomes.

Conclusion: Deep-learning systems for real-time image guidance exhibit strong technical performance and emerging workflow benefits. Priorities include multicenter prospective evaluations, standardized reporting of latency and external validation, rigorous human factors assessment, and open benchmarking to demonstrate generalizability and patient impact.

Keywords: Deep learning, image-guided surgery, intraoperative imaging, real-time guidance, convolutional neural networks, systematic review

INTRODUCTION

Surgery remains a cornerstone of modern healthcare, with approximately 313 million operations performed worldwide each year; paradoxically, only 6% of these are performed in the world's poorest nations[1]. Postoperative mortality is now recognized as a major global health burden, with an estimated 4.2 million deaths within 30 days of surgery each year, accounting for 7.7% of all deaths[2]. Real-time image guidance has been shown to mitigate intraoperative errors; for instance, three-dimensional (3-D) imaging prompts a revision of surgical plans in roughly 20% of orthopedic cases[3], while augmented-reality overlays improve spatial orientation and usability in laparoscopic liver surgery[4]. Deep learning (DL) has accelerated these advances in recent years. High-resolution optical coherence tomography (OCT) has > 90% sensitivity and specificity for breast-margin assessment[5], and stimulated Raman histology matched board-certified pathologists with 94.6% diagnostic accuracy in a multicenter randomized trial[6]. In workflow analytics, DL models now recognize surgical phases with > 90% accuracy across millions of laparoscopic frames[7,8]. Regulatory activity mirrors this momentum: the United States Food and Drug Administration (FDA) has authorized more than 690 artificial intelligence (AI)/machine learning (ML)-enabled medical devices to date, although most target preoperative or diagnostic imaging rather than intraoperative support[9]. However, translating these capabilities into routine operating-room practice remains difficult: adoption typically requires high up-front capital expenditure (CAPEX) for operating room (OR)-grade imaging and computing, substantial staff training and credentialing, workflow redesign to meet real-time constraints, and organizational change management to overcome resistance; interoperability and data governance hurdles further slow integration[10,11].

Despite encouraging test-bench metrics, external validation, generalizability, and real-time integration remain under-reported. Critical reviews highlight small, single-center datasets, opaque "black-box" reasoning, and inconsistent latency reporting, which hamper bedside adoption[12]. Even mature interpretability techniques lack consensus guidelines for surgical imaging[13]. Regulatory frameworks and evidence standards are evolving, and forthcoming United States advisory committee hearings have only begun to address the safety of generative AI in surgical devices[14]. Consequently, surgeons, developers, and policymakers lack a consolidated view of where DL already works reliably in the OR, where the evidence is insufficient, and which methodological decisions influence its clinical relevance. Emerging directions, such as generative AI for synthetic intraoperative data and multimodal fusion that combines endoscopic video with ultrasound and preoperative computed tomography (CT), may enhance robustness and context awareness but will require rigorous clinical validation and governance[15,16]. Given the high stakes of intraoperative decision-making and the sheer pace of algorithmic innovation, a systematic synthesis is needed to map the applications, technical pipelines, and reported clinical impacts of real-time DL systems in image-guided surgery. By linking performance metrics to workflow context and annotation quality, this study aims to clarify whether impressive accuracies translate into tangible reductions in revision rates, operating time, or margin-positive resections.

METHODS

Reporting framework and protocol registration

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement[17]. The protocol was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO) database (registration ID: CRD420251012412).

Eligibility criteria

We included peer-reviewed original studies, diagnostic test accuracy, cohort, quasi-experimental, and randomized controlled trial designs that evaluated DL-based systems for real-time intraoperative image guidance in human surgical procedures and were published between January 1, 2015, and December 31, 2024. Studies were eligible regardless of surgical specialty, imaging modality, geographic setting, or sample size. We excluded reviews, editorials, commentaries, conference abstracts, letters, single-patient case reports, and non-English articles to ensure methodological consistency in reporting.

Information sources, search strategy, and study selection

A comprehensive search was conducted in seven electronic databases [PubMed, Embase, Scopus, ScienceDirect, IEEE Xplore, Google Scholar, and the Directory of Open Access Journals (DOAJ)] from inception to December 31, 2024. Controlled vocabulary [e.g., MeSH (Medical Subject Headings), Emtree (Excerpta Medica Tree)] and free-text terms for DL ("deep learning," "neural network," "machine learning," "artificial intelligence") were combined with real-time imaging ("real-time," "live," "immediate") and surgical guidance ("image-guided," "surgery," "procedure," "augmented reality") using Boolean operators. All retrieved citations were imported into Rayyan systematic review software[17] for duplicate removal and screening. Two reviewers (MMA and OK) independently screened the titles and abstracts and then assessed the full texts for eligibility. Discrepancies were resolved through discussion, and the IA adjudicated the matter when necessary. The selection process is illustrated in Figure 1.

Quality appraisal and data extraction

The methodological quality of the included studies was assessed by two reviewers (OK and ZKO) using the Joanna Briggs Institute critical appraisal checklists appropriate for each design (diagnostic test accuracy, cohort, and quasi-experimental studies)[18]. Disagreements were resolved by consensus or consultation with MMH. Data extraction was performed independently by IA and MMH using a standardized form to capture the study characteristics (authors, year, country), surgical context, imaging modality, DL architecture, training protocol, preprocessing/augmentation methods, validation approach, annotation procedures, performance metrics, real-time capability, and reported or inferred clinical relevance. Real-time performance was defined variably across studies, most commonly as ≥ 20 frames per second (FPS) or

