AMERICAN JOURNAL OF MEDICAL AND CLINICAL SCIENCES, 2026
VOL 11, NO. 4, PAGE 1 – 5

RESEARCH ARTICLE

Open Access

Deep Learning Architectures for Surgical Intelligence in Minimally Invasive Gynecological Surgery: A Scoping Review

Mohamed Abdelrahman[1,2,3]*, Rawia Ahmed[1,2], Mohamed Elshaikh[1] ,Elmuiz Haggaz[1,2] and Simon Colreavy[3]

[1]Royal College of Physicians of Ireland.

[2]Limerick Hospital Groups.

[3]University of Limerick.

ABSTRACT

Background: The application of Deep Learning (DL) in Minimally Invasive Gynecological Surgery (MIGS) has accelerated rapidly, driven by the need for enhanced intra-operative navigation, surgical phase recognition, and predictive analytics. The efficacy of these applications is fundamentally determined by the underlying neural network architectures, yet a comprehensive technical synthesis of their comparative performance and clinical utility in the gynecological context is absent from the literature.

Objective: To provide a comprehensive scoping review of the DL architectures deployed in MIGS, evaluating their technical mechanisms, clinical applications, performance metrics, and limitations.

Methods: A scoping review was conducted following PRISMA-ScR guidelines, drawing from a systematic search of 55 included studies identified across five major databases. Studies explicitly detailing neural network architectures applied to gynecological surgical tasks were analyzed, with a focus on Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, Deep Neural Networks (DNNs), Transformer architectures, and hybrid models.

Results: CNNs remain the foundational architecture for spatial feature extraction in surgical video analysis, achieving accuracy rates of 85–97% in anatomical landmark identification and instrument detection. LSTM networks, frequently integrated with CNNs in hybrid architectures, significantly improve surgical phase recognition by capturing temporal dependencies across video sequences. Transformer architectures, including Vision Transformers (ViTs), have demonstrated superior performance in complex multi-modal tasks, outperforming CNN-LSTM models in surgical workflow understanding. Emerging applications include augmented reality navigation, automated endometriosis lesion recognition, and AI-driven prediction of perioperative outcomes.

Conclusion: The DL landscape in MIGS is evolving from static image analysis toward dynamic, spatio-temporal surgical understanding. While CNNs and LSTMs constitute the current standard, Transformers represent the emerging frontier. Critical priorities for future development include model interpretability, real-time edge deployment, and standardized benchmarking across diverse gynecological datasets.

ARTICLE HISTORY

Received 28 June 2026
Accepted 05 August 2026
Published 16 August 2026

KEYWORDS

Deep Learning, Convolutional Neural Networks, Transformers; LSTM, Surgical Phase Recognition, Minimally Invasive Gynecological Surgery, Computer Vision, Augmented Reality.

Introduction

The operating room of the 21st century is increasingly characterized by the convergence of advanced robotics, high-definition imaging, and Artificial Intelligence (AI). Within Minimally Invasive Gynecological Surgery (MIGS), this convergence has catalyzed a paradigm shift: from passive surgical tools to active, intelligent systems capable of perceiving, interpreting, and responding to the dynamic surgical environment in real time [1,2]. The rich data streams generated by modern laparoscopic and robotic platforms including high-definition video, kinematic data from robotic arms, and physiological monitoring provide an unprecedented substrate for AI-driven analysis.

At the core of this transformation lies Deep Learning (DL), a subset of Machine Learning (ML) based on multi-layered artificial neural networks capable of learning complex, hierarchical representations from raw data. Since the landmark demonstration of CNN superiority in the ImageNet challenge in 2012, DL has rapidly become the dominant paradigm for medical image analysis, and its application to surgical video has followed naturally [3]. However, the surgical domain presents unique challenges that distinguish it from standard computer vision tasks: the visual environment is dynamic, blood and tissue deformation alter scene appearance continuously, and the consequences of algorithmic errors are potentially life-threatening.

Contact: Mohamed Abdelrahman, Royal College of Physicians of Ireland, Phone: 00353873987464.

© 2026 The Authors. This is an open access article under the terms of the Creative Commons Attribution NonCommercial ShareAlike 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/).

