

Systematic Review

Open Access



# 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

Aman Goyal<sup>1,2</sup> , Mathew Mendoza<sup>3</sup>, Alfonso E. Munoz<sup>3</sup>, Christian Adrian Macias<sup>4,5,6</sup> , Adel Abou-Mrad<sup>7</sup> ,  
Luigi Marano<sup>8,9,10,11</sup> , Rodolfo J. Oviedo<sup>3,12,13</sup>

<sup>1</sup>Department of General Surgery, Mahatma Gandhi Medical College and Research Institute, Puducherry 607402, India.

<sup>2</sup>Department of Surgery, Adesh Institute of Medical Sciences and Research, Bathinda 151001, India.

<sup>3</sup>Department of Surgery, University of Houston Tilman J. Fertitta Family College of Medicine, Houston, TX 77001, USA.

<sup>4</sup>School of Medicine, Universidad Catolica de Santiago de Guayaquil, Guayaquil 090615, Ecuador.

<sup>5</sup>Department of Health and Sciences, Hillsborough College, Tampa, FL 33614, USA.

<sup>6</sup>Visiting External Researcher, Center for Space Emerging Technologies, Lima 15046, Peru.

<sup>7</sup>Department of Surgery, Centre Hospitalier Universitaire d'Orléans, Orléans 45000, France.

<sup>8</sup>Department of Medicine, Academy of Applied Medical and Social Sciences-AMiSNS: Akademia Medycznych i Społecznych Nauk Stosowanych, Elbląg 82-300, Poland.

<sup>9</sup>Department of General Surgery and Surgical Oncology, "Saint Wojciech" Hospital, "Nicolaus Copernicus" Health Center, Gdańsk 80-000, Poland.

<sup>10</sup>Department of Surgery, Dnipro State Medical University, Dnipro 49044, Ukraine.

<sup>11</sup>Department of Medicine, Surgery, and Neurosciences, University of Siena, Siena 53100, Italy.

<sup>12</sup>Department of Surgery, Nacogdoches Medical Center, Nacogdoches, TX 75962, USA.

<sup>13</sup>Department of Surgery, Sam Houston State University College of Osteopathic Medicine, Conroe, TX 77301, USA.

**Correspondence to:** Dr. Aman Goyal, Department of General Surgery, Mahatma Gandhi Medical College and Research Institute, Puducherry 607402, India. E-mail: doc.aman.goyal@gmail.com

**How to cite this article:** Goyal A, Mendoza M, Munoz AE, Macias CA, Abou-Mrad A, Marano L, Oviedo RJ. 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. <https://dx.doi.org/10.20517/ais.2024.108>

**Received:** 19 Dec 2024 **First Decision:** 2 Jul 2025 **Revised:** 5 Aug 2025 **Accepted:** 27 Aug 2025 **Published:** 12 Sep 2025

**Academic Editor:** Luca Milone **Copy Editor:** Xing-Yue Zhang **Production Editor:** Xing-Yue Zhang

## Abstract

**Introduction:** Artificial intelligence (AI) integration into surgical practice has advanced intraoperative precision, complication prediction, and procedural efficiency. While AI has demonstrated advancements in colorectal, cardiac, and other laparoscopic procedures, its application in inguinal hernia repair (IHR), one of the most commonly performed surgeries, remains underexplored. AI models demonstrate potential in real-time recognition



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



of surgical phases, anatomical structures, and instruments, particularly in transabdominal preperitoneal (TAPP), total extraperitoneal (TEP), and robotic inguinal hernia repair (RIHR). This systematic review evaluates the accuracy, applicability, and clinical impact of AI-based systems in real-time surgical phase recognition during IHR.

**Methods:** Following PRISMA 2020 guidelines and PROSPERO registration (CRD42024621178), a systematic search of PubMed, Scopus, Web of Science, Embase, Cochrane Library, and ScienceDirect was conducted on November 12, 2024. Studies utilizing AI models for real-time video-based surgical phase recognition in minimally invasive IHR (TAPP, TEP, and RIHR) were included. The screening process, data extraction task, and quality assessment using NOS (Newcastle-Ottawa Scale) were performed by three independent reviewers. Primary outcomes were AI performance metrics (accuracy, F1-score, precision, recall, and latency), and secondary outcomes included clinical phase recognition performance.

**Results:** Out of 903 records, six studies (2022-2024) were included, involving laparoscopic ( $n = 4$ ) and robotic-assisted ( $n = 2$ ) IHR from the United States ( $n = 2$ ), France ( $n = 2$ ), and Greece ( $n = 1$ ). A total of 774 videos (25-619 per study) underwent pre-processing (frame extraction or down-sampling). Annotation tools included CVAT, SuperAnnotate, and manual labeling. AI models (VTN, DETR, ResNet-50, YOLOv8) demonstrated accuracy between 74% and > 87%, with YOLOv8 achieving the highest F1-score (82%). Risk of bias was moderate to high, with Fleiss' kappa for inter-rater agreement at 0.82 (selection) and 0.49 (comparability).

**Conclusion:** AI and ML models demonstrate significant potential in achieving real-time surgical phase recognition during minimally invasive IHR. Despite promising accuracies, challenges such as heterogeneity in model performance, reliance on annotated datasets, and the need for real-time validation persist. Standardized benchmarks, multicenter studies, and hardware advancements will be essential to fully integrate AI into surgical workflows, improving surgical training, technical performance, and patient outcomes.

**Keywords:** Artificial intelligence, inguinal hernia repair, real-time surgical phase recognition, minimally invasive surgery, AI performance metrics

## INTRODUCTION

The revolutionary adoption of artificial intelligence (AI) in surgical applications has grown significantly in recent years, offering transformative potential across a variety of procedures. The use of AI in medicine dates back to the 1970s, with early systems assisting in bacterial detection, decision making, and imaging analysis<sup>[1,2]</sup>. As AI models and programming advanced, their application in surgical practice became increasingly feasible. In recent literature, the use of AI has been appraised to be valuable in areas such as colorectal surgery, laparoscopic cholecystectomy, and cardiac surgery<sup>[3,4]</sup>. However, most of these procedures utilize AI in specific aspects of their interventions, with notable effectiveness in assessing complication risks, avoiding intraoperative errors, and predicting mortality outcomes<sup>[3-5]</sup>.

An emerging field that has recently seen an increase in the incorporation of AI technologies is inguinal hernia repair (IHR), the most prevalent groin repair procedure<sup>[6]</sup>. By training AI models using compiled videos and imaging data to identify specific steps, anatomical structures, and instruments during procedures<sup>[7]</sup>, medical teams can leverage real-time assistance to enhance decision making, improve surgical efficiency, and support education and workflow optimization<sup>[8,9]</sup>. This advantage is novel and has recently been explored in studies across different AI systems.