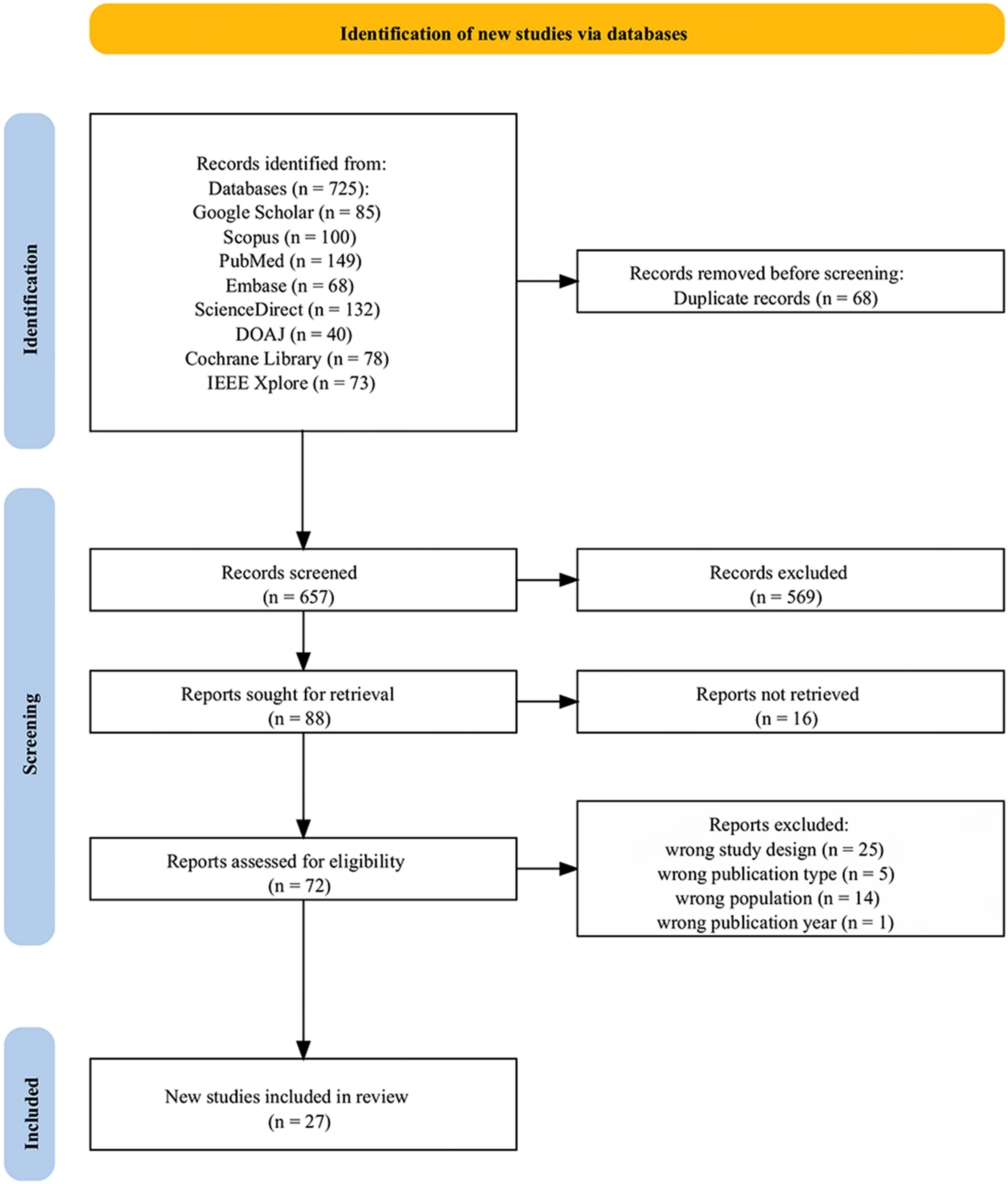

Figure 1. PRISMA flow chart of study selection. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; DOAJ: Directory of Open Access Journals; IEEE: Institute of Electrical and Electronics Engineers. The PRISMA flowchart was created with Rayyan (https://www.rayyan.ai/).

inference latency ≤ 1 s per frame; studies self-reporting "real-time" outside these bounds were still included but noted as such. The extracted data were cross-checked, and any inconsistencies were clarified through consultation with the MMA. No studies were excluded based on their risk of bias ratings, and the appraisal findings informed the narrative interpretation of the results.

Data synthesis

Substantial heterogeneity in study designs, imaging modalities, algorithmic architectures, training regimens, and outcome measures precluded quantitative meta-analysis. The findings were synthesized narratively as follows. We present tables detailing the study characteristics, technical pipelines, preprocessing strategies, annotation quality, core performance metrics, inference latencies, and real-time capabilities alongside the reported clinical impacts [Figure 2].

RESULTS

This systematic review synthesized data from 27 peer-reviewed studies published between 2017 and 2024, each evaluating the performance, training approaches, and clinical relevance of DL systems applied to surgical or image-guided interventions [Table 1]. The studies targeted a broad array of surgical procedures, with purely laparoscopic interventions being the most frequent (n = 7): laparoscopic hepatectomy[18], laparoscopic cholecystectomy[23,25,26,32], laparoscopic tubal sterilization[20], laparoscopic sigmoidectomy[8], and liver resection or staging laparoscopy with SmartLiver augmented reality (AR) guidance[15].

The imaging modalities were equally diverse. Red-green-blue (RGB) laparoscopy dominated minimally invasive workflows[8,18-20,25,26,32], while RGB endoscopy served endoscopic applications[19,33,36,37]. Ultrasound techniques, including transthoracic and transoesophageal ultrasound[21], 3-D intraoperative ultrasound[38], and intraoperative ultrasound[40], are prominent in cardiac and oncologic settings. OCT has enabled high-resolution tissue assessment in breast-conserving surgery[5,30] and *ex vivo* colorectal resection[34]. Cone-beam CT underpinned image-guided neurosurgery[24] and orthopedic procedures[39]. Confocal laser endomicroscopy provides *in vivo* histology[22], and stimulated Raman histology offers label-free intraoperative diagnosis[6].

Architecturally, convolutional neural networks (CNNs) and their variants form the backbone of studies. Segmentation tasks predominantly employed U-Net and its derivatives, U-Net for catheter and prostate segmentation[19] and U-Net+MobileNet for catheter overlays[35], with ensemble methods incorporating U-Net for colorectal tumor delineation[40]. Classification and margin assessment leveraged Inception-v3 for OCT-based breast margin evaluation[5] and Inception-Residual Network-v2 (ResNet) for stimulated Raman histology diagnosis[6]. Localization and tracking combined ResNet-18 with spatial-channel attention and You Only Look Once version 5 (YOLOv5) plus Deep Simple Online Real-Time Tracking (DeepSORT) in cardiac ultrasound[21]. Phase and instrument recognition utilized Xception pretrained on ImageNet for laparoscopic hepatectomy workflows[18] and ResNet-50 with squeeze-and-excitation plus long short-term memory (LSTM) for Cholec80 tool detection[32].

The performance metrics uniformly indicated high diagnostic quality [Table 2]. OCT-based margin assessment achieved 90% accuracy, 90% sensitivity, and 91.7% specificity[5], with follow-up studies reporting 94% accuracy and 96% sensitivity[30]. Colorectal optical biopsy models exceeded 99% specificity and reached 100% sensitivity[34]. Segmentation Intersection over Union (IoU) peaked at 0.95 for catheter overlays in urologic laparoscopy[19], and ultrasound-based tumor segmentation achieved a Dice coefficient of 0.84 in real-time colorectal workflows[40]. In pituitary surgery, landmark detection ran at 298 FPS with an IoU of 67% and 88.7% surgeon approval[33].

