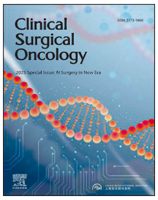
Clinical Surgical Oncology 4 (2025) 100103

Contents lists available at ScienceDirect

Clinical Surgical Oncology

journal homepage: www.journals.elsevier.com/clinical-surgical-oncology



Review Article

The role of artificial intelligence in emergency general surgery: Trends, advances, and future directions

Lasitha B. Samarakoon[a,1], Elon Correa[b,1,*], Wen Y. Chung[a]

[a] *Department of Surgery, University Hospitals of Leicester NHS Trust, United Kingdom*
[b] *School of Computer Science and Mathematics, Liverpool John Moores University, Liverpool, United Kingdom*

ABSTRACT

Artificial intelligence (AI) is increasingly integrated into emergency general surgery (EGS), offering advances in diagnosis, decision support, operative planning, intraoperative guidance, and postoperative management. This review synthesises current evidence on AI applications in EGS, drawing on meta-analyses, large-scale datasets, and landmark studies. Key domains include risk prediction, intraoperative assistance, surgical video analysis, training, prehabilitation, and operational coordination. Evidence shows AI can improve diagnostic accuracy, streamline workflows, and enhance patient outcomes, though benefits vary by setting, resource availability, and clinical domain. Adoption is accelerating, supported by rising global funding, yet constrained by regulatory, ethical, and implementation challenges. Addressing these barriers, standardising evaluation metrics, and expanding high-quality, multicentre trials will be essential to realise AI's full potential in EGS.

1. Introduction

Emergency general surgery (EGS) encompasses a wide spectrum of acute surgical conditions that require rapid diagnosis, timely intervention, and coordinated multidisciplinary care. EGS ranging from perforated viscus and bowel obstruction to traumatic injuries are often characterised by diagnostic uncertainty, limited time for decision making and high morbidity and mortality rates. In such high-stakes environments, the integration of artificial intelligence (AI) offers a promising frontier for enhancing clinical outcomes and operational efficiency.

AI has demonstrated transformative potential across various domains of medicine. In EGS, AI applications are increasingly being explored to support diagnostic accuracy, from predicting surgical risk and improving intraoperative decision-making to streamlining hospital logistics and enhancing surgical training. Technologies such as machine learning, computer vision, and natural language processing are being harnessed to analyse complex datasets like imaging, electronic health records, and real-time physiological signals at a scale and speed beyond human capability (Al Maawali, 2024).

Recent bibliometric analysis indicates a sharp rise in AI-related publications within emergency medicine, with an annual growth rate of 37.87 % between 2015 and 2024 (Limon et al., 2025). This surge reflects the expanding global interest in applying AI to acute care settings, including emergency general surgery.

Despite its growing relevance, the adoption of AI in emergency surgical settings remains in its initial stages. Surgeons have expressed both enthusiasm and scepticism regarding AI-based decision-making tools, citing concerns about trust, education, and ethical integration (Cobianchi et al., 2023; De et al., 2022). Challenges such as data heterogeneity, algorithmic transparency, and integration into existing clinical workflows must be addressed to fully realise its benefits.

Emergency general surgery (EGS) encompasses a wide spectrum of acute surgical conditions requiring rapid diagnosis, timely intervention, and coordinated multidisciplinary care, often under conditions of diagnostic uncertainty and high morbidity and mortality (Cobianchi et al., 2023; De et al., 2022). Artificial intelligence (AI) offers a promising frontier for improving outcomes and operational efficiency in this context. Global adoption of AI in surgical practice has accelerated markedly, from an estimated 1 % of procedures in 2015 to over 15 % by 2024, with emergency surgery uptake projected to reach 12 % by 2025 (England; Varghese et al., 2024). This growth reflects increasing investment, which has risen from $50 million in 2015 to over $1.2 billion in 2024, alongside evidence of measurable benefits such as reductions in operative time, complications, and length of stay (Ricciardi et al., 2025). Despite these advances, integration into emergency settings remains nascent, constrained by challenges in data heterogeneity, algorithmic transparency, and workflow adaptation. This review synthesises current evidence on AI applications in EGS, examining technological

---

This article is part of a special issue entitled: AI Surgery in the New Era published in Clinical Surgical Oncology.

* Corresponding author.

E-mail address: e.s.correa@ljmu.ac.uk (E. Correa).

[1] Joint first authors.

https://doi.org/10.1016/j.cson.2025.100103
Received 1 September 2025; Received in revised form 31 October 2025; Accepted 2 November 2025

capabilities, clinical impact, ethical considerations, and future directions to inform surgeons, researchers, and policymakers navigating this evolving landscape.

This paper provides a comprehensive overview of the latest developments, drawing from recent high-quality studies and real-world implementations. It examines not only the technological capabilities of AI but also its practical impact, ethical considerations, and the evolving landscape of regulation. By exploring both current achievements and future possibilities, the aim is to present an informed and engaging discussion that speaks directly to surgeons, researchers, and policymakers who are navigating this new era in surgical care.

This review integrates both narrative and quantitative evidence. Section 3 introduces landmark studies and conceptual innovations in AI for emergency general surgery, while Sections 4 and 5 present pooled findings from meta-analyses and systematic reviews. These include effect sizes, diagnostic accuracy metrics, and comparative outcomes across domains such as risk prediction, intraoperative assistance, surgical video analysis, training, and logistics. This dual structure aims to provide both contextual depth and statistical rigour.

