Computer Vision in Laparoscopic Surgery: Integration of Engineering Architecture and Clinical Decision-Making

Ulvi Rafizade [1[0009-0009-6005-897X]], Irada Imanova[2[0009-0001-6737-8652]] ,

Khagan Mammadli [3[0009-0004-8497-655X]]

[1] Azerbajan Technological University, SI Khatai Avenue 103 AZ2011, Ganja, Azerbaijan
u.refizade@atu.edu.az
[2] Odlar Yurdu University, K. Rahimov Str. 13 AZ1072, Baku, Azerbaijan
irada.imanova@oyu.edu.az
[3] Central Customs Hospital, K. K.Kazim zade str. 118 AZ1065, Baku, Azerbaijan
xaqan.mammadli@customshospital.az

Abstract. Laparoscopic surgery, a minimally invasive method utilizing fiber optic cameras, is central to the digital transformation of medicine. This article explores computer vision (CV) algorithms for workflow recognition, anatomical segmentation, and ergonomic monitoring. From an engineering perspective, the effectiveness of architectures such as 3D Convolutional Neural Networks (3DCNN), YOLOv9, and U-Net is analyzed for their ability to process spatio-temporal data and identify tissue at a pixel level. From a clinical standpoint, we discuss how these models prevent iatrogenic bile duct injuries (BDI) by automating the assessment of the Critical View of Safety (CVS).CV also addresses the 74% prevalence of musculoskeletal disorders among surgeons through real-time postural analysis. Our analysis demonstrates that AI systems can distinguish between novices and experts with an F1 score of 0.91 and identify operative steps with up to 85.6% accuracy. This synthesis of clinical and engineering judgment provides a robust framework for improving safety. Integration of vision technologies ensures quantifiable intraoperative precision.

Keywords: Computer Vision, Laparoscopic Surgery, Artificial Intelligence, Critical View of Safety, Explainable AI, Surgeon Ergonomics.

1 Introduction: Digitalization of Surgery

Laparoscopic surgery, as a cornerstone of minimally invasive operative techniques, requires surgeons to navigate the complex three-dimensional abdominal cavity using 2D monitor images. This transition from open surgery introduces significant challenges, most notably the loss of direct haptic (tactile) feedback, which limits the surgeon’s ability to palpate tissues or detect subtle anatomical anomalies by touch. In this digital environment, fiber-optic cameras serve a dual role: they function as the surgeon's essential "eyes" while simultaneously acting as high-fidelity sensors for artificial intelligence systems .From an engineering perspective, these intraoperative video streams represent an exceptionally rich data source, containing up to 25 times more information than high-resolution CT scans. From a clinical standpoint, this data is critical because approximately 2% of preventable medical errors occur in the operating room, often resulting from the incorrect visual perception of anatomical structures. By leveraging computer vision, engineers and clinicians aim to mitigate these risks—such as iatrogenic bile duct injuries—thereby transforming surgical video into a quantitative tool for real-time decision support and enhanced patient safety.

2 Methodology: Engineering Models and Clinical Labeling

2.1 Object detection and segmentation algorithms

Object Detection: Advanced object detection architectures, such as Faster R-CNN and YOLOv9, are instrumental in laparoscopy for localizing surgical instruments and identifying complex pathologies like endometriosis. Faster R-CNN is particularly noted for its exceptional diagnostic reliability, achieving a high test precision of 0.9811 in specialized comparative studies. While YOLOv9 excels in near real-time inference, it often exhibits higher sensitivity to class imbalances and overfitting when trained on smaller, custom datasets. Conversely, Faster R-CNN provides superior generalization and more aligned matches with medical ground-truth annotations. The implementation of these models enables precise tracking of various surgical tools, such as forceps and cauteries, facilitating better intraoperative guidance and reducing the risk of missing problematic blind spots in difficult anatomical regions

Semantic Segmentation: Semantic segmentation architectures like U-Net and DeepLabV3+ delineate anatomical structures such as the liver and gallbladder with pixel-level precision. This allows AI to generate color-coded "Go" (safe dissection zones) and "No-Go" (high-risk zones) masks. This intraoperative guidance helps surgeons avoid vital structures, significantly enhancing overall surgical safety