Several tools supported real-time decision support in the OR. The dissection-line guidance overlay developed by Smithmaitrie *et al.* achieved 95.7% surgeon acceptance[26], and multitask model developed by Jalal *et al.* for simultaneous phase and instrument recognition ran at 32 FPS with a tool mean average precision (mAP) of 95.6% and a phase F1 score of 70.1%[32]. The real-time AR registration model reported by

Table 1. Study characteristics, datasets, and technical details

Author(s)	Year	References	Study objective	Study design	Dataset (type & size)	Surgical context	Imaging modality	Deep-learning architecture	Training protocol
Padovan	2022	[19]	Develop real-time 3-D model registration from laparoscopic RGB video	Diagnostic test accuracy study	971 real frames (segmentation) + 115,000 synthetic frames (rotation) + 96 evaluation frames	Robot-assisted radical prostatectomy & partial nephrectomy	RGB endoscopy	U-Net (segmentation); ResNet-50 (rotation)	70%/15%/15% train/validation/test; synthetic rotation data
Boonkong	2024	[20]	Detect uterus and fallopian tubes (even when occluded) in laparoscopic sterilization	Diagnostic test accuracy study	800 manually annotated frames + 5 cadaveric laparoscopic videos	Laparoscopic tubal sterilization	RGB laparoscopy	YOLOv10 with key-point & ellipse heads	80%/10%/10% split; fine-tuned on YouTube data
Singla	2018	[5]	Distinguish malignant *vs.* healthy breast tissue using OCT	Diagnostic test accuracy study	219,000 image patches from 48 specimens	Breast-conserving surgery margin assessment	Optical coherence tomography	Inception-v3 (transfer learning)	Fine-tuning with augmentation
Zeye Liu	2023	[21]	Identify, localize, and track cardiac structures during surgery	Diagnostic test accuracy study	17,114 labeled ultrasound views (79 videos)	Structural-heart interventions	Transthoracic & transoesophageal ultrasound	ResNet-18 + spatial-channel attention; YOLOv5 + DeepSORT	Supervised learning
Kimimasa Sasaki	2022	[18]	Automatic recognition of surgical steps in laparoscopic hepatectomy	Diagnostic test accuracy study	40 videos (~8.1 million frames)	Laparoscopic hepatectomy	RGB laparoscopy	Xception pretrained on ImageNet	Supervised learning
Marc Aubreville	2017	[22]	Classify oral squamous-cell carcinoma from CLE images	Diagnostic test accuracy study	11,000 images from 12 patients (116 video sequences)	Oral cancer surgery	Confocal laser endomicroscopy	Custom CNN; Inception-v3 transfer learning baseline	From-scratch & transfer
Chinedu I. Nwoye	2023	[23]	Benchmark action-triplet detection (instrument + action + target)	Diagnostic test accuracy study	100,900 frames from 50 laparoscopic cholecystectomies	Laparoscopic cholecystectomy	RGB laparoscopy	Eleven submitted architectures (CNNs, Transformers, MIL, GNNs)	Supervised + weak supervision
Xiaoxuan Zhang	2023	[24]	Improve intra-operative CBCT with uncertainty-guided DL synthesis	Diagnostic test accuracy study	20 simulated CBCT cases + 7 patient cases	Image-guided neurosurgery (tumor, epilepsy, trauma)	Cone-beam CT	3-D Bayesian conditional GAN	Supervised with Monte-Carlo dropout
Schneider	2020	[15]	Assess feasibility of SmartLiver AR navigation	Cohort study	18 laparoscopic liver procedures	Liver resection or staging laparoscopy	3-D laparoscopic video + CT	CNN liver-surface segmentation	Supervised learning
Caballas	2020	[25]	Visual guidance for motion-based laparoscopic palpation	Diagnostic test accuracy study	428 polygon-annotated images	Laparoscopic cholecystectomy	RGB laparoscopy	YOLACT++	Supervised learning
Smithmaitrie	2024	[26]	Detect anatomical landmarks and guide dissection line	Diagnostic test accuracy study	3,200 frames from 40 videos	Laparoscopic cholecystectomy	RGB laparoscopy	YOLOv7	Supervised learning
Török	2018	[27]	Segment submucous fibroids and dissection plane	Diagnostic test accuracy study	6,288 images from 13 videos	Hysteroscopic fibroid resection	Hysteroscopic video	FCN-8s/16s/32s ensemble	Supervised learning
Lin	2018	[28]	Combine 3-D shape reconstruction and	Diagnostic test accuracy study	Qualitative dataset (size not specified)	Minimally invasive laryngeal surgery	Structured-light RGB + hyperspectral	CNN (shape); SSRNet (spectral super-	Supervised learning

			hyperspectral imaging				imaging	resolution)	
Blokker	2022	[29]	Classify glioma *vs.* normal brain with THG microscopy	Diagnostic test accuracy study	12,624 images (23 patients)	Brain tumor surgery	Third-harmonic-generation microscopy	Fully convolutional network	Monte-Carlo cross-validation
Mojahed	2019	[30]	Classify cancerous *vs.* non-cancerous breast OCT regions	Diagnostic test accuracy study	36,800 B-scans (46 specimens)	Breast-conserving surgery	Optical coherence tomography	Custom 11-layer CNN	Five-fold cross-validation
Tai	2021	[31]	AR-haptic guidance for precise lung biopsy	Quasi-Experimental study	341 COVID-19 patients + 1,598 controls + 24 surgeons	CT-guided lung biopsy	CT-based augmented reality	WPD-CNN-LSTM; ResNet	Five-fold cross-validation
Jalal	2023	[32]	Surgical phase and tool recognition in Cholec80	Diagnostic test accuracy study	80 laparoscopic cholecystectomy videos	Laparoscopic cholecystectomy	RGB laparoscopy	ResNet-50 + squeeze-excitation + LSTM	Supervised + weak supervision
Hollon	2020	[6]	Rapid brain-tumor diagnosis using SRH	Randomized control trial	2.5 million SRH patches (415 patients) + 278 trial patients	Brain tumor resection	Stimulated Raman histology	Inception-ResNet-v2	Supervised learning
Mao	2024	[33]	PitSurgRT: real-time localization in pituitary surgery	Diagnostic test accuracy study	635 annotated frames from 64 surgeries	Endoscopic trans-sphenoidal pituitary surgery	RGB endoscopy	HRNet dual-task heads	Five-fold cross-validation; staged training
Kitaguchi	2019	[8]	Automatic phase recognition in sigmoidectomy	Diagnostic test accuracy study	7.8 million frames (71 cases)	Laparoscopic sigmoidectomy	RGB laparoscopy	Inception-ResNet-v2 + LightGBM	Hold-out validation (63/8 cases)
Zeng	2020	[34]	PR-OCT classification of colorectal tissue	Diagnostic test accuracy study	26,000 OCT images (24 patients)	Colorectal resection (*ex-vivo*)	Swept-source OCT	RetinaNet (ResNet-18 + feature pyramid)	Supervised learning
Tanzi	2021	[35]	Catheter segmentation and 3-D overlay in RARP	Diagnostic test accuracy study	15,570 frames (five videos)	Robot-assisted radical prostatectomy	RGB laparoscopy	U-Net + MobileNet backbone	Supervised learning
Podlasek	2020	[36]	Real-time CNN polyp detection in colonoscopy	Diagnostic test accuracy study	79,284 frames + 2,678 photos	Diagnostic colonoscopy	RGB endoscopy	RetinaNet + EfficientNet-B4	Supervised learning
Sato	2022	[37]	Segment recurrent laryngeal nerve during oesophagectomy	Diagnostic test accuracy study	3,040 images (28 patients)	Thoracoscopic oesophagectomy	RGB endoscopy	DeepLab v3+	Transfer learning from PASCAL-VOC
Canalini	2019	[38]	Register intra-operative US volumes in glioma surgery	Diagnostic test accuracy study	31 3-D ultrasound volumes (RESECT + BITE datasets)	Glioma neurosurgery	3-D intra-operative ultrasound	3-D U-Net	Supervised learning
Mekki	2023	[39]	3-D localization of guidewires from two fluoroscopy views	Diagnostic test accuracy study	10,000 simulated images + 36 CBCT sets (five cadavers)	Orthopedic trauma guidewire placement	Fluoroscopy + cone-beam CT	Mask R-CNN + Key-point R-CNN	Simulated supervised training
Geldof	2023	[40]	Real-time tumor segmentation in colorectal ultrasound	Diagnostic test accuracy study	179 ultrasound images (74 patients)	Colorectal cancer surgery	Intra-operative ultrasound	Ensemble of MobileNetV2, ResNet-18/50, U-Net, Xception	Transfer learning; augmentation