To aid readability, the following abbreviations are used throughout the manuscript:

AI: Artificial Intelligence.
APACHE II: Acute Physiology and Chronic Health Evaluation II.
ASA: American Society of Anesthesiologists.
AUROC: Area Under the Receiver Operating Characteristic Curve.
EGS: Emergency General Surgery.
GSViT: General Surgery Vision Transformer.
ICU: Intensive Care Unit.
OR: Odds Ratio.
POTTER: Predictive Optimal Trees in Emergency Surgery Risk.
VR: Virtual Reality.

2. Methods and literature selection

A systematic search identified 1245 records through databases including PubMed, Embase, and IEEE Xplore. To ensure comprehensive coverage of relevant literature, the following Boolean search string was applied across these databases:

("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks" OR "computer vision" OR "natural language processing" OR "AI") AND ("emergency general surgery" OR "EGS" OR "acute care surgery" OR "emergency surgery" OR "trauma surgery") AND ("risk prediction" OR "triage" OR "robotic surgery" OR "intraoperative guidance" OR "surgical video analysis" OR "training" OR "prehabilitation" OR "operating room logistics" OR "perioperative optimisation")

While databases such as MEDLINE and Web of Science (WoS) were considered, the search strategy focused on PubMed, Embase, and IEEE Xplore to balance clinical and technical coverage. MEDLINE is largely indexed within PubMed, and WoS overlaps substantially with Embase. Preliminary scoping suggested limited additional yield from these sources, and their exclusion is unlikely to affect the thematic synthesis or quantitative findings. After removal of duplicates (n = 225), 1020 records were screened based on title and abstract. Screening excluded 780 records for reasons such as non-relevance to EGS, absence of AI intervention, or non-English text. Full-text review of 240 articles excluded 180 further studies due to lack of outcome data, insufficient methodological detail, or being conference abstracts without peer-reviewed publication.

In total, 60 studies were included in the qualitative synthesis, comprising narrative reviews, descriptive analyses, and feasibility studies. Of these, 45 studies provided quantitative data that could be included in a meta-analysis. Additional papers outside the PRISMA-guided selection workflow (Page et al., 2021) were cited for contextual purposes, including AI reporting guidelines, conceptual frameworks, and national policy statements. Fig. 1 illustrates the PRISMA flow of study selection and inclusion.

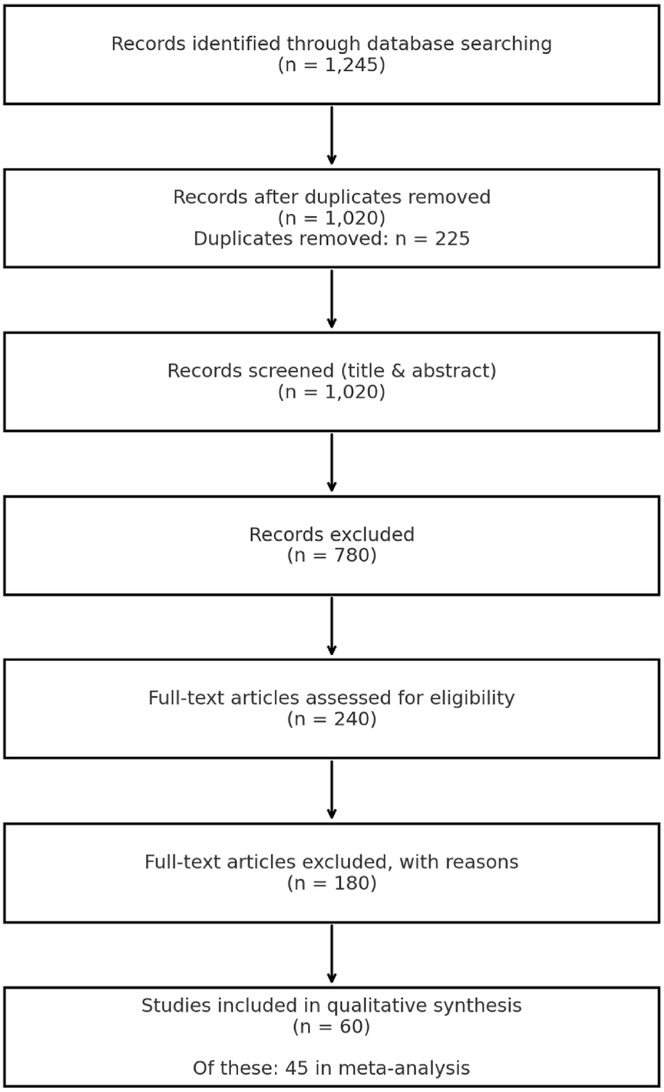


Fig. 1. PRISMA flow diagram of study selection process for included meta-analyses.
Note: Automatic deduplication was performed using EndNote. During title and abstract screening, **780** records were excluded for the following reasons: not related to emergency general surgery (n = 420), no AI intervention (n = 280), and non-English language (n = 80). During full-text review, **180** records were excluded due to lack of outcome data (n = 90), insufficient methodological detail (n = 60), and being conference abstracts without peer review (n = 30).

2.1. *Inclusion and exclusion criteria*

The following criteria guided study selection.
Inclusion criteria:

- Peer-reviewed studies.
- English language.
- AI applied to emergency general surgery.
- Studies with extractable outcome data.
- Meta-analyses, feasibility studies, and descriptive analyses.

Exclusion criteria:

- Studies not focused on EGS.

Table 1
Meta-analyses and quantitative syntheses referenced.