Current research presents this variety of AI models developed to monitor and assist in different bilateral and unilateral<sup>[8,10]</sup> IHR techniques, including transabdominal preperitoneal (TAPP), total extraperitoneal (TEP), and robotic inguinal hernia repairs (RIHR)<sup>[8-10]</sup>. These AI applications optimize procedural efficiency

and accuracy, significantly contributing to better surgical outcomes. Furthermore, the concurrent advancement of both laparoscopic and robotic techniques has been instrumental in advancing AI integration in hernia repair.

Due to the nascent nature of AI use in IHR, there is limited literature that compiles and aggregates the benefits and disadvantages of this practice. To the best of our knowledge, this review is the first systematic evaluation to compile and analyze the real-time intraoperative use of AI models in IHR. While recent discoveries show promising results, ongoing data collection and findings indicate that this area of practice and research is continuously evolving.

The purpose of this systematic review was to evaluate the latest use of AI models in the real-time intraoperative phase of IHR for their accuracy, applicability, and potential influence on clinical outcomes and surgical training. Additionally, the number of surgical phases recognized by each AI model is considered as a measure of phase-specific accuracy. This review provides a comprehensive and novel analysis of the recent performance and transformative potential of video-trained surgical phase-recognition AI systems, assessing their ability to enhance surgical outcomes and redefine traditional practices.

## **MATERIALS AND METHODS**

### **Study design**

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P 2020) guidelines<sup>[11]</sup> to assess the application of AI in real-time surgical phase recognition during minimally invasive IHR. Both robotic and laparoscopic approaches, including TAPP and TEP procedures, were included. The study protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) under the number CRD42024621178. Approval from an Institutional Review Board (IRB) committee was not necessary or indicated, as we used data from published primary studies available in the literature.

### **Information sources and search strategy**

A comprehensive search was performed in six electronic databases: PubMed, Scopus, Web of Science, Embase, Cochrane Library, and ScienceDirect. The final search was conducted on November 12, 2024. Language, geographical location, and date of publication were not imposed restrictions. The search terms included were derived from the keywords “artificial intelligence”, “machine learning”, “computer vision systems”, “inguinal hernia”, and “herniorrhaphies”. The search strategy, covering Boolean operators and keywords, is detailed in [Supplementary Table 1](#). Manual searches of reference lists and relevant studies were also performed to ensure comprehensive coverage.

### **Study selection**

The titles and abstracts of the identified studies were screened by three independent reviewers to ensure relevance based on the predefined inclusion and exclusion criteria. The final stage of eligibility was assessed by reviewing the full text of potentially relevant studies. Reviewers resolved any discrepancies through discussion and by consensus. Eligibility criteria included prospective and retrospective studies that used AI-based video recordings for surgical phase recognition in minimally invasive inguinal hernia repair, either laparoscopic or robotic. Both unilateral and bilateral hernia repairs were incorporated. Exclusion criteria involved case reports, case series, reviews, commentaries, conference abstracts, and editorials. Studies focusing solely on image-based AI models or evaluating skills without surgical phase recognition were excluded.

### Data extraction

Three independent reviewers performed the data extraction using a standardized data collection form. Extracted variables included:

- Study characteristics (authors, year of publication, country, study design)
- Procedure details (TAPP, TEP, laparoscopic vs. robotic, unilateral vs. bilateral)
- AI model specifics (model type, architecture, training dataset, ground-truth annotations)
- Performance metrics (accuracy, F1 score, precision, recall, latency)
- Surgical phase recognition outcomes
- Number of videos analyzed, surgeons involved, and video collection period

Any disagreements during data extraction were resolved by consensus or consultation with a third reviewer.

### Bias and quality assessment

The Newcastle-Ottawa Scale (NOS) was used to assess the risk of bias in the included studies. Three independent reviewers rated each study based on three domains: selection, comparability, and outcome assessment. Studies scoring 7-9 were considered high quality, 4-6 as moderate quality, and < 4 as low quality<sup>[12]</sup>. Inter-rater reliability was assessed using the Fleiss' kappa analysis. Discrepancies were resolved through discussion or input from a third reviewer.

### Outcomes and effect measures

The primary outcomes sought were AI model performance metrics: accuracy, F1 score, precision, recall, and latency. The F1 score, a metric that combines precision and recall, was specifically collected for studies employing classification models. Secondary outcomes included surgical phase recognition performance and details of edge computing devices, where applicable.

### Data synthesis and presentation

Due to the heterogeneity of AI models and outcome measures, a narrative synthesis was performed. Data were tabulated to summarize study characteristics and outcomes. Meta-analysis and sensitivity analysis were not conducted due to insufficient data and variability in study methodologies. Findings are presented in descriptive tables and figures to facilitate interpretation.

## RESULTS

### Study selection

A comprehensive search was conducted across six databases, yielding 903 records in the initial literature search. After removing 169 duplicates, 734 records remained for screening. Based on pre-determined inclusion and exclusion criteria, we conducted a title and abstract review, selecting six studies for full-text analysis. Of these, one study was excluded because it utilized images for surgical phase recognition<sup>[13]</sup>. After a thorough full-text review, data were extracted from five articles, all published in English. Additionally, a citation search was performed to ensure comprehensive coverage of relevant literature, identifying five additional references within the scope of our study. Upon reviewing these, one was excluded as it was a review article<sup>[14]</sup>, and three others were excluded for employing image-based surgical phase recognition

methods<sup>[15-17]</sup>. Consequently, one additional study was included in our analysis<sup>[7]</sup>. The detailed reasons for exclusion and the study selection process are illustrated in the PRISMA flowchart [Figure 1], created using the online tool developed by Haddaway *et al.*<sup>[18,19]</sup>.

### Study characteristics

This systematic review included six studies published between 2022 and 2024, focusing on laparoscopic or robotic-assisted IHR procedures. The studies were conducted in three countries: the United States ( $n = 2$ ), France ( $n = 2$ ), and Greece ( $n = 1$ ). The hernia repair techniques investigated across the included studies comprised laparoscopic total extraperitoneal (TEP), transabdominal preperitoneal (TAPP), and robotic-assisted inguinal hernia repair (RIHR). Four studies utilized laparoscopic approaches<sup>[7,8,10,20]</sup>, while two employed robotic-assisted techniques<sup>[9,21]</sup>.

The sample sizes varied significantly across the studies. One multicenter study, conducted in the USA, reported a total of 619 cases, comprising 270 unilateral and 349 bilateral hernias<sup>[10]</sup>. In contrast, the single-center studies mostly included unilateral IHR, except for one study by Takeuchi *et al.* (2022), which reported 86 unilateral and 33 bilateral cases<sup>[8]</sup>. The periods of video data collection, when provided, spanned from July 2019 to December 2022. However, two studies did not specify the time frame for their video collections<sup>[7,21]</sup>. All studies used an AI model architecture to detect the surgical phase recognition in patients undergoing minimal invasive IHR [Table 1].