3-D: Three-dimensional; RGB: red-green-blue; OCT: optical coherence tomography; CLE: confocal laser endomicroscopy; CNN: convolutional neural network; CNNs: convolutional neural networks; MIL: multiple instance learning; GNNs: graph neural networks; CBCT: cone-beam computed tomography; DL: deep learning; CT: computed tomography; GAN: generative adversarial network; AR: augmented reality; YOLACT++: You Only Look At CoefficienTs++; FCN: fully convolutional network; THG: third-harmonic generation; OCT: optical coherence tomography; COVID-19: Coronavirus Disease 2019; WPD: weighted pyramid decomposition; LSTM: long short-term memory; SRH: stimulated Raman histology; PR-OCT: polarization-resolved optical coherence tomography; RARP: robot-assisted radical prostatectomy; PASCAL-VOC: Pattern

Analysis, Statistical Modelling and Computational Learning - Visual Object Classes; US: ultrasound; CBCT: cone-beam computed tomography; R-CNN: region-based convolutional neural network.

Padovan *et al.* registered 3-D overlays at 25-30 FPS with intersection-over-union up to 0.95 and rotation errors ≤ 5°, improving intraoperative spatial orientation[19]. The guide-wire navigation system developed by Mekki *et al.* in orthopedic trauma achieved tip and directional errors of 1.8 mm and 2.7° in under 5 s per step, reducing radiation exposure while enhancing placement accuracy[39]. Nwoye established a public benchmark for fine-grained instrument-action-target analytics in laparoscopic cholecystectomy, reporting triplet average precision (AP) of 18.8%-35% on over 100,000 frames[23]. OCT-based margin assessment is promising for real-time margin status, potentially reducing re-excisions[5]. Pathologist-level diagnosis without an on-site neuropathologist was achieved in less than 150 s per case[6]. Intraoperative OCT segmentation can reduce re-excision via immediate feedback[30]. The colorectal ultrasound model has the potential to reduce positive margins intraoperatively[40]. Real-time cardiac structure tracking matches specialist performance and streamlines the workflow in areas where echocardiography expertise is scarce[21]. SmartLiver AR navigation has demonstrated feasibility and high surgeon acceptance[15]. A summary linking the core metrics to the reported clinical/workflow outcomes is provided in Table 3.

DISCUSSION

This systematic review synthesizes 27 peer-reviewed studies (2017-2024) encompassing laparoscopic, neurosurgical, breast, colorectal, cardiac, and other image-guided surgical workflows. The findings indicate that contemporary DL pipelines can achieve high diagnostic performance at clinically viable frame rates while also beginning to report task-level effects in the OR. This pattern aligns with a broader transition from isolated proofs-of-concept to context-aware assistants that integrate recognition of anatomy, instruments, and workflow, and begin to assess their effects on decision-making, ergonomics, and safety. Recent systematic reviews addressing DL applied to surgery have also emphasized that most applications are still early in the validation cycle, requiring more robust clinical evidence on which to ascertain real-world utility[41]. Simultaneously, developments can also be seen in surgical phase recognition, particularly in laparoscopic cholecystectomy where DL architectures had been able to reach robust frame-level classification and intraoperative usage[42]. Further studies have highlighted the need for and importance of harmonized evaluation schemes between institutions due to the effect of dataset splitting strategies and center-specific differences, which compromise the generalizability of models[43,44]. Contemporary syntheses in surgery and endoscopy similarly suggest that real-time feasibility is now common when models are engineered for throughput, although standardized reporting for end-to-end latency, human factors, and failure modes remains inconsistent across studies[45-47].

A central finding was the practical viability of video-rate inference in the OR. Many of the included systems paired lightweight backbones with task-specific pre- and post-processing, reflecting reports that one-stage detectors, efficient segmenters, and tracker stacks can sustain clinical frame rates on commodity hardware. External commentaries note that "FPS" claims are often hardware-dependent and rarely audited under OR load, complicating comparisons and necessitating shared measurement templates for total pipeline latency and integration overhead[48,49]. The task-level effects observed in this study are consistent with the growing evidence in several specialties. In neurosurgery, stimulated Raman histology combined with DL has repeatedly demonstrated near real-time intraoperative diagnosis with prospective clinical validation, supporting the feasibility claim without reiterating the metrics from our results[50].

Table 2. Preprocessing techniques, annotation quality, performance, and clinical relevance