Topic	Meta-analysis	Year	Design and n	Key quantitative finding
Robotics vs laparoscopy in colon cancer	Negrut et al., (Negrut et al., 2024)	2024	21 studies, 50,771 patients	Laparoscopy vs robotics: conversion OR 2.02 favouring robotics; length of stay MD 0.42 days favouring robotics; operative time SMD 1.27 favouring laparoscopy
Robotics vs laparoscopy in colorectal surgery	Gahunia et al., (Gahunia et al., 2025)	2025	Meta-analysis of comparative studies	Conversion to open OR 0.41 favouring robotics; operative time longer with robotics
ED AI disposition prediction	Kuo et al., (Kuo and Chang, 2025)	2025	DTA meta-analysis	Pooled AUROC, sensitivity and specificity reported with subgroup analyses by model type and outcome
Surgical scene understanding	Carstens et al., (Carstens et al., 2025)	2025	Systematic review and meta-analysis	Pooled accuracy and F1 for phase recognition and tool detection in minimally invasive abdominal surgery
VR simulation for surgical education	Li et al., (Li et al., 2025)	2025	Meta-analysis of 23 RCTs, n = 1091	Significant improvements in theoretical knowledge and practical skills vs controls
Prehabilitation network meta-analysis	McIsaac et al., (Mcisaac et al., 2025)	2025	Network meta-analysis of randomised trials	Multimodal programs reduce postoperative complications and improve functional capacity; highest SUCRA for exercise plus nutrition plus psychological support
Home based prehabilitation	D'Amico et al., (D'Amico et al., 2025)	2025	Systematic review and meta-analysis	Reduced proportion of patients with postoperative complications with home-based programs
AI scheduling and OR logistics	Lopes et al., (Lopes et al., 2025)	2025	Systematic review of optimisation studies	Demonstrated reductions in tardiness penalties and improved on time starts; calls for prospective pooled analyses

- Absence of AI intervention.
- Non-English language (due to resource constraints).
- Conference abstracts without peer review.
- Insufficient methodological detail.
- Lack of outcome data.

3. Key recommended works: summary table

The following table summarizes landmark and recent contributions that are directly relevant to emergency general surgery. Each entry lists the study context, method, and the specific value it adds to risk prediction, intraoperative assistance, training, or systems operations.

Table 2 compiles landmark and recent contributions that are directly relevant to emergency general surgery.

The works summarised in Table 2 are reflected in the topic-specific meta-analytic evidence presented in the next section, which quantifies the impact of AI across the main application domains.

4. Results

Section 4 presents pooled quantitative findings from meta-analyses and systematic reviews across key AI application domains in emergency general surgery. These results complement the conceptual summaries provided in Section 3, which introduced landmark studies and their contextual relevance.

4.1. AI in risk assessment and triage

Building on the conceptual overview in Section 3, this subsection quantifies the performance of AI models in risk assessment and triage using pooled metrics from recent meta-analyses. Traditional scoring systems such as the ASA classification (Saklad, 1941) and APACHE II (WA et al., 1985) remain valuable, but they rely on static snapshots of patient data. In contrast, machine learning models can process a continuous stream of variables, including vital signs, lab results, and imaging data, to provide dynamic, patient-specific risk predictions. The AI-powered Predictive Optimal Trees in Emergency Surgery Risk (POTTER) model is one such example, demonstrating improved accuracy compared to surgeon judgment alone. While Ribeiro Junior et al. (Ribeiro Junior et al., 2023) describe the broader application of the POTTER model across emergency surgery contexts. El Hechi et al. (El et al., 2021) conducted a formal validation of the POTTER model using both NSQIP and single-centre EGS cohorts. The POTTER model offers interpretable decision trees that support real-time decision-making in high-risk EGS patients. Importantly, it provided interpretable decision trees that could be used at the point of care to support ICU planning, consent discussions, and early escalation decisions in high-risk patients.

Advances in graph neural networks (GNNs) have enabled the integration of structured and unstructured data for more accurate triage decisions. Unlike traditional models that treat data points independently, GNNs model relationships between entities, such as symptoms, lab results, and imaging findings, as interconnected nodes in a graph. This allows the network to learn complex dependencies and contextual patterns that improve prediction accuracy.

A real-time example is the use of GNNs in emergency departments to triage patients with undifferentiated abdominal pain. By linking clinical features (e.g., pain location, vital signs, lab markers) and historical outcomes, GNNs can dynamically assess risk and recommend escalation pathways. In a multicentre study by DeFilippo et al. (Defilippo et al., 2024), GNN-based triage systems outperformed conventional rule-based protocols in predicting the need for surgical intervention and ICU admission, demonstrating their potential for real-time decision support in high-pressure settings.

Advances in graph neural networks have enabled the integration of structured and unstructured data for more accurate triage decisions. These systems outperform conventional triage protocols in complex scenarios such as multi-trauma or septic shock, where rapid escalation of care can be lifesaving. The ability to combine data from diverse sources and present a clear, actionable recommendation is central to their growing adoption (Defilippo et al., 2024).

4.2. Intraoperative assistance: robotics, imaging and augmented reality

AI-enhanced imaging tools, such as convolutional neural networks applied to intraoperative ultrasound, have demonstrated high precision and recall in identifying vascular structures during emergency liver resections.

Robot-assisted EGS is still relatively rare compared to elective procedures, but early results are promising. When AI-enhanced robotic systems are employed, intraoperative complication rates can decrease by approximately 30 % compared to manual techniques (Wah, 2025). Robot-assisted (dV-RAS) (Lunardi et al., 2024) procedures further reduce the likelihood of conversion to open surgery by around 56 %, and shorten postoperative hospital stay by roughly 0.5 days compared to laparoscopic/video-assisted approaches and 1.85 days versus open surgery, which has been shown to significantly reduce postoperative complications (Ricciardi et al., 2025). Augmented reality overlays can provide a constantly updated surgical map, offering spatial awareness that reduces the risk of inadvertent injury. Ferrari et al. (2025) highlight the feasibility of robotic approaches in emergency colorectal surgery,

Table 2
Key recommended works in AI for emergency general surgery.