### Results on individual studies

#### *Number of videos analyzed*

From the included studies in the systematic review, encompassing both laparoscopic and robotic surgery videos. The total number of videos analyzed varied across studies, ranging from 25 videos in the study by Zygomalas *et al.* (2024)<sup>[7]</sup> to 619 videos in the study by Ortenzi *et al.* (2023)<sup>[10]</sup>. Among these, the studies employed varying proportions of videos for specific tasks such as training, validation, and testing. A majority of the studies prioritized training datasets, with proportions ranging from 75% to over 85% of the total videos. Internal validation and testing were inconsistently reported, with only a few studies detailing their allocation strategies [Table 2].

#### *Number of surgeons involved*

The number of surgeons contributing to the studies varied. Ortenzi *et al.*'s study (2023)<sup>[10]</sup> had the largest cohort of surgeons involved ( $n = 85$ ), while Zygomalas *et al.*'s study (2024)<sup>[7]</sup> reported the smallest group ( $n = 5$ ). The studies by Takeuchi *et al.* (2023)<sup>[20]</sup> and Zang *et al.* (2023)<sup>[9]</sup> involved 11 and 8 surgeons, respectively. Similarly, Choksi *et al.* (2023)<sup>[21]</sup> included 8 surgeons in their study.

#### *Pre-processing of videos*

Prior to annotation, surgical videos underwent systematic pre-processing to ensure uniformity, reduce computational load, and optimize datasets for analysis. Ortenzi *et al.* [10] utilized a Surgical Intelligence platform (Theator Inc., Palo Alto, CA) to automatically capture, de-identify, and upload videos to a secure cloud infrastructure. Takeuchi *et al.* [20] sampled videos uniformly over time to prevent selection bias, with approximately  $18.12 \pm 11.59$  images extracted per video, resulting in 2,356 training images. Additionally, 10 test images per video (a total of 300) were selected to represent the CVMPO visual inspection phase. Choksi *et al.*<sup>[21]</sup> down-sampled the videos to 1 frame per second (FPS) to reduce memory usage, which has been shown to be sufficient for phase segmentation tasks. Similarly, Zang *et al.*<sup>[9]</sup> processed videos using the FFmpeg library to reduce the frame rate from 30 FPS to 1 FPS, a reduction that had minimal impact on phase recognition performance. Zygomalas *et al.*<sup>[7]</sup> extracted 1,095 images from 25 TAPP videos using Bounding Boxes for object labeling, resulting in a dataset of 800 images for training and 295 for testing

**Table 1. Baseline characteristics of included studies in the study**

Authors	Year of publication	Country	Video collection time period	Inguinal hernia procedure (TAPP/TEP/RIHR)	Laparoscopic/robotic	Number of surgeons involved	Single centric/multicentric	Unilateral/bilateral hernia
Ortenzi <i>et al.</i> <sup>[10]</sup>	2023	USA	October 25, 2019, and December 13, 2022	TEP	Laparoscopic	85	Multicentric	Yes (270 unilateral; 349 bilateral cases)
Takeuchi <i>et al.</i> <sup>[20]</sup>	2023	France	July 2019 and April 2021	TAPP	Laparoscopic	11	Single	Unilateral
Choksi <i>et al.</i> <sup>[21]</sup>	2023	USA	-	RIHR	Robotic	8	Single	Unilateral
Takeuchi <i>et al.</i> <sup>[8]</sup>	2022	France	July 2019-February 2021	TAPP	Laparoscopic	11	Single	Yes (86 unilateral; 33 bilateral)
Zang <i>et al.</i> <sup>[9]</sup>	2023	USA	Spring 2022	RALIHR (robotic assisted)	Robotic	8	Single	Unilateral
Zygomalas <i>et al.</i> <sup>[7]</sup>	2024	Greece	-	TAPP	Laparoscopic	5	Single	Unilateral

TAPP: Transabdominal preperitoneal; TEP: totally extraperitoneal; RIHR: robotic inguinal hernia repair.

**Table 2. Pre-processing and number of videos analyzed in included studies**

Authors	Pre-processing of videos	Total	For training	Internal validation	Test
Ortenzi <i>et al.</i> <sup>[10]</sup>	Surgical Intelligence platform used for automatic video capture, de-identification, and cloud upload	619	371	93	155
Takeuchi <i>et al.</i> <sup>[20]</sup>	Uniform video sampling to prevent selection bias; ~18.12 ± 11.59 images extracted per video. 2,356 training images; 300 test images representing the CVMPO visual inspection phase	160	130	-	30
Choksi <i>et al.</i> <sup>[21]</sup>	Videos down-sampled to 1 frame per second (FPS) to reduce memory usage for phase segmentation	211	106	30	60
Takeuchi <i>et al.</i> <sup>[8]</sup>	Videos annotated and organized using the Indexity Annotation Platform at IRCAD France and IHU Strasbourg	119	89	-	30
Zang <i>et al.</i> <sup>[9]</sup>	Frame rate reduced from 30 FPS to 1 FPS using the FFmpeg library to minimize computational load	209	188	-	21
Zygomalas <i>et al.</i> <sup>[7]</sup>	25 TAPP videos processed to extract 1,095 images, annotated using Bounding Boxes (11 object classes)	25	800*	-	295*

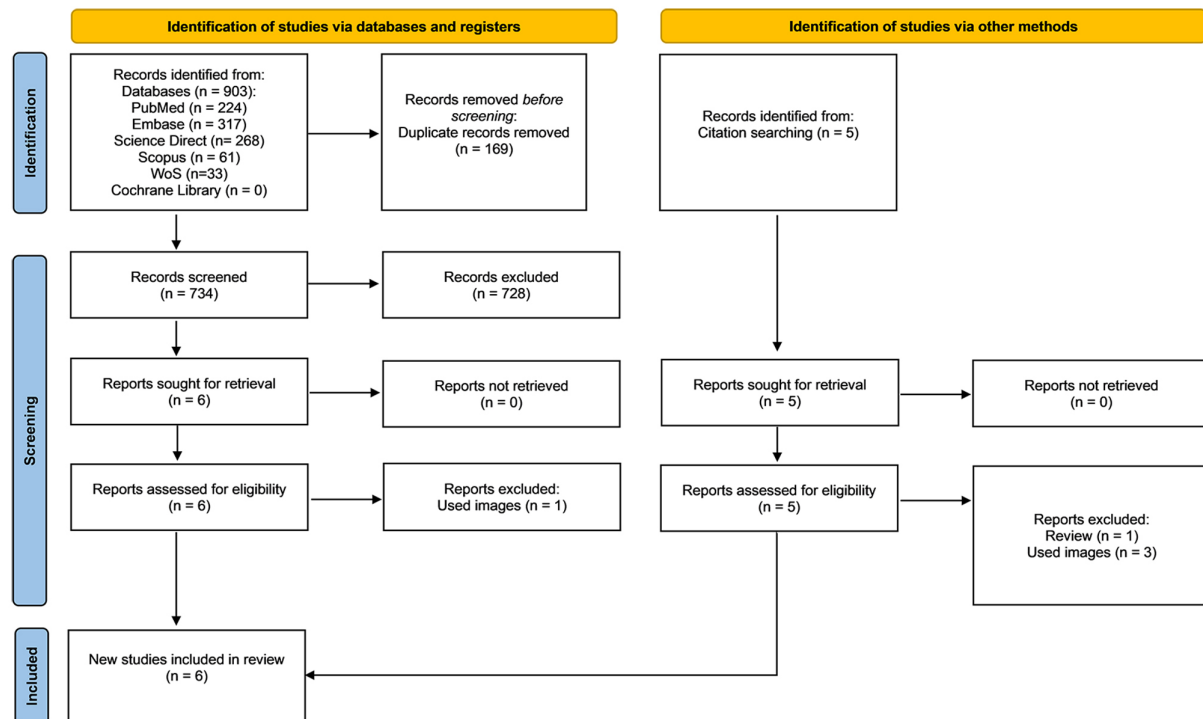
\*Zygomalas *et al.* randomly selected images from 25 study videos for training and testing the AI model. AI: Artificial intelligence.