Author(s)	References	Preprocessing/augmentation	Core performance metrics	Real-time capability	Reported/inferred clinical relevance
Padovan	[19]	Manual frame selection; three-class masks; synthetic rotation augmentation	IoU: 0.95 (catheter), 0.73 (prostate), 0.86 (kidney); rotation error ≤ ± 5°	25-30 FPS	Enables accurate intra-operative 3-D overlay for improved spatial orientation
Boonkong	[20]	Rotation, flip, blur, and colour adjustments (Albumentations)	Multi-class F1 > 0.90; ellipse fit error < 4 points	30 FPS	Rapid identification of occluded reproductive anatomy, enhancing safety
Singla	[5]	Curvature flattening, intensity normalization, patch extraction (150 × 150 px)	Accuracy 90%, Sens 90%, Spec 91.7%	≈ 1 s per B-scan	Promising for real-time margin status, potentially reducing re-excisions
Zeye Liu	[21]	Image cropping to 227 × 227 px; normalization	AUC ≥ 0.93, frame-accuracy ≥ 0.85	< 40 ms inference	Matches specialist performance; streamlines workflow where echo expertise is scarce
Kimimasa Sasaki	[18]	30 fps frame extraction; resizing; codec normalization	Accuracy 0.89-0.95; F1 up to 0.97	21 FPS	Provides context-aware phase display and potential automated alerts
Marc Aubreville	[22]	Patch extraction, zero-mean whitening, 2 × rotation	Accuracy 88%, AUC 0.96	50 ms per frame (prototype)	Near-real-time tumor delineation during ablation
Chinedu I. Nwoye	[23]	Frame extraction; bounding-box and weak labels	Triplet AP 18.8%-35%	Not tested live	Establishes public benchmark for detailed OR analytics
Xiaoxuan Zhang	[24]	Metal-artefact correction, scatter in-painting, Gaussian smoothing	SSIM ↑ 15%-22%; lesion Dice ↑ ≤ 25%	< 1 min per volume (parallel pipeline)	Improves soft-tissue contrast and registration accuracy
Schneider	[15]	Stereo reconstruction, iterative closest point alignment, camera calibration	Registration error: manual 10.9 ± 4.2 mm; semi-auto 13.9 ± 4.4 mm	Setup 5-10 min then live overlay	Demonstrates feasibility and surgeon acceptance of AR navigation
Caballas	[25]	Polygon mask annotation	Box AP 92.2%, Mask AP 88.4%	20.6 FPS	Proof-of-concept for real-time palpation cues
Smithmaitrie	[26]	Image resizing; Mosaic augmentation	Landmark mAP 0.85; precision 0.88	Deployed live	95.7% surgeon acceptance of guidance overlay
Török	[27]	500 × 500 px tiles; ensemble prediction fusion	Pixel accuracy 86.2%	Offline	Could aid plane visualization though not yet live
Lin	[28]	Structured-light projection; RGB-HSI fusion	3-D recon 12 FPS; HSI 2 FPS	Yes	Enables dual-modality AR overlay in theatre
Blokker	[29]	Frequency-domain noise filtering; normalization	Accuracy 79%, AUC 0.77, Spec 95.9%	35 ms per 1 k×1 k image	Rapid histology-like feedback during resection
Mojahed	[30]	Down-sampling; z-score normalization; dropout	Accuracy 94%, Sens 96%	0.1 s per B-scan	Could reduce re-excisions via immediate feedback
Tai	[31]	Wavelet-packet decomposition; normalization	AR-guided lung biopsy	Accuracy 97%, RMSE 0.013	900 Hz haptic loop
Jalal	[32]	Resizing to 375 × 300 px; spatiotemporal pooling	Tool mAP 95.6%; phase F1 70.1%	32 FPS	OR decision support and automated video indexing
Hollon	[6]	300 × 300 px sliding window; affine transforms	CNN accuracy 94.6% *vs.* pathologist 93.9%	≤ 150 s per case	Pathologist-level diagnosis without on-site neuropathologist
Mao	[33]	Shift, zoom, rotation, brightness/contrast augmentation	IoU (sella) 67%; 298 FPS (TensorRT)	298 FPS	88.7% surgeons deem output clinically useful
Kitaguchi	[8]	Frame extraction every 1/30 s; manual phase labels	Phase accuracy 91.9%; action accuracy 82-89%	32 FPS	Real-time phase display and coaching
Zeng	[34]	Resizing to 608 × 608 px; patient-wise split; Xavier initialisation	Sens 100%, Spec 99.7%, AUC 0.998	Instant B-scan	Real-time neoplasia triage
Tanzi	[35]	Manual masks; resizing to 416 × 608 px	IoU 0.89 ± 0.08; overlay error ≈ 4 px	8 FPS	Improves biopsy localization in robotic surgery
Podlasek	[36]	Resizing to 224 × 224 px; flips, rotations; class balance	Detection 94%; F1 0.73-0.94	24-57 FPS (GPU	Increases ADR on commodity hardware

				dependent)	
Sato	[37]	Surgeon-annotated masks; data augmentation	Dice 0.58 (AI) *vs.* 0.62 (experts)	30 FPS	Assists nerve preservation, especially for less experienced surgeons
Canalini	[38]	Manual sulci masks; 3-D patch training	mTRE reduced 3.5 → 1.4 mm	Offline (~2 min)	Low-cost alternative to intra-operative MRI
Mekki	[39]	Log transform; affine projection; noise injection	Tip error 1.8 mm; dir. error 2.7°	< 5 s per step-and-shoot	Reduces radiation and improves accuracy
Geldof	[40]	Cropping, normalization, rotation, gamma correction; gradient-weighted Dice loss	Dice 0.84; margin error 0.67 mm; AUC 0.97	Near real-time	Potential to reduce positive margins intra-operatively

IoU: Intersection over Union; FPS: frames per second; 3-D: three-dimensional; OCT: optical coherence tomography; AUC: area under the curve; CLE: confocal laser endomicroscopy; OSCC: oral squamous cell carcinoma; AP: average precision; OR: operating room; CT: computed tomography; SSIM: structural similarity index measure; AR: augmented reality; CNN: convolutional neural network; mAP: mean average precision; RGB: red-green-blue; HSI: hue-saturation-intensity; THG: third - harmonic generation; RMSE: root mean square error; F1: F1-score; SS-OCT: swept-source optical coherence tomography; GPU: graphics processing unit; ADR: adenoma detection rate; RLN: recurrent laryngeal nerve; AI: artificial intelligence; US: ultrasound; mTRE: mean target registration error; MRI: magnetic resonance imaging; CBCT: cone-beam computed tomography.

Table 3. Summary of model performance and reported clinical or workflow outcomes

Clinical task and setting	Performance and runtime	Reported clinical or workflow outcome	References
Breast margin assessment during breast-conserving surgery using optical coherence tomography	Accuracy 90%, sensitivity 90%, specificity 91.7%; one second per cross-sectional scan	Supports intraoperative margin status, with potential to reduce re-excisions	[5]
Breast margin assessment follow-up using optical coherence tomography	Accuracy 94%, sensitivity 96%; 1/10 of a second per cross-sectional scan	Immediate margin feedback, with potential to reduce re-excisions	[30]
Intraoperative brain tumor diagnosis using stimulated Raman histology	Model accuracy comparable to pathologists; no more than 150 s per case	Enables pathologist-level diagnosis without an on-site neuropathologist	[6]
Landmark localization in endoscopic trans-sphenoidal pituitary surgery	Intersection over union 67% for the Sella; 298 frames per second	Deemed clinically useful by most surgeons in user assessment	[33]
Dissection guidance in laparoscopic cholecystectomy	High landmark detection performance; operated in real time	95.7% surgeon acceptance of the guidance overlay	[26]
Phase and instrument recognition in laparoscopic cholecystectomy	Mean average precision for tools 95.6%; F1 score 70.1% for phases; 32 frames per second	Decision support and automated video indexing in the operating room	[32]
Three-dimensional registration for augmented reality overlays in urologic laparoscopy	Intersection over union as high as 0.95; rotation error not exceeding 5°; 25 to 30 frames per second	Improved intraoperative spatial orientation with stable overlays	[19]
Cardiac structure tracking during intraoperative ultrasound	Area under the receiver operating characteristic curve at least 0.93; frame-level accuracy at least 0.85; per-frame processing below 40 ms	Matches specialist performance and streamlines workflow where expertise is scarce	[21]
Tumor segmentation during colorectal surgery with intraoperative ultrasound	Dice coefficient 0.84; margin error 0.67 mm; operated near real time	Potential to reduce positive margins during surgery	[40]
Guidewire navigation in orthopedic trauma under fluoroscopy and cone-beam computed tomography	Tip error 1.8 mm; directional error 2.7°; per step under 5 s	Reduced radiation exposure and improved placement accuracy	[39]
Augmented reality navigation feasibility in laparoscopic liver surgery	Registration error around 11 to 14 mm; live overlay after a short setup period	Feasibility demonstrated with high surgeon acceptance	[15]
Optical biopsy for colorectal tissue using swept-source optical coherence tomography (*ex vivo*)	Sensitivity 100%; specificity 99.7%; instantaneous per scan	Real-time triage of neoplasia during specimen assessment	[34]
Catheter segmentation and overlay in robotic prostatectomy	Intersection over union 0.89 with a standard deviation of 0.08; overlay error 4 pixels; 8 frames per second	Improved biopsy localization during robotic surgery	[35]
Polyp detection during diagnostic colonoscopy	Detection accuracy 94%; F1 score from 0.73% to 0.94%; 24 to 57 frames per second depending on hardware	Real-time detection on common hardware, with potential to increase adenoma detection rate	[36]