Despite the proliferation of DL applications in MIGS, a systematic technical synthesis of the architectures employed their mechanisms, comparative performance, and specific clinical utility in gynecological surgery is lacking. This scoping review addresses this gap by mapping the DL architectural landscape in MIGS, drawing on evidence from 55 included studies identified through a comprehensive systematic search. The review is structured around the four principal DL architectures deployed in this domain: CNNs, LSTM networks, DNNs, and Transformers, with attention to hybrid models and emerging applications.

Methods

This scoping review was conducted in accordance with the PRISMA Extension for Scoping Reviews (PRISMA-ScR) guidelines. The evidence base was drawn from a comprehensive systematic search conducted across PubMed/MEDLINE, Embase, Scopus, Web of Science, and the Cochrane Library, yielding 55 included studies after screening 4,196 initial records and removing 639 duplicates. Studies were included in this technical analysis if they explicitly described the neural network architecture employed and reported quantitative performance metrics (e.g., accuracy, sensitivity, specificity, F1-score, mean Average Precision). Data on architecture type, training dataset characteristics, task type, performance metrics, and clinical application were extracted and synthesized narratively, with tabular summaries provided for key findings.

Deep Learning Architectures in MIGS: A Technical Review

Convolutional Neural Networks (CNNs)

Architecture and Mechanism

Convolutional Neural Networks (CNNs) are the cornerstone of computer vision in surgery. Designed to process grid-structured data such as images, CNNs excel at hierarchical spatial feature extraction. The architecture consists of alternating convolutional layers which apply learnable filters to generate feature maps capturing edges, textures, and progressively complex anatomical patterns and pooling layers, which reduce spatial dimensions and confer translation invariance. The final fully connected layers perform classification or regression based on the extracted feature representations.

Applications in Gynecological MIGS

CNNs are the most widely deployed DL architecture in MIGS (24 of 55 included studies, 43.6%). Their primary applications include:
(1) anatomical structure segmentation and identification.
(2) surgical instrument detection and tracking.
(3) endometriosis lesion recognition.
(4) tissue classification.

Paracchini et al. [4] conducted a systematic review of AI models for surgical phase, instrument, and anatomical structure identification in the operating room, demonstrating that CNN-based models achieved mean accuracy rates of 85–97% across these tasks in gynecological laparoscopy. Gkrozou et al. [5] similarly reported high performance of CNN variants (ResNet, VGG, YOLO) in laparoscopic gynecological applications, with YOLO-based real-time instrument detection achieving frame rates compatible with intra-operative use (>25 fps).

A particularly impactful application is the automated visual recognition of endometriosis lesions. Chen et al. [6] demonstrated the feasibility of CNN-based AI for automatic recognition of endometriotic lesions during laparoscopy, reporting sensitivity and specificity values of 87.3% and 91.2%, respectively performance approaching that of experienced endometriosis surgeons. Similarly, Liu et al. [7] applied CNN-based instance segmentation to MRI data for preoperative planning of laparoscopic myomectomy for broad ligament fibroids, demonstrating improved surgical efficiency and safety.

Long Short-Term Memory (LSTM) Networks

Architecture and Mechanism

While CNNs excel at spatial analysis of individual frames, surgical procedures are inherently sequential processes with critical temporal dependencies. Long Short-Term Memory (LSTM) networks, a specialized form of Recurrent Neural Network (RNN), were designed to address the vanishing gradient problem that prevents standard RNNs from learning long-range temporal dependencies. The LSTM memory cell is regulated by three gates: the forget gate (determining which historical information to discard), the input gate (controlling new information incorporation), and the output gate (regulating the information passed to subsequent time steps). This gating mechanism enables LSTMs to maintain relevant context over extended surgical sequences while filtering irrelevant information.

Applications in Gynecological MIGS

LSTMs are most commonly deployed in hybrid CNN-LSTM architectures for surgical phase recognition and workflow analysis. The CNN component extracts spatial features from individual video frames, which are then processed sequentially by the LSTM to model the temporal progression of the procedure. Liao et al. [8] conducted a systematic review of AI-assisted phase recognition and skill assessment in laparoscopic surgery, identifying CNN-LSTM architectures as the dominant approach, with phase recognition accuracy ranging from 78% to 94% across studies. Levin et al. [9] applied a surgical intelligence system incorporating temporal modeling to automatically identify key steps in laparoscopic hysterectomy, demonstrating high agreement with expert annotation and revealing clinically relevant insights into time allocation during procedures.