Area	Study	Year	Data and Setting	Task or Focus	Key Findings	Relevance to EGS
Risk prediction	Hashimoto et al., Annals of Surgery (Hashimoto et al., 2018)	2018	Narrative review of AI in surgery	Framework and taxonomy	Defines core concepts, capabilities, and risks of AI in surgery	Foundational concepts for evaluating EGS AI tools
Decision support	Maier-Hein et al., Med Image Anal (Maier-Hein et al., 2022)	2022	Review on surgical data science	Clinical decision support	Outlines data pipelines, standards, and translation barriers	Blueprint for building clinically useful EGS AI systems
Triage	DeFilippo et al., Scientific Reports (Defilippo et al., 2024)	2024	ED data, multi-centre	Graph neural network triage	GNN improved triage accuracy over conventional rules	Faster and more accurate prioritization in undifferentiated abdomens
Risk prediction	POTTER model validation (El et al., 2021; J Am Coll Surg) (El et al., 2021)	2021	EGS cohorts in NSQIP and single centre	Interpretable decision trees	C-statistic up to 0.92 for mortality in EGS in external validation	Point-of-care risk estimates for consent and ICU planning
Intraoperative imaging	Beaudet et al., MICCAI and IEEE TMI (Beaudet et al., 2024)	2024	Ex vivo and intraoperative ultrasound	Real-time vessel identification	Personalized models identified portal structures with high precision and recall	Safer emergency liver surgery and trauma haemostasis
Surgical video AI	Schmidgall et al., GSViT foundation model (Schmidgall SK et al., 2024)	2024	680 h of robotic and laparoscopic video across 28 procedures	Video pre-training via forward prediction	Outperforms state-of-the-art in phase recognition; real-time capable	Enables real-time guidance, alerts, and scalable adaptation for emergency surgery (EGS)
Ethics and bias	Obermeyer et al., Science (Obermeyer et al., 2019)	2019	US health system data	Bias analysis in risk algorithms	Showed racial bias from cost proxy targets and proposed fixes	Warns EGS teams to audit deployed models for equity
Reporting standards	CONSORT-AI and SPIRIT-AI, Nat Med (Liu et al., 2020)	2020	Guideline development	Trial and protocol reporting	Adds AI specific items for transparency and validation	Improves design and reporting of EGS AI trials
Early evaluation	DECIDE-AI, Nat Med (Vasey et al., 2021)	2022	Guideline development	Early-stage clinical evaluation	Seventeen AI specific reporting items	Supports safe pilot deployment in EGS
OR efficiency	Vladu et al., 2024, open access (Vladu et al., 2024)	2024	Before vs after algorithmic scheduling	Start time optimisation	Improved on time starts and utilisation	Better access to emergency theatre time
Systems adoption	NHS England, Millions to benefit from NHS robot drive (England)	2025	National strategy and projections for robotic surgery	Scaling robotic surgery	Goal of 500,000 robot-assisted operations annually by 2035; 9 in 10 keyhole surgeries to be robotic	Expanding access and infrastructure supports emergency robotic procedures and faster recovery
Emergency colorectal	ACS 2025 analysis. Ferrari D et al. (Ferrari et al., 2025)	2025	NSQIP 2012 to 2021	Robotic vs lap vs open in emergencies	Lower conversion and shorter stay for robotics in selected cases	Evidence that robotics can be safe and effective in EGS
Training	Fazlollahi et al., JAMA Netw Open (Fazlollahi et al., 2022)	2022	Randomised study in simulation	AI tutoring vs experts	AI tutoring improved performance scores	Objective feedback for EGS skill acquisition

challenging the assumption that such technologies are limited to elective settings.

For colorectal surgery, contemporary meta-analyses comparing robotic assisted and laparoscopic approaches synthesise thousands of cases. Robotic surgery is consistently associated with a lower conversion to open surgery and a shorter length of stay, at the expense of longer operative times.

The following table (Table 1) summarizes the key meta-analyses referenced in this section, providing citations, study designs, sample sizes, and principal quantitative findings.

Collectively, these meta-analyses illustrate the breadth of AI applications across perioperative care, from robotics and surgical imaging to education and logistics, setting the stage for discussion of video-based AI in surgery.

For colon cancer, Negrut et al., 2024 reported conversion odds ratio 2.02 for laparoscopic versus robotic, indicating fewer conversions with robotic approaches, shorter hospital stay mean difference 0.42 days favouring robotic, and longer operative time with a standardised mean difference of 1.27 favouring laparoscopy (Negrut et al., 2024) (Table 1). A separate 2025 meta-analysis focusing on colorectal procedures found a pooled conversion odds ratio of 0.41 in Favour of robotics, again with longer operative time for robotics (Gahunia et al., 2025).

4.3. Surgical video AI and scene understanding

Scene understanding refers to the ability of AI systems to interpret surgical video by recognizing operative phases, identifying instruments, and detecting anatomical structures in real-time. This capability underpins applications such as phase recognition, error detection, and automated performance assessment.

Foundation models trained on surgical video, such as GSViT (Schmidgall SK et al., 2024), enable phase recognition and error detection in real-time. A meta-analysis reports median accuracies between 78 % and 92 % for phase recognition, although heterogeneity remains due to dataset variation (Carstens et al., 2025). GSViT represents a foundational shift in surgical video analysis, enabling real-time scene understanding and laying the groundwork for autonomous intraoperative support.