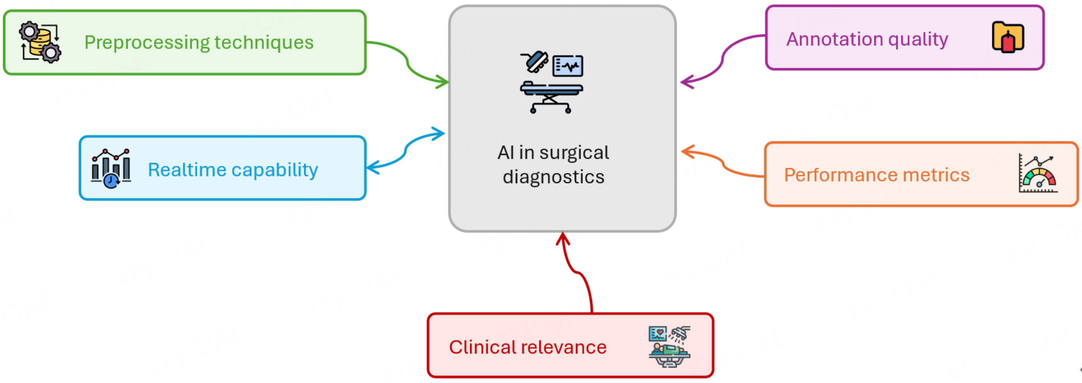

Figure 2. AI in surgical diagnostics: techniques and applications. AI: Artificial intelligence. The figure is available at this link: https://simadeduso-my.sharepoint.com/:p:/g/personal/momustafahmed_simad_edu_so/IQC6ZWrC08ocR4mCaevyYq6bAY8GGfRtkb42bPbiwSHFmtM?e=H9QjU2.

Studies and meta-analyses of breast-conserving surgery indicate that intraoperative OCT can inform margin status and may reduce re-excisions when integrated into the live workflow, aligning with our interpretation of clinical utility rather than offline accuracy alone[51,52]. In hepatic surgery, recent clinical experiences and narrative reviews have reported high surgeon acceptance of AR navigation when registration quality and display ergonomics are adequate, echoing the acceptance signals in our model[49,50]. In comparative studies of surgical workflow and skill analysis ML algorithms, it has been found that benchmark standardization and multicentric validation are crucial, as performance can vary significantly depending on dataset design and institutional variability[53,54]. Cardiac imaging reviews similarly describe automated structure tracking and measurements that approach expert performance and can alleviate bottlenecks where echocardiography expertise is scarce, corresponding to the workflow benefits we noted for the intraoperative echocardiography[41,55].

However, the generalizability of these findings remains a limiting factor. Multicenter studies consistently show that models trained on single-center surgical videos degrade when tested elsewhere and that technique and camera heterogeneity drive performance variability. Comparative analyses and multicenter datasets of laparoscopic procedures quantify this gap and recommend explicit external validation and domain-robust training over reliance on pretraining alone. Our interpretation that per-site excellence does not guarantee portability is consistent with these results and broader reviews cataloging dataset biases and split pathologies in common benchmarks[56-58]. Human factors significantly influence the adoption of technology, comparable to the impact of accuracy. Reviews of automation bias and clinician trust underscore the importance of interface design, transparency regarding uncertainty, and the establishment of clear guidelines for enhancing team performance and safety, particularly under time constraints. These findings support our caution that explanation features should elucidate system limitations and that escalation and fallback mechanisms must be integrated into OR assistants rather than being assumed[59-61]. Equity considerations are also pertinent to the use of surgical AI. The United Kingdom independent review on equity in medical devices highlights the need to address differential performance across subgroups and the sociotechnical factors affecting access and outcomes, reinforcing the importance of subgroup reporting and post-deployment monitoring in surgical contexts[62].

Regulatory pathways are beginning to align with these considerations. In the United States, AI-enabled surgical software is generally regulated as software for medical devices under existing device frameworks. The FDA's final guidance on Predetermined Change Control Plans outlines expectations for pre-specifying permissible model updates, verification and validation plans, and real-world monitoring, which is particularly relevant to video systems that will undergo iterations post-launch. The Good Machine Learning Practice (GMLP) principles, co-published by the FDA, Health Canada, and the Medicines and Healthcare products Regulatory Agency (MHRA), along with the International Medical Device Regulators Forum's (IMDRF) 2025 final GMLP document, emphasize lifecycle quality, data governance, and transparency. In the European Union, the new Medical Device Coordination Group (MDCG) Document 2025-6 Frequently Ask Questions (FAQ) clarifies how the high-risk obligations of AI Act will be assessed alongside the Medical Device Regulation (MDR) or *In Vitro* Diagnostic Regulation (IVDR) during the conformity assessment, necessitating that teams plan technical documentation and post-market surveillance to satisfy both regulatory regimes[59-62].

This study has several limitations. The restriction to English-language sources may introduce language and publication bias by excluding pertinent non-English evidence, potentially skewing our conclusions. Additionally, the inclusion was limited to complete calendar years from January 1, 2017, to December 31, 2024, to ensure a consistent sampling frame. Although several studies from 2025 were discussed to contextualize current trajectories, they were not eligible for inclusion in the evidence synthesis and were cited narratively. We adopted a pragmatic operationalization in which real-time was most interpreted as at least 20 FPS or at most 1 s per frame; studies that self-described as real-time outside these bounds were retained and explicitly flagged. Frames-per-second and per-frame latencies were extracted when available and coded as not reported when absent. This variability introduces measurement noise that can affect cross-study comparability, underscoring the value of reporting frameworks for early clinical AI evaluations that emphasize the explicit documentation of technical performance and human-AI interaction during live use.

In conclusion, the evidence suggests that contemporary DL pipelines can meet OR constraints and are beginning to offer clinically significant support. However, their widespread applicability and accuracy are contingent on rigorous multicenter validation, explicit management of domain shifts, meticulous human factors engineering, and adherence to evolving regulatory frameworks that prioritize lifecycle quality. A revised version of this review should expand the search to include non-English sources and gray literature and extend the inclusion criteria to encompass the most recent calendar year, as several studies from 2025 have already expanded the evidence base for label-free pathology, AR navigation, and real-time workflow analysis.

DECLARATIONS

Acknowledgements

The icons used in the Graphical Abstract and Figure 2 were obtained from Flaticon (https://www.flaticon.com/).

Author contributions

Conceptualized and designed the review: Ahmed MM, Kasimieh O
Conducted the literature review, quality appraisal, and data extraction: Hassan MM, Ali I, Othman ZK
Led the methodology and drafted the manuscript: Ahmed MM
Supported data verification and synthesis: Maulion PM
Provided critical revisions for intellectual content: Okesanya OJ, Branda F, Kasimieh O, Babalola AE, Ukoaka BM
Supervised this work: Lucero-Prisno III DE
All authors read and approved the final manuscript.