[Table 2]. These pre-processing steps ensured consistent, high-quality data input for subsequent annotation and model development.

### Annotation software

The term annotation refers to the process of manually marking the start and end points of each procedural step. The included studies utilized various annotation methods and software to process the videos prior to testing and internal validation. Manual annotation was employed in multiple studies, including those conducted by Ortenzi *et al.* (2023)<sup>[10]</sup>, Choksi *et al.* (2023)<sup>[21]</sup>, and Zang *et al.* (2023)<sup>[9]</sup>. Computer Vision Annotation Tool (CVAT) was specifically utilized in the study by Takeuchi *et al.* (2023)<sup>[20]</sup>, while Indexity Annotation Platform was adopted in a separate study by Takeuchi *et al.* (2022)<sup>[8]</sup>. In contrast, SuperAnnotate v1.1.0 was used in the study by Zygomalas *et al.* (2024)<sup>[7]</sup>. These annotation tools facilitated detailed video labeling, which is critical for creating training, validation, and testing datasets essential for





**Figure 1.** Search outputs based on PRISMA guidelines<sup>[19]</sup>. PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis.

subsequent analyses. The choice of annotation tools varied based on study design, with manual annotation being predominant, potentially due to its precision in labeling surgical steps and events. However, automated platforms like CVAT and SuperAnnotate highlight the integration of advanced technologies to streamline the annotation process, ensuring both accuracy and efficiency in preparing datasets for machine learning-based analyses [Table 3].

#### AI models and architectures

A range of AI models and architectures were employed across the included studies to train systems for recognizing annotated structures within surgical videos or images. Ortenzi *et al.*<sup>[10]</sup> utilized a Surgical Intelligence Platform developed by Theator Inc. (Palo Alto, CA), integrating a Video Transformer Network (VTN) and a Long Short-Term Memory (LSTM) network to adapt the model for sequential data. This approach enabled effective temporal learning from video data. Takeuchi *et al.*<sup>[20]</sup> implemented DETR (Detection Transformer), leveraging its advanced transformer architecture to process visual inputs for identifying structures. Similarly, Choksi *et al.*<sup>[21]</sup> relied on ResNet-50, a well-established convolutional neural network (CNN), to process images with robust feature extraction capabilities.

In studies combining temporal and probabilistic modeling, Takeuchi *et al.*<sup>[8]</sup> adopted a hybrid architecture combining Temporal Convolutional Networks (TeCNO) with a Hidden Markov Model (HMM). This setup was effective for handling sequential data while modeling probabilistic dependencies. Zang *et al.*<sup>[9]</sup> explored a multi-model approach incorporating ResNet-50, Video Swin Transformer, and Perceiver IO, demonstrating the potential of combining different architectures to enhance the performance of AI systems. Meanwhile, Zygomalas *et al.*<sup>[7]</sup> utilized a YOLOv8 (<https://yolov8.com/>) deep learning model, known for its efficiency in real-time object detection, for analyzing annotated video data [Table 3].

**Table 3. Comparison of AI models, annotation tools, and accuracy metrics for surgical phase recognition**

Authors	Annotation software	AI model type	Model architecture	Number of surgeons	Overall accuracy	F1 score	Precision
Ortenzi <i>et al.</i> (2023) <sup>[10]</sup>	Manual annotation	Surgical Intelligence Platform (Theator Software Co.)	Video Transformer Network (VTN), Long Short-Term Memory (LSTM) network (for model adaptation)	85	88.8% (89.16% with an overall 89.59% for unilateral; 83.17% with an overall 88.45% for bilateral)	-	-
Takeuchi <i>et al.</i> (2023) <sup>[20]</sup>	CVAT (Computer Vision Annotation Tool)	DETR (Detection Transformer)	-	11	77.10%	75.40%	51.20%
Choksi <i>et al.</i> (2023) <sup>[21]</sup>	Manual annotation	ResNet-50 (Microsoft Research)	-	8	68.7% (59.8% - 78.2%)	-	-
Takeuchi <i>et al.</i> (2022) <sup>[8]</sup>	Indexiv Annotation Platform	TeCNO (Transsion Holdings) & Hidden Markov Model (HMM)	Temporal Convolutional Network (MS-TCN) + Hidden Markov Model (HMM)	11	Unilateral: 88.81%; Bilateral: 85.82%	-	-
Zang <i>et al.</i> (2023) <sup>[9]</sup>	Manual annotation	CV/DL	ResNet-50, Video Swin Transformer, Perceiver IO	8	85%	75.81%	-
Zygomalas <i>et al.</i> (2024) <sup>[7]</sup>	SuperAnnotate v1.1.0	YOLOv8 (Ultralytics Inc.)	-	5	87.30%	82%	> 50%

AI: Artificial intelligence.

These models were trained using annotated datasets, allowing the systems to learn and identify specific structures accurately. The choice of model architecture was influenced by the nature of the input data (videos or images) and the specific requirements of the task, such as temporal modeling, feature extraction, or real-time analysis.

### *Surgical phase recognition*

Significant heterogeneity exists in the identification and classification of surgical phases during IHR across included studies. While some studies focus on broad procedural steps such as “balloon dissection” and “mesh placement”<sup>[10]</sup>, others incorporate a more granular breakdown, identifying anatomical landmarks like the “pubic symphysis” and “iliopsoas muscle”<sup>[20]</sup>. Additionally, phases related to peritoneal manipulation, including “peritoneal flap incision” and “closure,” were emphasized in works by Takeuchi *et al.*<sup>[8]</sup> and Zang *et al.*<sup>[9]</sup>.