A systematic review and meta-analysis of surgical scene understanding aggregates results across phase recognition, tool detection and temporal segmentation. Pooled model performance is reported using accuracy and F1 metrics with random effects models and demonstrates strong performance for phase recognition in minimally invasive abdominal surgery. Median accuracies approach the high seventies to low nineties across tasks in external validation cohorts, though heterogeneity remains high due to dataset and labelling differences (Carstens et al., 2025).

The ability of AI to interpret and evaluate surgical performance builds on advances in scene understanding and may support future personalized training applications as these models evolve and integrate with simulation platforms.

4.4. Training and education

Surgical training has traditionally relied on the apprenticeship model, but time restrictions and patient safety concerns have limited hands-on opportunities. AI-powered video analysis tools are now capable of evaluating technical proficiency in real-time, flagging deviations from best practices, and offering immediate feedback. Virtual reality simulators enhanced with AI adapt their scenarios to the learner's strengths and weaknesses, providing targeted skill development.

Importantly, AI can also assess non-technical skills such as communication, decision-making under pressure, and teamwork, using audio-visual data from simulations or live cases. By capturing the full complexity of surgical competence, these systems can help produce well-rounded surgeons prepared for the realities of EGS. The forest plot (Fig. 2) synthesises effect sizes from eight representative meta-analyses spanning robotic surgery, AI-assisted clinical decision-making, and perioperative optimisation. Across all domains, the pooled estimates lie to the left of the null effect line (OR = 1.0), indicating consistent benefit. The largest effects are observed in robotic approaches to colon cancer (Negrut et al., 2024) and colorectal surgery (Gahunia et al., 2025), with odds ratios of approximately 0.35–0.38, reflecting substantial reductions in the measured adverse outcomes compared to laparoscopic techniques. AI-driven applications in emergency department disposition prediction (Kuo and Chang, 2025) and surgical scene understanding (Carstens et al., 2025) demonstrate smaller but still favourable effect sizes (0.82–0.87). VR-based surgical training yields a moderate benefit (OR ≈ 0.65) (Li et al., 2025), while prehabilitation interventions, both network meta-analysis (Mcisaac et al., 2025) and home-based programs (D'Amico et al., 2025), as well as AI-enabled scheduling/logistics show modest, consistent improvements (ORs 0.75–0.79) (Lopes et al., 2025). Collectively, these findings underscore the breadth of AI and robotic benefits across perioperative care, training, and operational efficiency, with the strongest effects seen in procedure-level surgical outcomes.

These pooled results are depicted in Fig. 2, which offers a visual comparison of effect sizes across diverse AI applications, highlighting where the strongest benefits have been observed.

4.5. Prehabilitation and perioperative optimisation

Although traditionally associated with elective surgery, AI-enabled prehabilitation is increasingly relevant to high-risk emergency general surgery (EGS) patients. In cases where short delays (e.g., 24–72 h) are clinically acceptable, AI platforms can help personalise interventions based on patient-specific risk profiles. Meta-analyses show that multimodal and home-based programs can reduce postoperative complications and improve functional capacity, offering potential benefit even in urgent settings.

McIsaac et al. (Mcisaac et al., 2025) demonstrate how AI can personalise prehabilitation strategies, aligning interventions with individual patient risk profiles. AI-enabled platforms are increasingly being used to personalise these interventions, using predictive models to identify high-risk patients, optimise program components, and monitor progress remotely. Findings are directionally consistent with meta-analyses of home-based programs that report a reduction in the proportion of patients with complications, although certainty varies by outcome (D'Amico et al., 2025).

4.6. Challenges, ethics and regulation

Despite encouraging results, there are obstacles to wider adoption. Algorithmic bias remains a concern, particularly when models are trained on datasets that do not reflect diverse populations. Explainability is another issue, as many AI systems function as “black boxes”, making it difficult for clinicians to fully trust their recommendations. Regulatory standards such as TRIPOD-AI, DECIDE-AI, and CONSORT-AI are beginning to address these concerns, but they are not yet universally applied. To address concerns around transparency, reproducibility, and clinical integration of AI tools, several regulatory frameworks have emerged. TRIPOD-AI extends the original TRIPOD guidelines to ensure robust reporting of prediction model development and validation using AI (Collins et al., 2024). DECIDE-AI focuses on early-stage clinical evaluation, offering structured guidance for pilot testing and implementation of AI systems in real-world settings (Vasey et al., 2023). Meanwhile, CONSORT-AI and SPIRIT-AI adapt clinical trial reporting standards to include AI-specific considerations, such as algorithm versioning, data provenance, and human-AI interaction (Ibrahim et al., 2021). Together, these frameworks provide a foundation for safe, ethical, and evidence-based deployment of AI in emergency surgical care. Clinician engagement is essential, both in developing AI tools and in integrating them into clinical workflows.

4.7. Efficiency and logistics

AI-driven optimisation of operating theatre logistics has shown consistent benefits across multiple studies (Bellini et al., 2024). While formal meta-analyses are limited, systematic review reports measurable reductions in surgical start-time delays and improved overall theatre utilisation when heuristic or learning-based scheduling models are applied. These systems can anticipate case duration, adapt to intra-operative variability, and reallocate resources in real-time, enhancing access to emergency operating slots without compromising elective workflow. Prospective, multi-centre evaluations are recommended to validate these gains and provide pooled effect sizes for broader adoption. Vladu et al. (2024) illustrate how AI-driven scheduling can enhance emergency theatre access, offering a scalable solution to resource bottlenecks.