Availability of data and materials
Not applicable.

Financial support and sponsorship
None.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

REFERENCES

1. Meara JG, Leather AJ, Hagander L, et al. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *Lancet.* 2015;386:569-624. DOI PubMed
2. Nepogodiev D, Martin J, Biccard B, Makupe A, Bhangu A; National Institute for Health Research Global Health Research Unit on Global Surgery. Global burden of postoperative death. *Lancet.* 2019;393:401. DOI PubMed
3. Keil H, Beisemann N, Swartman B, et al. Intraoperative revision rates due to three-dimensional imaging in orthopedic trauma surgery: results of a case series of 4721 patients. *Eur J Trauma Emerg Surg.* 2023;49:373-81. DOI PubMed PMC
4. Wang X, Yang J, Zhou B, Tang L, Liang Y. Integrating mixed reality, augmented reality, and artificial intelligence in complex liver surgeries: enhancing precision, safety, and outcomes. *ILIVER.* 2025;4:100167. DOI PubMed PMC
5. Singla N, Dubey K, Srivastava V. Automated assessment of breast cancer margin in optical coherence tomography images via pretrained convolutional neural network. *J Biophotonics.* 2019;12:e201800255. DOI PubMed
6. Hollon TC, Pandian B, Adapa AR, et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat Med.* 2020;26:52-8. DOI PubMed PMC
7. Li Y, Zhao Z, Li R, Li F. Deep learning for surgical workflow analysis: a survey of progresses, limitations, and trends. *Artif Intell Rev.* 2024;57:10929. DOI
8. Kitaguchi D, Takeshita N, Matsuzaki H, et al. Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. *Surg Endosc.* 2020;34:4924-31. DOI PubMed
9. Aboy M, Minssen T, Vayena E. Navigating the EU AI Act: implications for regulated digital medical products. *NPJ Digit Med.* 2024;7:237. DOI PubMed PMC
10. Magalhães R, Oliveira A, Terroso D, et al. Mixed reality in the operating room: a systematic review. *J Med Syst.* 2024;48:76. DOI PubMed PMC
11. Nair M, Svedberg P, Larsson I, Nygren JM. A comprehensive overview of barriers and strategies for AI implementation in healthcare: mixed-method design. *PLoS One.* 2024;19:e0305949. DOI PubMed PMC
12. Ennab M, Mcheick H. Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions. *Front Robot AI.* 2024;11:1444763. DOI PubMed PMC
13. Salahuddin Z, Woodruff HC, Chatterjee A, Lambin P. Transparency of deep neural networks for medical image analysis: a review of interpretability methods. *Comput Biol Med.* 2022;140:105111. DOI PubMed
14. U. S. Food & Drug Administration. FDA Roundup: September 17, 2024. Available from: https://www.fda.gov/news-events/press-announcements/fda-roundup-september-17-2024?utm_source=chatgpt.com [accessed 16 December 2025].
15. Schneider C, Thompson S, Totz J, et al. Comparison of manual and semi-automatic registration in augmented reality image-guided liver surgery: a clinical feasibility study. *Surg Endosc.* 2020;34:4702-11. DOI PubMed PMC
16. Koetzier LR, Wu J, Mastrodicasa D, et al. Generating synthetic data for medical imaging. *Radiology.* 2024;312:e232471. DOI PubMed PMC
17. PRISMA 2020. PRISMA statement. Available from: https://www.prisma-statement.org/ [accessed 16 December 2025].
18. Sasaki K, Ito M, Kobayashi S, et al. Automated surgical workflow identification by artificial intelligence in laparoscopic hepatectomy: experimental research. *Int J Surg.* 2022;105:106856. DOI PubMed
19. Padovan E, Marullo G, Tanzi L, et al. A deep learning framework for real-time 3D model registration in robot-assisted laparoscopic