Notably, certain steps critical to surgical workflows, such as “needleholder” handling and recognition of vascular or ligamentous structures, were unique to specific models<sup>[7]</sup>. Late procedural phases, including “stationary idle” and “blurry” recognition, were inconsistently reported, underscoring variability in endpoint definition [Table 4]. Establishing a standardized framework for phase segmentation remains critical to improving model performance, ensuring comparability, and advancing automation in inguinal hernia surgery.



**Table 4. Recognition of surgical phases across studies**

Authors	Ortenzi <i>et al.</i> (2023) <sup>[10]</sup>	Takeuchi <i>et al.</i> (2023) <sup>[20]</sup>	Choksi <i>et al.</i> (2023) <sup>[21]</sup>	Takeuchi <i>et al.</i> (2022) <sup>[8]</sup>	Zang <i>et al.</i> (2023) <sup>[9]</sup>	Zygomalas <i>et al.</i> (2024) <sup>[7]</sup>
Phase 1	Balloon dissection	Pubic symphysis	Adhesiolysis	Preparation	Adhesiolysis	Mesh
Phase 2	Access to the preperitoneal space	Direct hernia orifice	Peritoneal Scoring	Peritoneal flap incision	Peritoneal Scoring	Grasp
Phase 3	Preperitoneal dissection	Cooper ligament	Preperitoneal dissection	Peritoneal flap dissection	Preperitoneal dissection	Bowel
Phase 4	Hernia sac reduction	Iliac vein	Reduction of hernia	Hernia dissection	Reduction of hernia	Needleholder
Phase 5	Mesh placement	Doom triangle	Mesh positioning	Mesh deployment	Mesh positioning	Umbilical ligament
Phase 6	Mesh fixation	Deep inguinal ring	Mesh placement	Mesh fixation	Mesh placement	Inf Epi Vessel
Phase 7	-	Iliopsoas muscle	Positioning suture	Closing of peritoneal flap, additional closure	Positioning suture	Spermatic Vessel
Phase 8	-	-	Securing mesh	No step	Primary hernia repair	Cooper
Phase 9	-	-	Primary hernia repair	-	Catheter insertion	Bipolar
Phase 10	-	-	Catheter insertion	-	Peritoneal closure	Vas Deferens
Phase 11	-	-	Peritoneal closure	-	Transitory idle	Scissors
Phase 12	-	-	Transitory idle	-	Stationary idle	-
Phase 13	-	-	Stationary idle	-	Out of body	-
Phase 14	-	-	Out of body	-	Blurry	-
Phase 15	-	-	Blurry	-	-	-

#### *Accuracy and performance metrics across studies*

The accuracy and performance metrics of the included studies are summarized in Table 3. Overall, there was variability in reported accuracy and performance across studies, influenced by methodological differences and dataset characteristics. Ortenzi *et al.*<sup>[10]</sup> and Zygomalas *et al.*<sup>[7]</sup> demonstrated the highest accuracy, exceeding 87%, while Choksi *et al.*<sup>[21]</sup> reported lower accuracy with considerable variation across cases. F1 scores, reported in a subset of studies, further highlighted predictive performance, with Zygomalas *et al.*<sup>[7]</sup> achieving the highest score at 82%. Precision metrics were also noted, with values ranging from 51% to above 50% in studies by Takeuchi *et al.*<sup>[20]</sup> and Zygomalas *et al.*<sup>[7]</sup>.

#### *Risk of bias*

The risk of bias for all six included studies was assessed using the Newcastle-Ottawa Scale (NOS) across three domains: Selection, Comparability, and Outcome Assessment. Three independent reviewers conducted the assessment, and consensus scores were derived after resolving discrepancies. Two studies (Ortenzi *et al.*, 2023<sup>[10]</sup>; Takeuchi *et al.*, 2023<sup>[20]</sup>) were classified as high quality, while the remaining four were rated as moderate quality based on their total NOS scores. The scores for each domain and their totals are presented in Table 5.

To evaluate the reliability of these assessments, inter-rater agreement was calculated using Fleiss' kappa for each domain. The selection domain exhibited the strongest agreement, with  $\kappa = 0.82$  (95% CI: 0.74-0.90), reflecting high consistency across reviewers. Comparatively, the comparability domain demonstrated moderate agreement, with  $\kappa = 0.49$  (95% CI: 0.37-0.61), while the outcome domain showed fair agreement, with  $\kappa = 0.44$  (95% CI: 0.32-0.56). The overall agreement was 0.89, 0.73, and 0.67 for the selection, comparability, and outcome domains, respectively, with corresponding expected agreement values of 0.39, 0.47, and 0.41. Notably, Choksi *et al.* (2023)<sup>[21]</sup> had the lowest inter-rater agreement within the outcome domain ( $\kappa = 0.33$ ), highlighting variability in the reviewers' interpretation of the NOS criteria for this study.

**Table 5. Newcastle-Ottawa Scale (NOS) for risk of bias assessment**

Authors	Reviewers	Selection	Comparability	Outcomes	Total	Final score after consensus
Ortenzi <i>et al.</i> (2023) <sup>[10]</sup>	1st	4	2	2	8	8
	2nd	4	1	2	7	
	3rd	4	2	2	8	
Takeuchi <i>et al.</i> (2023) <sup>[20]</sup>	1st	3	1	3	7	7
	2nd	4	1	3	8	
	3rd	3	0	1	4	
Choksi <i>et al.</i> (2023) <sup>[21]</sup>	1st	3	1	2	6	6
	2nd	3	2	3	8	
	3rd	3	2	1	6	
Takeuchi <i>et al.</i> (2022) <sup>[8]</sup>	1st	3	1	2	6	6
	2nd	4	2	2	8	
	3rd	4	1	1	6	
Zang <i>et al.</i> (2023) <sup>[9]</sup>	1st	3	1	2	6	6
	2nd	3	1	2	6	
	3rd	3	1	2	6	
Zygomalas <i>et al.</i> (2024) <sup>[7]</sup>	1st	2	1	2	5	5
	2nd	2	1	2	5	
	3rd	2	1	1	4	

Bias sources were further analyzed to identify recurring issues across the studies. The most frequent concerns included attrition bias due to incomplete outcome data and single-center study designs, which limited the diversity of datasets. Additionally, imbalanced case distributions, where the number of videos analyzed per surgeon varied, were observed. In some instances, inadequate case annotations impacted the robustness of outcome assessments. These limitations collectively contribute to the moderate quality ratings assigned to the majority of studies in this analysis [Table 6].