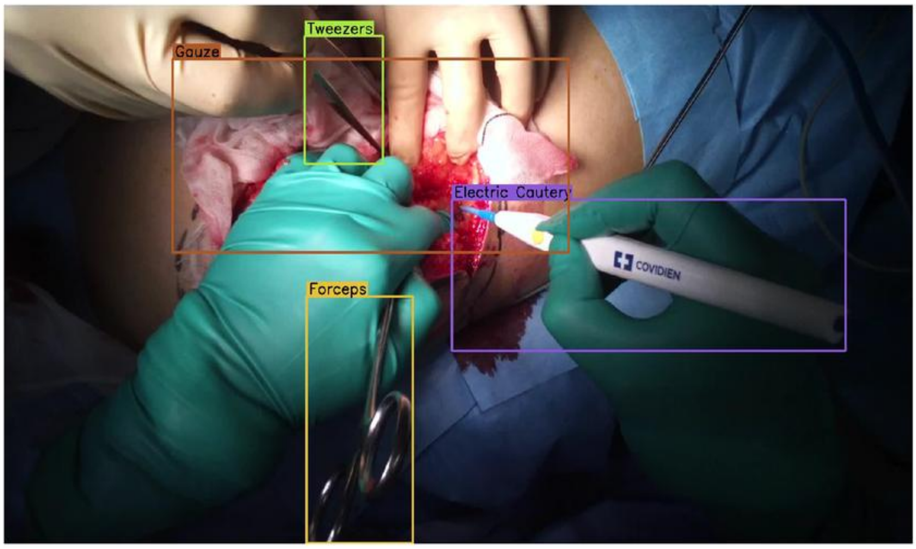


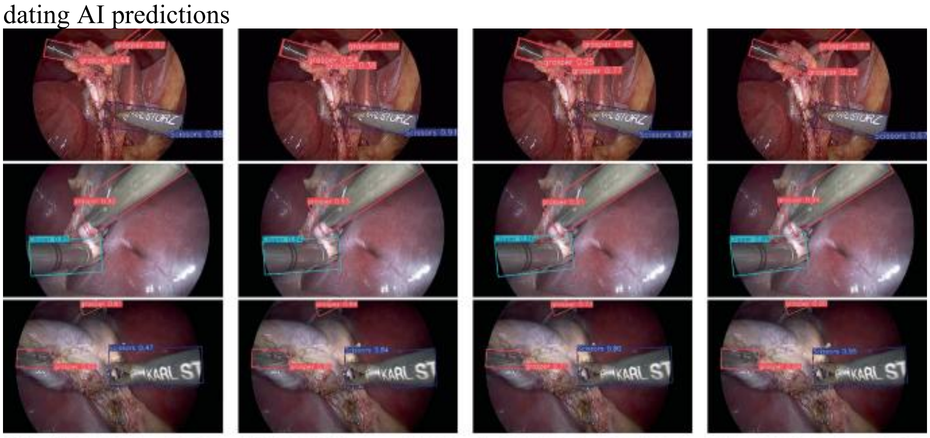
Fig. 1. Deep learning-based surgical instrument localization. YOLO-based architectures allow for near real-time tracking of multiple tools during complex maneuvers.

2.2 Spatiotemporal (Spatial-Temporal) Modeling

Surgery is a dynamic process; therefore, video sequences rather than single frames must be analyzed to capture the temporal evolution of the procedure:

CNN-LSTM Hybrids: This synthesis utilizes CNNs to extract intricate spatial features from surgical frames, while Long Short-Term Memory (LSTM) networks are integrated to learn critical temporal dependencies. This hybrid architecture significantly enhances the AI's ability to accurately identify specific surgical phases and track complex instrument trajectories over time. Such spatiotemporal modeling is essential for capturing the dynamic nature of procedures and improving overall surgical safety.

3DCNN and Weakly Supervised Learning: Utilizing specialized datasets like the Laparoscopic Surgical Performance Dataset (LSPD), engineering models such as 3D Convolutional Neural Networks (3DCNN) analyze the complex spatiotemporal coordination of instrument movements over time. This approach is instrumental in distinguishing between novice, trainee, and expert skill levels with high precision, achieving an impressive F1 score of 0.91 and an AUC of 0.92. Unlike traditional 2D models, 3DCNNs effectively capture the temporal evolution of surgical maneuvers, facilitating automated, non-subjective skill assessment. The comparative performance of various architectures, including those utilizing oriented bounding boxes (OBB) for enhanced instrument localization, is visualized in Fig. 2a. Moreover, Fig. 2b illustrates the specialized annotation interfaces used by clinicians to label intraoperative sequences—a critical but labor-intensive process—to provide the necessary "ground truth" for validating AI predictions



(a) YOLOv8s-OBB (b) YOLOv8m-OBB (c) YOLOv8l-OBB (d) YOLOv8x-OBB

Fig. 2a. Performance comparison of YOLOv8-OBB variants in surgical instrument detection.