Walczak and Mikhail [10] demonstrated a distinct application of LSTM-based neural networks for predicting estimated blood loss and transfusion requirements in gynecological surgery from preoperative patient data, achieving predictive accuracy superior to traditional logistic regression models. This application exemplifies the versatility of temporal modeling architectures beyond video analysis.

Deep Neural Networks (DNNs) for Predictive Modeling

Architecture and Applications

Deep Neural Networks (DNNs)—feedforward networks with multiple hidden layers are the predominant architecture for tabular and structured clinical data analysis in MIGS. Unlike CNNs and LSTMs, which are optimized for spatial and temporal data respectively, DNNs are well-suited to learning complex non-linear relationships in high-dimensional clinical datasets comprising patient demographics, laboratory values, imaging characteristics, and surgical parameters.

In the MIGS context, DNNs have been applied primarily to:
1. Preoperative risk stratification and complication prediction.
2. Prediction of surgical outcomes (operative time, blood loss, conversion to open surgery).
3. Postoperative outcome modeling.

Bar-El et al. [11] developed a DNN-based predictive model for perioperative outcomes in endometriosis surgery, demonstrating superior discrimination compared to traditional scoring systems. Snyder et al. [12] employed ML algorithms including gradient boosting and DNN variants to develop a preoperative endometriosis prediction model from clinical variables, achieving an area under the receiver operating characteristic curve (AUC-ROC) of 0.84. Chen et al. [13] developed a machine learning-based model for prediction of deep vein thrombosis after gynecological laparoscopy, demonstrating the breadth of DNN applications in perioperative risk management.

Transformer Architectures

Architecture and Mechanism

The Transformer architecture, introduced by Vaswani et al. in 2017 for natural language processing, has emerged as a transformative paradigm in computer vision following the development of Vision Transformers (ViTs) and their surgical adaptations. Unlike CNNs, which process images through local convolutional operations, Transformers employ self-attention mechanisms that model global relationships across the entire input simultaneously. The multi-head self-attention mechanism computes attention weights for all pairwise relationships between input tokens (image patches or video frames), enabling the model to capture both local detail and long-range context without the inductive biases inherent to convolution.

Emerging Applications in MIGS

While Transformer applications in MIGS remain nascent compared to CNNs and LSTMs, their adoption is accelerating. Spatio-temporal Transformers have demonstrated superior performance to CNN-LSTM models in complex surgical workflow understanding tasks, particularly where long-range temporal dependencies and multi-modal data integration are required. Collins et al. [14] developed an augmented reality guided laparoscopic surgery system for the uterus employing advanced spatial modeling techniques, demonstrating the feasibility of real-time AR overlay in gynecological laparoscopy a task that benefits from the global context modeling capacity of Transformer-based approaches.

The interpretability advantage of Transformers is particularly relevant in the surgical context: attention maps generated by Transformer models can visually highlight the specific regions of the surgical field that the model focuses on when making predictions, providing a form of explainability that is critical for clinical acceptance. Nwoye et al. [15] reviewed AI for emerging technology in surgery, noting the growing adoption of Transformer-based architectures for surgical scene understanding tasks, with performance metrics consistently exceeding those of CNN-based predecessors on benchmark datasets.

Hybrid and Multimodal Architectures

The most clinically impactful DL systems in MIGS are increasingly hybrid architectures that combine the complementary strengths of multiple network types. The CNN-LSTM combination for surgical phase recognition exemplifies this approach. More recently, architectures integrating CNNs with Transformer encoders (CNN-Transformer hybrids) have demonstrated state-of-the-art performance on surgical video benchmarks, leveraging CNNs for efficient local feature extraction and Transformers for global temporal context modeling. Saadati and Amirma zlaghani [16] developed a system for automated surgical operation in endometriosis treatment through AI and robotic vision, employing a hybrid architecture that combined convolutional feature extraction with sequential decision-making, demonstrating the feasibility of semi-autonomous surgical task execution.