## DISCUSSION

In this systematic review, we evaluated studies utilizing AI and machine learning (ML) techniques to train models for real-time surgical phase recognition during minimally invasive IHR procedures. To our knowledge, this is the first systematic review to focus on this critical intersection of AI and surgical workflow, underscoring its potential to transform the future of surgery. AI holds immense promise for healthcare, offering opportunities to reduce labor demands and improve the quality of clinical analysis and decision making<sup>[22]</sup>. Over the past few years, the role of artificial intelligence in hernia surgery has been examined from broad perspectives. For instance, Lima *et al.* (2024)<sup>[23]</sup> performed a qualitative systematic review of 13 peer-reviewed studies published between 2020 and 2023. Their analysis suggests that machine learning platforms hold promise as tools for predicting surgical outcomes and identifying risk factors to mitigate postsurgical complications.

The integration of AI into healthcare is already reshaping patient care across multiple domains. For instance, AI-based systems like Babylon Health are enhancing both in-person and virtual consultations by providing timely and accurate assistance<sup>[24]</sup>. Similarly, AI-powered tools such as Sense.ly, which combines a virtual assistant with an intuitive, human-like interface, are advancing medication management and health monitoring<sup>[25]</sup> in AI-driven diagnostics. A particularly notable example of AI-driven diagnostics is Caption Health's FDA-approved software, which enables healthcare providers to perform cardiac ultrasound imaging without extensive sonographic expertise. Acting as a "co-pilot," this AI software offers real-time guidance, instructing users on transducer manipulation and ensuring the acquisition of diagnostic-quality images<sup>[26]</sup>.

**Table 6. Fleiss' kappa analysis for inter-rater agreement**

Metric	Selection domain	Comparability domain	Outcome domain
Overall agreement	0.89	0.73	0.67
Expected agreement	0.39	0.47	0.41
<b>Fleiss' kappa (<math>\kappa</math>)</b>	<b>0.82</b>	<b>0.49</b>	<b>0.44</b>

In surgical practice, AI-driven phase recognition has emerged as a critical tool for workflow analysis and the evaluation of technical performance. Automation of operative phase segmentation and classification has the potential to train residents, improve skills, reduce iatrogenic complications, improve efficiency, and minimize costs. While various surgical specialties are exploring AI applications, laparoscopic procedures have been the primary focus of most studies. For example, significant advancements have been reported for surgeries such as laparoscopic sleeve gastrectomy<sup>[27,28]</sup>, sigmoidectomy<sup>[29]</sup>, myotomy<sup>[30]</sup>, and cholecystectomy<sup>[31]</sup>.

AI functions primarily through ML, a paradigm that allows computers to learn and adapt to complex tasks without being explicitly programmed<sup>[32]</sup>. Traditional algorithms, which rely on predefined rules, often fall short when tackling intricate problems in healthcare where variability and complexity abound. ML addresses this limitation by enabling computers to analyze large volumes of high-quality annotated data, iteratively improving their performance. Rather than relying on rigid rules to identify surgical phases, AI models are trained using vast annotated datasets, allowing algorithms to discern patterns and optimize phase recognition. Advanced techniques such as deep learning and neural networks further enhance AI's capabilities by autonomously processing unstructured data, such as surgical videos, and extracting temporal and spatial dependencies to predict surgical workflows.

The *Cholec80* dataset has been widely recognized as a benchmark for surgical phase recognition, providing a foundation for the development of deep learning models. Early attempts in 2016, such as EndoNet, achieved 75.2% accuracy by leveraging convolutional neural networks (CNN) for feature extraction combined with support vector machines and hidden Markov models for temporal phase inference<sup>[9]</sup>. Subsequent advancements incorporated temporal dependencies, as seen with models like SV-RCNet, which utilized long short-term memory (LSTM) networks to achieve 81.6% accuracy in 2018<sup>[33]</sup> and TeCNO (88.6%, 2020) using 3D convolutions<sup>[34]</sup>.

The introduction of attention mechanisms further enhanced the performance of deep learning models, as demonstrated by TMRNet (90.1%, 2021)<sup>[35]</sup> and Trans-SVNet (90.3%, 2021)<sup>[36]</sup>. Most recently, in 2023, a model incorporating 3D convolutions with positional encodings achieved an accuracy of 92.3%, representing the highest reported performance on the *Cholec80* dataset to date<sup>[37]</sup>.

Building on these advancements, the current review highlights the expanding applicability of AI models across various surgical procedures, particularly in laparoscopic and robotic hernia repairs. In the study by Takeuchi et al. (2022), the TeCNO + HMM model, which integrates temporal convolutional networks and hidden Markov models, demonstrated accuracies of 88.81% for unilateral and 85.82% for bilateral TAPP repairs<sup>[8]</sup>. These findings align closely with the performance of earlier attention-based models such as TMRNet and Trans-SVNet, which achieved accuracies of 90.1% and 90.3%, respectively, in cholecystectomy workflows<sup>[35,36]</sup>. Similarly, Takeuchi et al. (2023) applied the transformer-based DETR model to analyze 211

TAPP videos, achieving an accuracy of 77.1%, with an F1 score of 75.4% and precision of 51.2%<sup>[20]</sup>. While these results underscore the trade-off between precision and scalability, they also highlight the adaptability of transformer models in surgical phase recognition.

The study by Zygomalas *et al.* (2024) utilized the YOLOv8 deep learning model to analyze laparoscopic TAPP repair procedures, reporting an accuracy of 87.3% and an F1 score of 82%<sup>[7]</sup>. However, unlike the TeCNO and ResNet-50 models, phase segmentation was not explicitly addressed, underscoring ongoing variability in study endpoints. Notably, the study by Choksi *et al.* (2023) implemented the ResNet-50 model in RIHR, reporting an accuracy of 68.7% and identifying seven procedural phases<sup>[21]</sup>. This study also provided a critical latency period of 250 milliseconds, a key consideration for real-time applications, especially when paired with edge computing devices like the Nvidia Jetson Xavier™.

In contrast, Zang *et al.* (2023) analyzed RALIHR using a CV/DL (Computer Vision/Deep Learning) model, achieving an accuracy of 85% with an F1 score of 78.51%<sup>[9]</sup>. This study also utilized Nvidia A100 as its edge computing device, further highlighting the role of hardware advancements in improving model performance. Finally, the study by Ortenzi *et al.* (2023) stands out for its scale, analyzing 619 TEP repairs across multiple centers and involving 85 surgeons<sup>[10]</sup>. By employing the Surgical Intelligence Platform incorporating Video Transformer Networks (VTN) and LSTM networks, this study achieved a mean accuracy of 88.8%, with 89.59% for unilateral cases and 88.45% for bilateral cases. These findings reaffirm the significance of large-scale datasets and multi-institutional collaborations in enhancing AI model generalizability and minimizing bias.

Despite promising results, this systematic review has several limitations. First, the included studies exhibited considerable heterogeneity in terms of AI model architectures (e.g., CNN-based models like ResNet-50, transformer-based models such as DETR and VTN), dataset sizes (ranging from 25 to 619 videos), surgical approaches (TAPP, TEP, RIHR), and annotation protocols (manual vs. semi-automated tools such as CVAT and SuperAnnotate). This heterogeneity extended to evaluation metrics, with accuracy ranging from 74% to over 87%, and F1 scores from 65% to 82%, with some studies lacking complete reporting of precision and recall. Such variability not only precluded meta-analysis but also limited the comparability and generalizability of model performance across surgical contexts.