surgery. *Int J Med Robot.* 2022;18:e2387. DOI PubMed PMC
20. Boonkong A, Khampitak K, Kaewfoongrungsi P, Namkhun S, Hormdee D. Applying deep learning for occluded uterus and fallopian tube detection for laparoscopic tubal sterilization. *IEEE Access.* 2024;12:183182-94. DOI
21. Liu Z, Li W, Li H, et al. Automated deep neural network-based identification, localization, and tracking of cardiac structures for ultrasound-guided interventional surgery. *J Thorac Dis.* 2023;15:2129-40. DOI PubMed PMC
22. Aubreville M, Knipfer C, Oetter N, et al. Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning. *Sci Rep.* 2017;7:11979. DOI PubMed PMC
23. Nwoye CI, Yu T, Sharma S, et al. CholecTriplet2022: Show me a tool and tell me the triplet - an endoscopic vision challenge for surgical action triplet detection. *Med Image Anal.* 2023;89:102888. DOI PubMed
24. Zhang X, Sisniega A, Zbijewski WB, et al. Combining physics-based models with deep learning image synthesis and uncertainty in intraoperative cone-beam CT of the brain. *Med Phys.* 2023;50:2607-24. DOI PubMed PMC
25. Caballas KG, Bolingot HJM, Libatique NJC, Tangonan GL. Development of a visual guidance system for laparoscopic surgical palpation using computer vision. 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES); 2021 Mar 1-3; Langkawi Island, Malaysia. New York: IEEE; 2021. pp. 88-93. DOI
26. Smithmaitrie P, Khaonualsri M, Sae-Lim W, Wangkulangkul P, Jearanai S, Cheewatanakornkul S. Development of deep learning framework for anatomical landmark detection and guided dissection line during laparoscopic cholecystectomy. *Heliyon.* 2024;10:e25210. DOI PubMed PMC
27. Török P, Harangi B. Digital image analysis with fully connected convolutional neural network to facilitate hysteroscopic fibroid resection. *Gynecol Obstet Invest.* 2018;83:615-9. DOI PubMed
28. Lin J, Clancy NT, Qi J, et al. Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks. *Med Image Anal.* 2018;48:162-76. DOI PubMed
29. Blokker M, Hamer PCW, Wesseling P, Groot ML, Veta M. Fast intraoperative histology-based diagnosis of gliomas with third harmonic generation microscopy and deep learning. *Sci Rep.* 2022;12:11334. DOI PubMed PMC
30. Mojahed D, Ha RS, Chang P, et al. Fully automated postlumpectomy breast margin assessment utilizing convolutional neural network based optical coherence tomography image classification method. *Acad Radiol.* 2020;27:e81-6. DOI PubMed PMC
31. Tai Y, Qian K, Huang X, Zhang J, Jan MA, Yu Z. Intelligent intraoperative haptic-AR navigation for COVID-19 lung biopsy using deep hybrid model. *IEEE Trans Industr Inform.* 2021;17:6519-27. DOI PubMed PMC
32. Jalal NA, Alshirbaji TA, Docherty PD, et al. Laparoscopic video analysis using temporal, attention, and multi-feature fusion based-approaches. *Sensors.* 2023;23:1958. DOI PubMed PMC
33. Mao Z, Das A, Islam M, et al. PitSurgRT: real-time localization of critical anatomical structures in endoscopic pituitary surgery. *Int J Comput Assist Radiol Surg.* 2024;19:1053-60. DOI PubMed PMC
34. Zeng Y, Xu S, Chapman WC Jr, et al. Real-time colorectal cancer diagnosis using PR-OCT with deep learning. *Theranostics.* 2020;10:2587-96. DOI PubMed PMC
35. Tanzi L, Piazzolla P, Porpiglia F, Vezzetti E. Real-time deep learning semantic segmentation during intra-operative surgery for 3D augmented reality assistance. *Int J Comput Assist Radiol Surg.* 2021;16:1435-45. DOI PubMed PMC
36. Podlasek J, Heesch M, Podlasek R, Kilisiński W, Filip R. Real-time deep learning-based colorectal polyp localization on clinical video footage achievable with a wide array of hardware configurations. *Endosc Int Open.* 2021;9:E741-8. DOI PubMed PMC
37. Sato K, Fujita T, Matsuzaki H, et al. Real-time detection of the recurrent laryngeal nerve in thoracoscopic esophagectomy using artificial intelligence. *Surg Endosc.* 2022;36:5531-9. DOI PubMed
38. Canalini L, Klein J, Miller D, Kikinis R. Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery. *Int J Comput Assist Radiol Surg.* 2019;14:1697-713. DOI PubMed PMC
39. Mekki L, Sheth NM, Vijayan RC, et al. Surgical navigation for guidewire placement from intraoperative fluoroscopy in orthopedic surgery. *Phys Med Biol.* 2023;68:215001. DOI PubMed
40. Geldof F, Pruijssers CWA, Jong LS, Veluponnar D, Ruers TJM, Dashtbozorg B. Tumor segmentation in colorectal ultrasound images using an ensemble transfer learning model: towards intra-operative margin assessment. *Diagnostics.* 2023;13:3595. DOI PubMed PMC
41. Kenig N, Monton Echeverria J, Muntaner Vives A. Artificial intelligence in surgery: a systematic review of use and validation. *J Clin Med.* 2024;13:7108. DOI PubMed PMC
42. Yang HY, Hong SS, Yoon J, et al. Deep learning-based surgical phase recognition in laparoscopic cholecystectomy. *Ann Hepatobiliary Pancreat Surg.* 2024;28:466-73. DOI PubMed PMC
43. Kostiuchik G, Sharan L, Mayer B, Wolf I, Preim B, Engelhardt S. Surgical phase and instrument recognition: how to identify appropriate dataset splits. *Int J Comput Assist Radiol Surg.* 2024;19:699-711. DOI PubMed PMC
44. Goyal A, Mendoza M, Munoz AE, et al. Artificial intelligence for real-time surgical phase recognition in minimal invasive inguinal hernia repair: a systematic review on behalf of TROGSS - the robotic global surgical society. *Art Int Surg.* 2025;5:450-64. DOI
45. Bellini V, Russo M, Domenichetti T, Panizzi M, Allai S, Bignami EG. Artificial intelligence in operating room management. *J Med Syst.* 2024;48:19. DOI PubMed PMC
46. Hollon TC, Orringer DA. An automated tissue-to-diagnosis pipeline using intraoperative stimulated Raman histology and deep learning. *Mol Cell Oncol.* 2020;7:1736742. DOI PubMed PMC
47. Duan Y, Guo D, Zhang X, et al. Diagnostic accuracy of optical coherence tomography for margin assessment in breast-conserving

surgery: a systematic review and meta-analysis. *Photodiagnosis Photodyn Ther.* 2023;43:103718. DOI PubMed
48. Fan S, Zhang H, Meng Z, Li A, Luo Y, Liu Y. Comparing the diagnostic efficacy of optical coherence tomography and frozen section for margin assessment in breast-conserving surgery: a meta-analysis. *J Clin Pathol.* 2024;77:517-27. DOI PubMed
49. Ramalhinho J, Bulathsinhala S, Gurusamy K, Davidson BR, Clarkson MJ. Assessing augmented reality displays in laparoscopic liver surgery - a clinical experience. *Surg Endosc.* 2025;39:5863-71. DOI PubMed PMC
50. Roman J, Sengul I, Němec M, et al. Augmented and mixed reality in liver surgery: a comprehensive narrative review of novel clinical implications on cohort studies. *Rev Assoc Med Bras.* 2025;71:e20250315. DOI PubMed PMC
51. Raissi-dehkordi N, Raissi-dehkordi N, Xu B. Contemporary applications of artificial intelligence and machine learning in echocardiography. *npj Cardiovasc Health.* 2025;2:64. DOI
52. Hirata Y, Kusunose K. AI in echocardiography: state-of-the-art automated measurement techniques and clinical applications. *JMA J.* 2025;8:141-50. DOI PubMed PMC
53. Wagner M, Müller-Stich BP, Kisilenko A, et al. Comparative validation of machine learning algorithms for surgical workflow and skill analysis with the HeiChole benchmark. *Med Image Anal.* 2023;86:102770. DOI PubMed
54. Lavanchy JL, Ramesh S, Dall'Alba D, et al. Challenges in multi-centric generalization: phase and step recognition in Roux-en-Y gastric bypass surgery. *Int J Comput Assist Radiol Surg.* 2024;19:2249-57. DOI PubMed PMC
55. Abdelwanis M, Alarafati HK, Tammam MMS, Simsekler MCE. Exploring the risks of automation bias in healthcare artificial intelligence applications: a Bowtie analysis. *Journal of Safety Science and Resilience.* 2024;5:460-9. DOI
56. Tun HM, Rahman HA, Naing L, Malik OA. Trust in artificial intelligence-based clinical decision support systems among health care workers: systematic review. *J Med Internet Res.* 2025;27:e69678. DOI PubMed PMC
57. Sadeghi Z, Alizadehsani R, Cifci MA, et al. A review of Explainable Artificial Intelligence in healthcare. *Comput Electr Eng.* 2024;118:109370. DOI
58. GOV.UK. Equity in medical devices: independent review - final report. Available from: https://www.gov.uk/government/publications/equity-in-medical-devices-independent-review-final-report [accessed 16 December 2025].
59. European Commission. MDCG 2025-6 - FAQ on interplay between the medical devices regulation & *in vitro* diagnostic medical devices regulation and the artificial intelligence act (June 2025). Available from: https://health.ec.europa.eu/latest-updates/mdcg-2025-6-faq-interplay-between-medical-devices-regulation-vitro-diagnostic-medical-devices-2025-06-19_en [accessed 16 December 2025].
60. US Food and Drug Administration. Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence [accessed 16 December 2025].
61. US Food and Drug Administration. Good machine learning practice for medical device development: guiding principles. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles [accessed 16 December 2025].
62. International Medical Device Regulations Forum. Good machine learning practice for medical device development: guiding principles. AUTHORING GROUP: Artificial Intelligence/Machine Learning-enabled Working Group, 2025. Available from: https://www.imdrf.org/sites/default/files/2025-01/IMDRF_AIML%20WG_GMLP_N88%20Final_0.pdf [accessed 16 December 2025].