Discussion

This scoping review maps the DL architectural landscape in MIGS, revealing a field in rapid transition. The dominance of CNN-based approaches reflects both their proven efficacy for spatial image analysis and the availability of pre-trained models (e.g., ResNet, VGG, EfficientNet) that can be fine-tuned on relatively small surgical datasets through transfer learning. However, the limitations of CNNs particularly their inability to natively model temporal dynamics—have driven the widespread adoption of CNN-LSTM hybrid architectures for the fundamentally sequential task of surgical phase recognition.

The emergence of Transformer architectures represents the most significant recent development in surgical AI. Their capacity for global context modeling and multi-modal data integration positions them as the natural successor to CNN-

Table 1: Comparative Summary of Deep Learning Architectures in MIGS.

Architecture	Primary Strength	Main MIGS Application	Typical Accuracy	Key Limitation
CNN	Spatial feature extraction	Instrument detection, anatomical segmentation, lesion recognition	85–97%	No temporal modeling; static frame analysis
CNN-LSTM (Hybrid)	Spatio-temporal modeling	Surgical phase recognition, workflow analysis	78–94%	High computational cost; sequential processing
DNN	Tabular data modeling	Complication prediction, outcome modeling, risk stratification	AUC 0.78–0.91	Requires large structured datasets
Transformer / ViT	Global context, multi-modal fusion	Workflow understanding, AR navigation, multi-modal analysis	88–96%*	High data requirements; computationally intensive

*Transformer accuracy figures represent emerging benchmark results; clinical validation in MIGS is ongoing. CNN = Convolutional Neural Network; LSTM = Long Short-Term Memory; DNN = Deep Neural Network; ViT = Vision Transformer; AUC = Area Under the ROC Curve.

LSTM models for complex surgical understanding tasks. However, their substantially higher data and computational requirements present practical barriers to clinical deployment, particularly in resource-limited settings. The development of efficient Transformer variants (e.g., Swin Transformer, DeiT) and knowledge distillation techniques for model compression will be critical for translating these architectural advances into clinical practice.

A critical challenge across all DL architectures in MIGS is the scarcity of large, annotated surgical datasets. The manual annotation of surgical video is extraordinarily time-consuming, requiring expert surgeons to label each frame with anatomical structures, instrument types, and procedural phases. This data bottleneck has driven interest in self-supervised and semi-supervised learning approaches, which can leverage large volumes of unannotated surgical video to pre-train models that are subsequently fine-tuned with limited labeled data. The development of shared, multi-institutional surgical video datasets analogous to ImageNet for general computer vision is an urgent priority for the field [17].

Model interpretability and explainability represent another critical dimension for clinical adoption. The "black box" nature of deep neural networks has been a persistent barrier to surgeon acceptance and regulatory approval. Techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) for CNNs and attention visualization for Transformers provide visual explanations of model predictions, but their integration into clinical workflows remains limited. Future DL systems for MIGS must be designed with interpretability as a first-class requirement, not an afterthought [18].

Conclusion

Deep Learning has transformed the technical capabilities available to gynecological surgeons, enabling real-time anatomical guidance, automated workflow analysis, and data-driven outcome prediction. CNNs and CNN-LSTM hybrids constitute the current standard for spatial and spatio-temporal surgical analysis, respectively, while Transformer architectures represent the emerging frontier for complex multi-modal surgical understanding. The continued advancement of DL in MIGS requires coordinated progress on four fronts: the development of large, annotated surgical datasets; the creation of lightweight, real-time-capable architectures for edge deployment; the integration of interpretability mechanisms for clinical acceptance; and the rigorous prospective validation of AI systems against patient-centered clinical outcomes. The technical foundations are in place; the imperative now is translation.

Acknowledgements

Main author who did write this article would like to thank Dr Simon Colreavy my Health informatics MSc Supervisor, Dr. Rawia Ahmed, Dr. Elmuiz Haggaz and Dr. Mohamed Elshaikh, the reviewers of the paper before submission.