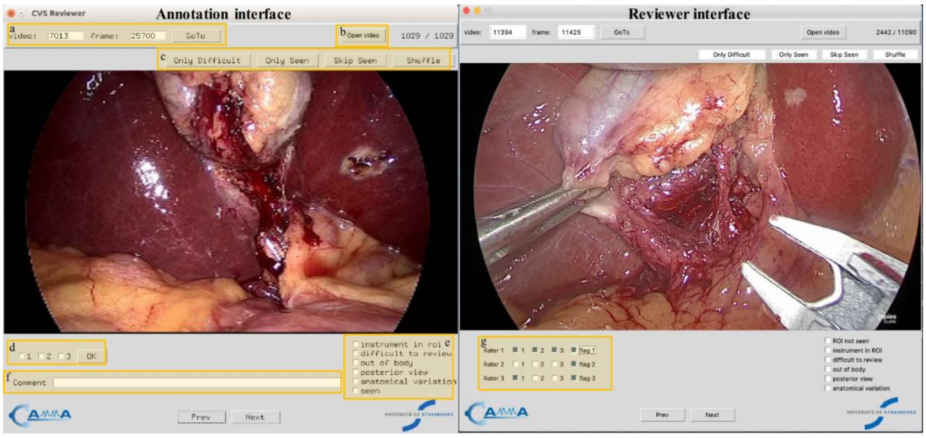


Fig. 2b. Clinical annotation and reviewer interface for intraoperative dataset validation.

3 Clinical Application and Results

3.1 Surgical Workflow and Safety

Artificial intelligence (AI) systems have emerged as transformative tools for managing surgical workflows and ensuring patient safety during laparoscopic procedures. In specialized operations such as laparoscopic sleeve gastrectomy (LSG), deep learning architectures can autonomously identify seven critical operative phases, ranging from port placement and liver retraction to stomach stapling and final inspection. Research indicates that these models achieve a maximum accuracy of 85.6% in identifying these steps, facilitating the automated generation of standardized surgical reports and optimizing overall operating room efficiency. Beyond workflow management, AI plays a pivotal role in preventing iatrogenic bile duct injuries (BDI) during laparoscopic cholecystectomy (LC). By automating the assessment of the Critical View of Safety (CVS)—which relies on identifying the cystic duct, cystic artery, and hepatocystic triangle—AI provides real-time navigational support. Systematic reviews demonstrate that these algorithms can detect biliary structures with an exceptional precision of 98%. Furthermore, evidence suggests that 70% of AI-prompted annotation shifts lead surgeons to adopt safer dissection strategies. These quantifiable insights directly address anatomical cognitive errors, which are primary causes of surgical complications, thereby establishing a new standard for intraoperative safety and quantifiable precision

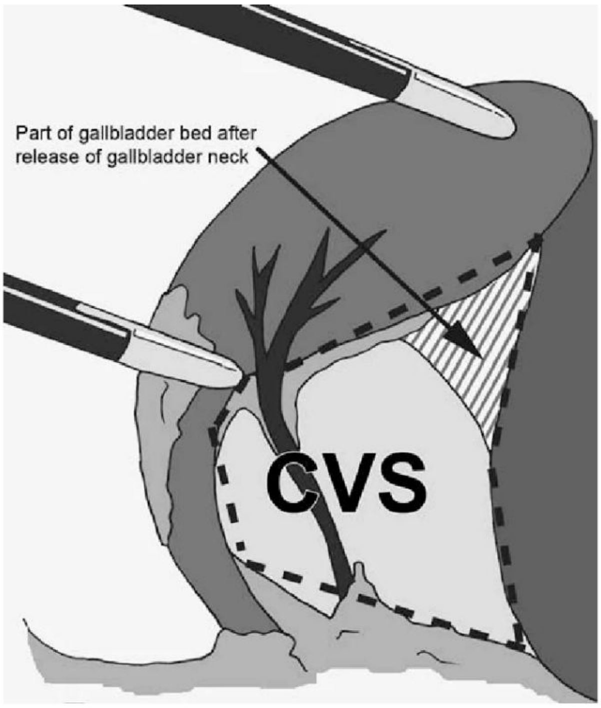


Fig. 3. AI-assisted identification of the Critical View of Safety (CVS) in laparoscopic cholecystectomy. The model provides color-coded "Go" and "No-Go" masks to assist the surgeon in safe biliary dissection.