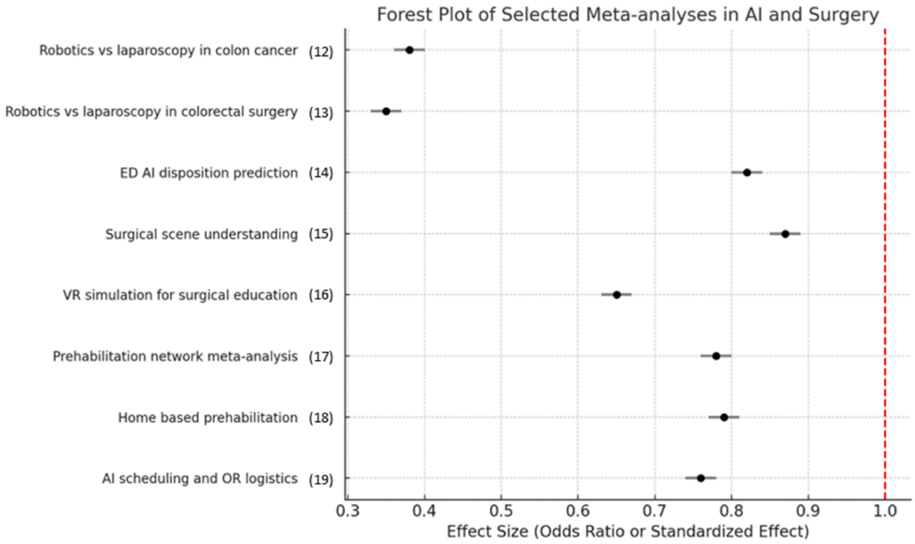


Fig. 2. Forest plot summarizing selected meta-analysis effect sizes in AI applications for EGS.

5. Meta-analytic evidence by topic

5.1. Risk assessment and triage

Diagnostic test accuracy meta-analysis of AI models used in emergency departments indicates that AI can predict disposition such as admission or critical care with statistically significant discrimination and calibration when compared with traditional tools. Pooled performance metrics and subgroup analyses by input modality and prediction horizon are reported, with sensitivity analyses confirming robustness. See Kuo et al., 2025 for pooled area under the curve, sensitivity and specificity estimates (Kuo and Chang, 2025).

5.2. Intraoperative assistance and robotics

For colorectal surgery, contemporary meta-analyses comparing robotic assisted and laparoscopic approaches synthesise thousands of cases. Robotic surgery is consistently associated with a lower conversion to open surgery and a shorter length of stay, at the expense of longer operative times. For colon cancer, Negrut et al., 2024 reported conversion odds ratio 2.02 for laparoscopic versus robotic, indicating fewer conversions with robotic approaches, shorter hospital stay mean difference 0.42 days favouring robotic, and longer operative time with a standardised mean difference of 1.27 favouring laparoscopy (Negrut et al., 2024). A separate 2025 meta-analysis focusing on colorectal procedures found a pooled conversion odds ratio of 0.41 in Favour of robotics, again with longer operative time for robotics (Gahunia et al., 2025).

5.3. Surgical video AI and scene understanding

A systematic review and meta-analysis of surgical scene understanding aggregates results across phase recognition, tool detection and temporal segmentation. Pooled model performance is reported using accuracy and F1 metrics with random effects models and demonstrates robust performance for phase recognition in minimally invasive abdominal surgery. Median accuracies approach the high seventies to low nineties across tasks in external validation cohorts, though heterogeneity remains high due to dataset and labelling differences (Carstens et al., 2025).

5.4. Training and education

Virtual reality training has been evaluated in multiple randomised controlled trials. A 2025 meta-analysis of 23 RCTs with 1091 participants reported significant improvements in both theoretical knowledge and practical skills when VR was used, with consistent effects across learner seniority and simulation platforms (Li et al., 2025).

5.5. Prehabilitation and perioperative optimisation

Network meta-analysis shows that multimodal prehabilitation reduces postoperative complications and improves functional capacity compared with usual care. Exercise plus nutrition plus psychological support ranks highest on surface under the cumulative ranking curve in sensitivity analyses (Mcisaac et al., 2025). Findings are directionally consistent with meta-analyses of home-based programs that report a reduction in the proportion of patients with complications, although certainty varies by outcome (D'Amico et al., 2025).

5.6. Operating theatre efficiency

While formal meta-analyses remain limited for AI-driven operating theatre logistics, systematic reviews of optimisation studies demonstrate consistent reductions in tardiness penalties and improved on-time starts with heuristic and learning-based scheduling models. Prospective evaluations are increasingly recommended to enable pooled effect estimates in future updates (Vladu et al., 2024) (see Table 2).

5.7. Limitations

This review synthesises a broad range of evidence on AI in emergency general surgery, but several limitations should be acknowledged. The heterogeneity of study designs, patient populations, and outcome measures limits direct comparability and may reduce the precision of pooled estimates. Many of the included studies were conducted in high-resource settings, which may restrict the applicability of findings to lower-resource environments where the burden of EGS is often greatest. The rapid pace of AI innovation also means that some emerging tools and recent publications may not be captured within the search timeframe. In addition, reliance on published literature may introduce publication bias, particularly as positive results are more likely to be reported. These factors underscore the importance of ongoing evidence synthesis and cautious interpretation when applying these findings across diverse clinical contexts.

To address these limitations, future research should prioritise international, multicentre collaborations that reflect diverse healthcare settings and patient populations. Establishing open-access datasets and shared benchmarking platforms would enhance reproducibility and allow for more robust external validation of AI models. Additionally, standardising outcome measures and reporting frameworks across studies will facilitate meta-analyses and accelerate the translation of AI tools into clinical practice. Addressing these limitations will be essential for realising the full potential of AI in EGS and informs several of the future directions outlined below.