References

1. Pipes GM, Hung AJ, Kim KH. Artificial Intelligence and Gynecologic Surgery. Obstet Gynecol. 2025; 146(5): 672-678.
2. Pavone M, Goglia M, Rosati A, Innocenzi C, Bizzarri N, et al. Unveiling the real benefits of robot-assisted surgery in gynaecology: from telesurgery to image-guided surgery and artificial intelligence. Facts Views Vis Obgyn. 2025; 17(1): 50-60.
3. Hager G, Stoyanov D, Taylor RH, Maier-Hein L, Speidel S, et al. Deep learning for surgical computer vision. Nat Rev Methods Primers. 2023; 3(1): 18.
4. Paracchini S, Taliento C, Pellecchia G, Tius V, Tavares M, et al. Artificial intelligence in the operating room: A systematic review of AI models for surgical phase, instruments and anatomical structure identification. Acta Obstet Gynecol Scand. 2025; 104(11): 2054-2064.
5. Gkrozou F, Bais V, Skentou C, Kalaitzopoulos DR, Grigoriadis G, et al. Applications of Deep Learning Models in Laparoscopy for Gynecology. Medicina (Kaunas). 2025; 61(8): 1460.
6. Chen S, Song J, Sun Y, Xiao J. Initial Results in the Automatic Visual Recognition of Endometriosis Lesions by Artificial Intelligence During Laparoscopy: A Proof-of-Concept Study. J Minim Invasive Gynecol. 2026.
7. Liu F, Chen M, Pan H, Li B, Bai W. Artificial intelligence for instance segmentation of MRI: advancing efficiency and safety in laparoscopic myomectomy of broad ligament fibroids. Front Oncol. 2025.
8. Liao W, Zhu Y, Zhang H, Wang D, Zhang L, et al. Artificial intelligence-assisted phase recognition and skill assessment in laparoscopic surgery: a systematic review. Front Surg. 2025.
9. Levin I, Ferman JR, Bar O, Ayoun DB, Cohen A, et al. Introducing surgical intelligence in gynecology: Automated identification of key steps in hysterectomy. Int J Gynecol Obstet. 2024; 166(3): 1273-1278.
10. Walczak S, Mikhail E. Predicting Estimated Blood Loss and Transfusions in Gynecologic Surgery Using Artificial Neural Networks. IJHISI. 2021.
11. Bar-El L, Albahra S, Kumar SV, Ockunzzi S, Chin A, et al. Harnessing Artificial Intelligence to Personalize Endometriosis Surgery: Development of a Predictive Model for Perioperative Outcomes. JMIG. 2025; 32(11): 69-70.
12. Snyder DL, Sidhom S, Chatham CE, Tillotson SG, Zapata RD, et al. Utilizing Artificial Intelligence: Machine Learning Algorithms to Develop a Preoperative Endometriosis Prediction Model. JMIG. 2025; 32(9): 784-792.
13. Chen X, Hou M, Wang DX. Machine learning-based model for prediction of deep vein thrombosis after gynecological laparoscopy: A retrospective cohort study. Medicine. 2024; 103(1): 36717.
14. Collins T, Pizarro D, Gasparini S, Bourdel N, Chauvet P, et al. Augmented Reality Guided Laparoscopic Surgery of the Uterus. IEEE Trans Med Imaging. 2021; 40(1): 371-380.
15. Nwoye E, Woo WL, Gao B, Anyanwu T. Artificial Intelligence for Emerging Technology in Surgery: Systematic Review and Validation. IEEE Rev Biomed Eng. 2023; 16: 241-259.

16. Saadati S, Amirmazlaghani M. Revolutionizing endometriosis treatment: automated surgical operation through artificial intelligence and robotic vision. J Robot Surg. 2024; 18(1): 383.
17. Maier-Hein L, Eisenmann M, Sarikaya D, März K, Collins T, et al. Surgical data science – from concepts toward clinical translation. Med Image Anal. 2022; 76: 102306.
18. Collins JW, Marcus HJ, Ghazi A, Sridhar A, Hashimoto D, et al. Ethical implications of AI in robotic surgical training: A Delphi consensus statement. Eur Urol Focus. 2022; 8(2): 613-622.