3.2 AI-Based Monitoring of Surgeon Ergonomics

The physical well-being of the surgical team is a critical component of operative success. Laparoscopic surgery requires prolonged periods of static, non-ergonomic positioning, leading to a high prevalence of work-related musculoskeletal disorders (WMSD), reported in up to 74% of laparoscopic surgeons. Common complaints include generalized fatigue, carpal tunnel syndrome, and cervical spondylosis. To address this, computer vision (CV) software leverages real-time postural analysis to monitor surgeon ergonomics without the need for wearable sensors. By utilizing frontal and side-view imaging, AI models detect specific anatomical landmarks such as the tragus of the ear, the acromion process, and the C7 vertebra. These markers allow the software to calculate critical parameters including the craniovertebral angle (CVA) and craniohorizontal angle (CHA) in real time. Clinical studies highlight a significant concern: 93.3% of surgeons exhibit "forward head posture" (FHP) during procedures, a condition characterized by excessive strain on the cervical spine. Advanced CV systems can provide immediate "red alerts" when high-risk postures are detected, prompting surgeons to make necessary corrections. By integrating ergonomic monitoring into the digital operating room, AI helps mitigate physical fatigue and potential operative errors, ensuring the long-term professional health of the surgical workforce.

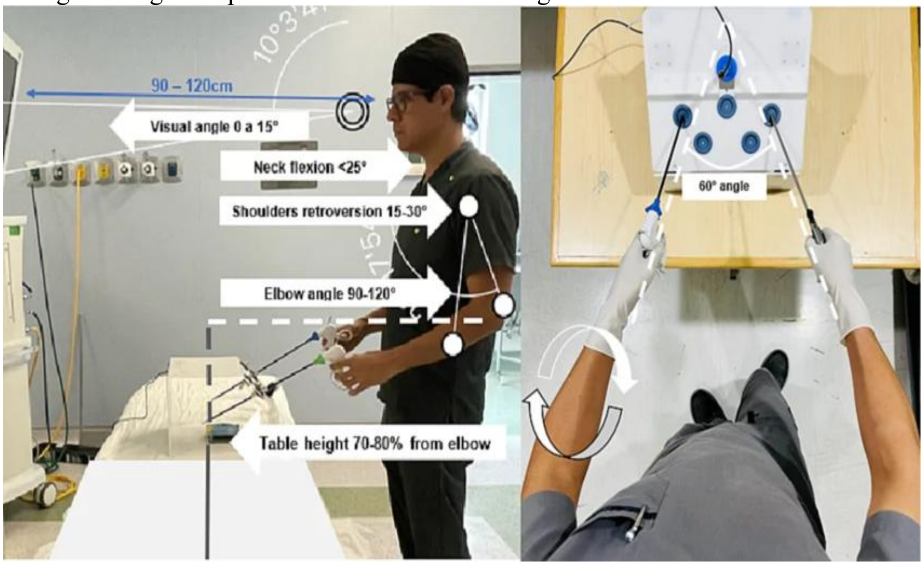


Fig. 4. Real-time monitoring of surgeon ergonomics. Computer vision software detects neck and shoulder landmarks to calculate the craniovertebral angle (CVA) and prevent musculoskeletal disorders.

3.3 Intraoperative Navigation: The "ImNavi" Trial

The "ImNavi" trial, a pioneering multicenter randomized controlled trial (RCT) conducted in Japan, represents a significant leap in intraoperative navigation using advanced computer vision. This study focuses on the clinical application of semantic segmentation, utilizing specialized algorithms known as UreterNet and NerveNet to identify the ureter and autonomic nerves—specifically the hypogastric nerves and aortic plexus—in real-time during left-sided colorectal resections. In a laparoscopic environment where surgeons navigate without haptic feedback, the risk of iatrogenic ureteral injury is estimated between 0.5% and 1.5%. Moreover, autonomic nerve damage can lead to devastating post-operative dysfunctions, including a significant decline in male erectile and ejaculatory performance.