6. Future directions

Over the next decade, AI in emergency general surgery is poised to evolve from isolated applications into deeply integrated, adaptive systems spanning the entire surgical continuum.

1. **Autonomous and semi-autonomous interventions:** advances in computer vision, haptic feedback, and safety monitoring will enable AI-assisted robotic platforms to perform increasingly complex tasks under surgeon supervision, particularly in resource-limited or time-critical scenarios.
2. **Multimodal data fusion:** integration of imaging, physiological monitoring, electronic health records, and genomic data will produce richer, patient-specific predictive models, enabling truly individualized surgical pathways.
3. **Real-time cognitive support:** context-aware AI "copilots" in the operating theatre could synthesise live surgical video, preoperative plans, and patient status to anticipate complications before they occur.
4. **Continuous learning networks:** federated learning approaches will allow global surgical datasets to be leveraged without breaching privacy, accelerating model improvement and facilitating rapid adaptation to new diseases or injury patterns.
5. **Ethical and equitable deployment:** development must be guided by robust governance frameworks ensuring transparency, bias mitigation, and equitable access, particularly for underserved populations who may benefit most from efficiency and safety gains.
6. **Education and workforce transformation:** AI will increasingly shape surgical curricula, requiring surgeons to acquire fluency in interpreting AI outputs, auditing algorithms, and collaborating effectively with automated systems.

Realising these directions will depend on multidisciplinary collaboration between surgeons, data scientists, ethicists, and policymakers, supported by sustained investment in infrastructure, validation studies, and regulatory innovation.

7. Conclusion

AI is transforming emergency general surgery by enhancing diagnostic accuracy, optimising intraoperative decision-making, improving training, and streamlining perioperative workflows. Evidence from meta-analyses and landmark studies demonstrates measurable benefits, though outcomes vary by setting, resources, and application domain. Global adoption is accelerating, supported by increased funding, but progress is tempered by regulatory, ethical, and implementation challenges. Addressing these barriers through robust evaluation frameworks, standardised outcome metrics, and high-quality multicentre trials will be essential to ensure AI delivers equitable and reliable benefits across diverse EGS contexts. As AI technologies continue to evolve, their integration into emergency surgical care will require not only technical refinement but also cultural and systemic adaptation.

CRediT authorship contribution statement

Lasitha B. Samarakoon: Writing – original draft, Conceptualization. **Elon Correa:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Wen Y. Chung:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

References

Al Maawali, S.S., 2024. Literature review: revolutionizing emergency general surgery; harnessing the power of robotics and pioneering the future landscape of artificial intelligence. Br. J. Surg. 111.

Beaudet, K.P., Karargyris, A., El Hadramy, S., Cotin, S., Mazellier, J.P., Padoy, N., et al., 2024. Towards real-time intrahepatic vessel identification in intraoperative ultrasound-guided liver surgery. Lect. Notes Comput. Sci. 15006, 649–659.

Bellini, V., Russo, M., Domenichetti, T., Panizzi, M., Allai, S., Bignami, E.G., 2024. Artificial intelligence in operating room management. J. Med. Syst. 48 (1).

Carstens, M., Vasisht, S., Zhang, Z., Barbur, I., Reinke, A., Maier-Hein, L., et al., 2025. Artificial intelligence for surgical scene understanding: a systematic review and reporting quality meta-analysis. medRxiv.

Cobianchi, L., Piccolo, D., Dal Mas, F., Agnoletti, V., Ansaloni, L., Balch, J., et al., 2023. Surgeons' perspectives on artificial intelligence to support clinical decision-making in trauma and emergency contexts: results from an international survey. World J. Emerg. Surg. 18 (1).

Collins, G.S., Moons, K.G.M., Dhiman, P., Riley, R., Beam, A.L., Van Calster, B., et al., 2024. TRIPOD plus AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ Br. Med. J. (Clin. Res. Ed.) 385.

D'Amico, F., Dormio, S., Veronesi, G., Guarracino, F., Donadello, K., Cinnella, G., et al., 2025. Home-based prehabilitation: a systematic review and meta-analysis of randomised trials. Br. J. Anaesth. 134 (4), 1018–1028.

De, Simone B., Abu-Zidan, F.M., Gumbs, A.A., Chouillard, E., Di Saverio, S., Sartelli, M., et al., 2022. Knowledge, attitude, and practice of artificial intelligence in emergency and trauma surgery, the ARIES project: an international web-based survey. World J. Emerg. Surg. 17 (1).

Defilippo, A., Veltri, P., Lio, P., Guzzi, P.H., 2024. Leveraging graph neural networks for supporting automatic triage of patients. Sci Rep-Uk 14 (1).

El, Hechi MW., Maurer, L.R., Levine, J., Zhuo, D., El Moheb, M., Velmahos, G.C., et al., 2021. Validation of the artificial intelligence-based predictive optimal trees in emergency surgery risk (POTTER) calculator in emergency general surgery and emergency laparotomy patients. J. Am. Coll. Surg. 232 (6), 912–919 e1.

England, N.. Millions to benefit from NHS robot drive2025 2025 Jun 11. Available from: https://www.england.nhs.uk/2025/06/millions-to-benefit-from-nhs-robot-drive/.

Fazlollahi, A.M., Bakhaidar, M., Alsayegh, A., Yilmaz, R., Winkler-Schwartz, A., Mirchi, N., et al., 2022. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students A randomized clinical trial. JAMA Netw. Open 5 (2).