Second, internal validation was lacking or inconsistently reported in several studies. The absence of rigorous methods such as k-fold cross-validation or independent test sets raises concerns regarding overfitting and diminishes the robustness of the reported outcomes. Third, most studies were observational and retrospective in design, introducing selection bias and limiting the applicability of findings to real-time, prospective surgical workflows. Fourth, surgeons' experience, both in performing the procedures and in annotating surgical videos, was either underreported or not described in most studies. This omission may influence the quality of annotations and affect model training and reliability, especially for complex or skill-dependent phases. Finally, manual annotation is labor-intensive and time-consuming, and inconsistencies in labeling standards, phase definitions, and inter-annotator agreement can negatively impact model reproducibility. Lastly, institutional variability in operative techniques, equipment, and surgical team experience further limits the external validity and clinical scalability of the current AI models reviewed.

Future research should prioritize the development of standardized, publicly available, multi-institutional datasets with consistent surgical phase definitions and annotation protocols. This would enable robust benchmarking of AI models across institutions and procedural techniques. Establishing open repositories for annotated surgical videos, including TAPP, TEP, and RIHR procedures, could facilitate reproducibility

and fairness in model comparison. In addition, integrating quantitative measures of heterogeneity, such as  $I^2$  statistics,  $\tau^2$  values, and the coefficient of variation of reported accuracies into future reviews could help identify underlying sources of inconsistency and guide model selection.

Advanced training strategies such as federated learning and self-supervised approaches may reduce the dependence on manually labeled data while ensuring data privacy across institutions. Further research should focus on the development of real-time AI systems with low-latency performance using edge computing hardware such as Nvidia Jetson or A100 GPUs. These systems should be validated prospectively in real-world, multicenter clinical settings and evaluated not only for model accuracy but for their impact on surgical workflow, efficiency, safety, and outcomes.

Equally important is the pursuit of explainable AI to foster trust and transparency. Future models should incorporate interpretable outputs, including heatmaps, attention weights, and confidence scores, to assist surgical decision making and training. Regulatory pathways, including alignment with FDA and international guidelines on software as a medical device, must be actively considered during development and deployment. Interdisciplinary collaboration among surgeons, engineers, regulatory authorities, and ethicists will be crucial in ensuring safe, effective, and equitable integration of AI into operative practice.

From our perspective, this systematic review represents a foundational step toward shaping the future of AI in surgical education and operative care. We envision that real-time AI-assisted surgical phase recognition will become an essential tool in residency training, competency evaluation, and intraoperative decision support. For early-career robotic surgeons and general surgery trainees, such systems could offer objective, adaptive feedback that enhances skill acquisition. The future of surgical AI lies not only in advancing algorithmic performance but also in establishing the human-AI partnership at the core of modern surgical practice. With early, coordinated investment in algorithm development, dataset curation, and prospective validation, we believe this vision can be realized.

In conclusion, This systematic review underscores the transformative potential of AI and ML in advancing real-time surgical phase recognition during minimally invasive IHR. While significant progress has been achieved across laparoscopic and robotic procedures, the application of AI in IHR remains in its early stages. The included studies demonstrate encouraging performance metrics, with models like TeCNO, ResNet-50, and Video Transformer Networks achieving promising accuracies. However, challenges such as heterogeneity in model architectures, reliance on labor-intensive annotated datasets, and variability in surgical techniques must be addressed to enhance the clinical utility of AI. Future research should focus on developing standardized benchmarks, conducting prospective multicenter validations, and leveraging hardware advancements to enable latency-free, real-time applications. Collaborative efforts among surgeons, AI developers, and policymakers will be pivotal to overcoming existing limitations, ensuring ethical implementation, and integrating AI-driven systems into surgical workflows. By doing so, AI has the potential to revolutionize surgical training, improve technical performance, and ultimately enhance patient outcomes in IHR and beyond.

## DECLARATIONS

### Authors' contributions

Conceptualization and study design: Goyal A, Oviedo RJ

Data acquisition and curation: Goyal A, Mendoza M, Munoz AE, Macias CA

Statistical analysis and data interpretation: Goyal A, Mendoza M, Munoz AE, Macias CA

Manuscript drafting, data visualization: Goyal A, Mendoza M, Munoz AE, Macias CA

Manuscript editing, revision and formatting: Abou-Mrad A, Marano L, Oviedo RJ

Systematic literature review and methodology: Goyal, A, Mendoza M, Munoz AE, Macias CA

Critical review, supervision, and manuscript refinement: Abou-Mrad A, Marano L, Oviedo RJ

### Availability of data and materials

The datasets generated and/or analyzed during this study are not publicly available but may be obtained from the corresponding author upon reasonable request.

### 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.

### Copyright

© The Author(s) 2025.

## REFERENCES

1. Shortliffe EH. Mycin: a knowledge-based computer program applied to infectious diseases. In: Proceedings of the Annual Symposium on Computer Application in Medical Care; 1977 Oct 3-5; Washington, D.C., USA. pp. 66-9. [PMC](#)
2. Kuperman GJ, Gardner RM, Pryor TA. HELP: a dynamic hospital information system. New York: Springer Science & Business Media; 2013. Available from: [https://books.google.com/books?hl=zh-CN&lr=&id=T1fSBwAAQBAJ&oi=fnd&pg=PR7&dq=HELP:+a+dynamic+hospital+information+system&ots=BqlXJePDPo&sig=vh15EtzBbZJMGQJU6dzph\\_3UkD4#v=onepage&q=HELP%3A%20a%20dynamic%20hospital%20information%20system&f=false](https://books.google.com/books?hl=zh-CN&lr=&id=T1fSBwAAQBAJ&oi=fnd&pg=PR7&dq=HELP:+a+dynamic+hospital+information+system&ots=BqlXJePDPo&sig=vh15EtzBbZJMGQJU6dzph_3UkD4#v=onepage&q=HELP%3A%20a%20dynamic%20hospital%20information%20system&f=false) [Last accessed 12 September 2025].
3. Spinelli A, Carrano FM, Laino ME, et al. Artificial intelligence in colorectal surgery: an AI-powered systematic review. *Tech Coloproctol*. 2023;27:615-29. [DOI](#) [PubMed](#)
4. Ouyang D, Theurer J, Stein NR, et al. Electrocardiographic deep learning for predicting post-procedural mortality: a model development and validation study. *Lancet Digit Health*. 2024;6:e70-8. [DOI](#) [PubMed](#)
5. Le MH, Le TT, Tran PP. AI in surgery: navigating trends and managerial implications through bibliometric and text mining odyssey. *Surg Innov*. 2024;31:630-45. [DOI](#) [PubMed](#)
6. Burcharth J, Pedersen M, Bisgaard T, Pedersen C, Rosenberg J, Burney RE. Nationwide prevalence of groin hernia repair. *PLoS ONE*. 2013;8:e54367. [DOI](#) [PubMed](#) [PMC](#)
7. Zygomalas A, Kalles D, Katsiakakis N, Anastasopoulos A, Skroubis G. Artificial intelligence assisted recognition of anatomical landmarks and laparoscopic instruments in transabdominal preperitoneal inguinal hernia repair. *Surg Innov*. 2024;31:178-84. [DOI](#) [PubMed](#)
8. Takeuchi M, Collins T, Ndagijimana A, et al. Automatic surgical phase recognition in laparoscopic inguinal hernia repair with artificial intelligence. *Hernia*. 2022;26:1669-78. [DOI](#) [PubMed](#)
9. Zang C, Turkcan MK, Narasimhan S, et al. Surgical phase recognition in inguinal hernia repair - AI-based confirmatory baseline and exploration of competitive models. *Bioengineering*. 2023;10:654. [DOI](#) [PubMed](#) [PMC](#)
10. Ortenzi M, Rapoport Ferman J, Antolin A, et al. A novel high accuracy model for automatic surgical workflow recognition using artificial intelligence in laparoscopic totally extraperitoneal inguinal hernia repair (TEP). *Surg Endosc*. 2023;37:8818-28. [DOI](#) [PubMed](#) [PMC](#)
11. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. [DOI](#) [PubMed](#) [PMC](#)
12. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of case-control studies in meta-analyses. *Eur J Epidemiol*. 2011;25:603-5. Available from: [https://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp). [Last accessed on 12 Sep 2025]