The system functions by superimposing color-coded segmentation masks onto a sub-monitor, providing the surgical team with real-time visual confirmation without obstructing the main monitor. The trial's primary objective is to measure the recognition time—defined as the duration between a target organ's initial appearance on the monitor and its identification by the surgeon. By evaluating both quantitative metrics, such as blood loss and operation time, and subjective factors, like surgeon confidence and fatigue, the ImNavi trial aims to provide objective evidence of AI's ability to reduce technical errors. Ultimately, this research provides a robust framework for integrating AI-driven decision support into standard surgical workflows to improve long-term patient outcomes.

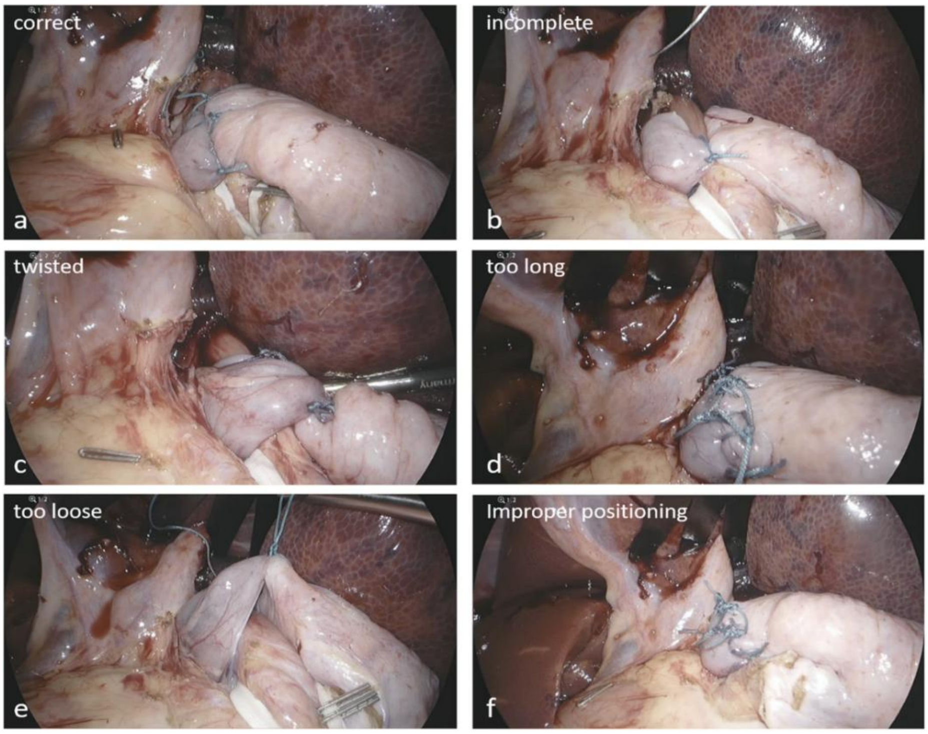


Fig. 5. Evaluation of surgical technique and intraoperative guidance. The AI system analyzes procedural correctness (e.g., "Correct" vs. "Twisted" maneuvers) to provide real-time navigation feedback, similar to the protocols established in the "ImNavi" clinical trials.

3.4 Discussion: Challenges and Future Directions

The widespread adoption of CV faces several barriers:

The "Black Box" Problem: The "Black Box" problem refers to the inherent opacity of deep learning models, whose internal logic remains uninterpretable to human observers. In high-stakes environments like laparoscopic surgery, this lack of transparency often hinders clinical trust and slows the adoption of AI-driven decision support. To address this, current research is shifting toward Explainable AI (XAI), which aims to provide surgeons with the biological or visual rationale behind algorithmic predictions. A primary method involves utilizing techniques like Grad-CAM or saliency maps to generate "heatmaps". As illustrated in Fig. 6, these visualizations highlight the specific anatomical regions—such as vessel tortuosity or organ boundaries—that the model prioritizes when identifying pathologies or safe dissection zones. By transforming the "black box" into a transparent decision-making process, XAI allows clinicians to audit AI outputs, recognize potential errors like overconfidence, and ensure patient safety through quantifiable intraoperative precision.