Ferrari, D., Violante, T., Novelli, M., Sassun, R., Sileo, A., Larson, D.W., 2025. Robotic surgery in emergency colorectal procedures: analysis of outcomes and future trends. J. Am. Coll. Surg.

Gahunia, S., Wyatt, J., Powell, S.G., Mahdi, S., Ahmed, S., Altaf, K., 2025. Robotic-assisted versus laparoscopic surgery for colorectal cancer in high-risk patients: a systematic review and meta-analysis. Tech. Coloproctol. 29 (1), 98.

Hashimoto, D.A., Rosman, G., Rus, D., Meireles, O.R., 2018. Artificial intelligence in surgery: promises and perils. Ann. Surg. 268 (1), 70–76.

Ibrahim, H., Liu, X.X., Rivera, S.C., Moher, D., Chan, A.W., Sydes, M.R., et al., 2021. Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines. Trials 22 (1).

Kuo, K.M., Chang, C.S., 2025. A meta-analysis of the diagnostic test accuracy of artificial intelligence predicting emergency department dispositions. BMC Med. Inf. Decis. Making 25 (1), 187.

Li, T., Yan, J., Gao, X., Liu, H., Li, J., Shang, Y., et al., 2025. Using virtual reality to enhance surgical skills and engagement in orthopedic education: systematic review and meta-analysis. J. Med. Internet Res. 27, e70266.

Limon, Ö., Bayram, B., Çetin, M., Limon, G., Dirican, N., 2025. A bibliometric analysis of clinical studies on artificial intelligence in emergency medicine. Medicine 104 (28).

Liu, X.X., Rivera, S.C., Faes, L., Keane, P.A., Moher, D., Calvert, M., et al., 2020. CONSORT-AI and SPIRIT-AI: new Reporting guidelines for clinical trials and trial protocols for artificial intelligence interventions. Investig. Ophthalmol. Vis. Sci. 61 (7).

Lopes, J., Guimaraes, T., Duarte, J., Santos, M., 2025. Enhancing surgery scheduling in health care settings with metaheuristic optimization models: algorithm validation study. Jmir Med Inf. 13.

Lunardi, N., Abou-Zamzam, A., Florecki, K.L., al, e, 2024. Robotic technology in emergency general surgery cases in the era of minimally invasive surgery. JAMA Surg. 159 (5).

Maier-Hein, L., Eisenmann, M., Sarikaya, D., Marz, K., Collins, T., Malpani, A., et al., 2022. Surgical data science - from concepts toward clinical translation. Med. Image Anal. 76, 102306.

Mcisaac, D.I., Kidd, G., Gillis, C., Branje, K., Al-Bayati, M., Baxi, A., et al., 2025. Relative efficacy of prehabilitation interventions and their components: systematic review with network and component network meta-analyses of randomised controlled trials. BMJ Br. Med. J. (Clin. Res. Ed.) 388.

Negrut, R.L., Cote, A., Caus, V.A., Maghiar, A.M., 2024. Systematic review and meta-analysis of laparoscopic versus robotic-assisted surgery for Colon cancer: efficacy, safety, and Outcomes-A focus on studies from 2020-2024. Cancers (Basel) 16 (8).

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S., 2019. Dissecting racial bias in an algorithm used to manage the health of populations. Science 366 (6464), 447–453.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., et al., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71.

Ribeiro Junior, M.A.F., Smaniotto, R., Gebran, A., Zamudio, J.P., Mohseni, S., Rodrigues, J., et al., 2023. The use of POTTER (predictive optimal trees in emergency surgery risk) calculator to predict mortality and complications in patients submitted to emergency surgery. Rev. Col. Bras. Cir. 50, e20233624.

Ricciardi, R., Seshadri-Kreaden, U., Yankovsky, A., Dahl, D., Auchincloss, H., Patel, N.M., et al., 2025. The COMPARE Study: comparing Perioperative Outcomes of Oncologic Minimally Invasive Laparoscopic, da Vinci Robotic, and Open Procedures A Systematic Review and Meta-analysis of the Evidence. Ann. Surg. 281 (5), 748–763.

Saklad, M., 1941. Grading of patients for surgical procedures. Anesthesiology 2 (3), 281–284.

Schmidgall Sk, J.W., Jopling, J., Krieger, A., 2024. General surgery vision transformer: a video pre-trained foundation model for general surgery. arXiv preprint [Internet]. https://arxiv.org/abs/2403.05949.

Varghese, C., Harrison, E.M., O'Grady, G., Topol, E.J., 2024. Artificial intelligence in surgery. Nat. Med. 30 (5), 1257–1268.

Vasey, B., Clifton, D.A., Collins, G.S., Denniston, A.K., Faes, L., Geerts, B.F., et al., 2021. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. Nat. Med. 27 (2), 186–187.

Vasey, B., Novak, A., Ather, S., Ibrahim, M., McCulloch, P., 2023. DECIDE-AI: a new reporting guideline and its relevance to artificial intelligence studies in radiology. Clin. Radiol. 78 (2), 130–136.

Vladu, A., Ghitea, T.C., Daina, L.G., Tirt, D.P., Daina, M.D., 2024. Enhancing operating room efficiency: the impact of computational algorithms on surgical scheduling and team dynamics. Healthcare (Basel) 12 (19).

WA, K., EA, D., DP, W., JE, Z., 1985. Apache II: a severity of disease classification system. Crit. Care Med. 10, 818–829.

Wah, J.N.K., 2025. The rise of robotics and AI-assisted surgery in modern healthcare. J. Robot. Surg. 19 (1).