13. Cui P, Zhao S, Chen W, Peng J. Identification of the vas deferens in laparoscopic inguinal hernia repair surgery using the convolutional neural network. *J Healthc Eng*. 2021;2021:1-10. DOI PubMed PMC
14. Taha A, Enodien B, Frey DM, Taha-mehlitz S. The development of artificial intelligence in hernia surgery: a scoping review. *Front Surg*. 2022;9:908014. DOI PubMed PMC
15. Elfanagely O, Mellia JA, Othman S, Basta MN, Mauch JT, Fischer JP. Computed tomography image analysis in abdominal wall reconstruction: a systematic review. *Plast Reconstr Surg Glob Open*. 2020;8:e3307. DOI PubMed PMC
16. Elhage SA, Deerenberg EB, Ayuso SA, et al. Development and validation of image-based deep learning models to predict surgical complexity and complications in abdominal wall reconstruction. *JAMA Surg*. 2021;156:933. DOI PubMed PMC
17. Niebuhr H, Born O. Image tracking system eine neue Technik für die sichere und kostensparende laparoskopische operation. *Chirurg*. 2000;71:580-4. DOI PubMed
18. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: an R package and Shiny app for producing PRISMA 2020 compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Syst Rev*. 2022;18:e1230. DOI PubMed PMC
19. Moher D, Shamseer L, Clarke M, et al. PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1. DOI PubMed PMC
20. Takeuchi M, Collins T, Lipps C, et al. Towards automatic verification of the critical view of the myopectineal orifice with artificial intelligence. *Surg Endosc*. 2023;37:4525-34. DOI PubMed
21. Choksi S, Szot S, Zang C, et al. Bringing artificial intelligence to the operating room: edge computing for real-time surgical phase recognition. *Surg Endosc*. 2023;37:8778-84. DOI PubMed
22. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med*. 2020;3:126. DOI PubMed PMC
23. Lima DL, Kasakewitch J, Nguyen DQ, et al. Machine learning, deep learning and hernia surgery. Are we pushing the limits of abdominal core health? *Hernia*. 2024;28:1405-12. DOI PubMed
24. Gagandeep K, Rishabh M, Vyas S. Artificial intelligence (AI) startups in health sector in India: challenges and regulation in India. In: Goyal D, Kumar A, Piuri V, Paprzycki M, Editors. Proceedings of the Third International Conference on Information Management and Machine Intelligence; 2021 Dec 23-24; Jaipur, India. Singapore: Springer; 2023. pp. 203-15. DOI
25. Pharmaphorum. Sensely and Mayo Clinic take virtual nurse one step further. Available from: <https://pharmaphorum.com/news/sensely-mayo-clinic-develop-virtual-doctor>. [Last accessed on 12 Sep 2025].
26. Pharmaphorum. FDA approves Caption Health's AI-driven cardiac ultrasound software. Available from: <https://pharmaphorum.com/news/fda-approves-caption-healths-ai-driven-cardiac-imaging-software>. [Last accessed on 12 Sep 2025].
27. Hashimoto DA, Rosman G, Witkowski ER, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg*. 2019;270:414-21. DOI PubMed PMC
28. Zhang B, Ghanem A, Simes A, Choi H, Yoo A. Surgical workflow recognition with 3DCNN for sleeve gastrectomy. *Int J Comput Assist Radiol Surg*. 2021;16:2029-36. DOI PubMed PMC
29. 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*. 2019;34:4924-31. DOI PubMed
30. Ward TM, Hashimoto DA, Ban Y, et al. Automated operative phase identification in peroral endoscopic myotomy. *Surg Endosc*. 2020;35:4008-15. DOI PubMed PMC
31. Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Padoy N. EndoNet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans Med Imaging*. 2016;36:86-97. DOI PubMed
32. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev*. 1959;3:210-29. DOI
33. Jin Y, Dou Q, Chen H, et al. SV-RCNet: workflow recognition from surgical videos using recurrent convolutional network. *IEEE Trans Med Imaging*. 2018;37:1114-26. DOI PubMed
34. Czempiel T, Paschali M, Keicher M, et al. TeCNO: surgical phase recognition with multi-stage temporal convolutional networks. In: Martel AL, Abolmaesumi P, Stoyanov D, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2020; 2020 Oct 4-8; Lima, Peru. Cham: Springer; 2020. pp. 343-52. DOI
35. Jin Y, Long Y, Chen C, Zhao Z, Dou Q, Heng PA. Temporal memory relation network for workflow recognition from surgical video. *IEEE Trans Med Imaging*. 2021;40:1911-23. DOI PubMed
36. Gao X, Jin Y, Long Y, Dou Q, Heng PA. Trans-SVNet: accurate phase recognition from surgical videos via hybrid embedding aggregation transformer. In: De Bruijne M, Cattin PC, Cotin S, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2021. 2021 Sep 27-Oct 1; Strasbourg, France. Cham: Springer; 2021. pp. 593-603. DOI
37. Park M, Oh S, Jeong T, Yu S. Multi-stage temporal convolutional network with moment loss and positional encoding for surgical phase recognition. *Diagnostics*. 2022;13:107. DOI PubMed PMC