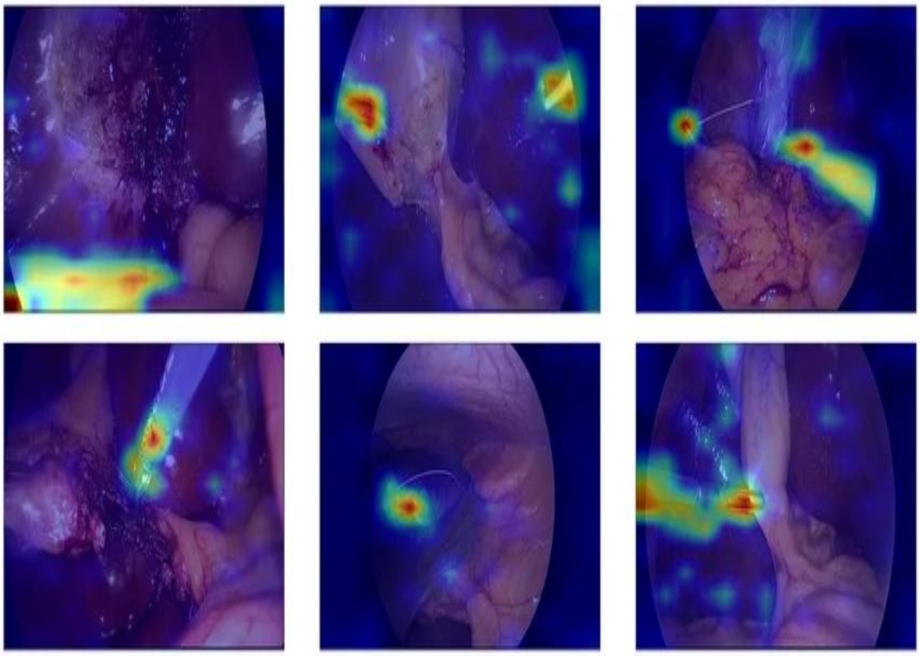


Fig. 6. Grad-CAM visualizations for Explainable AI (XAI) in surgery. The heatmap highlights the critical visual features, allowing surgeons to understand the internal decision-making logic of the deep neural network during anatomical identification.

Annotation Bottleneck: The annotation bottleneck is a critical challenge in the development of AI for laparoscopic surgery, as high-performing deep learning models typically require vast datasets of video manually labeled with high-fidelity annotations. This process is exceptionally costly and time-consuming, primarily because it necessitates the direct involvement of expert clinicians whose specialized knowledge is required to provide accurate "ground truth". The opportunity cost of diverting medical professionals from clinical duties to label thousands of frames is significant. Furthermore, surgical video presents unique complexities compared to static images, such as dynamic occlusions, varying perspectives, and diverse pathological states, which often lead to inter-observer variability. Even among experts, inconsistencies in labeling anatomical boundaries or surgical phases can compromise the reliability of the training data. To address this, current research is shifting toward weakly-supervised, semi-supervised, and self-supervised learning paradigms. These approaches, utilized in datasets like Endo700k and LSPD, aim to learn from unlabeled data or coarse labels, thereby significantly reducing the reliance on detailed manual annotation and accelerating the translation of AI tools into clinical practice.

Vision-Language Models (VLM): Represent a transformative frontier in surgical AI, moving beyond static image analysis toward interactive, multimodal systems. Emerging technologies such as SurgicalGPT allow for sophisticated, bidirectional communication, enabling surgeons to ask natural language questions—such as "Is this the ureter?"—and receive immediate, AI-driven visual and verbal confirmations. As demonstrated in Fig. 7, these systems integrate advanced computer vision with natural language processing to provide real-time semantic labeling of complex surgical scenes. Technically, models like SurgicalGPT utilize transformer-based architectures that process word tokens alongside visual features, mimicking human cognitive processes to better infer answers based on intraoperative imagery. Advanced methods like Chain-of-Look prompting further enhance this by decomposing tasks into interpretable steps through sequential video reasoning. Unlike traditional segmentation models that merely provide color-coded masks, VLMs can explain their reasoning, offering a biological or visual rationale for their conclusions. This capability is critical for bridging the transparency gap between algorithmic predictions and high-stakes clinical decision-making. By facilitating an intuitive "conversation" between the surgeon and the digital assistant, VLMs are expected to significantly reduce cognitive load during high-stakes procedures, establishing a new standard for intraoperative decision support.

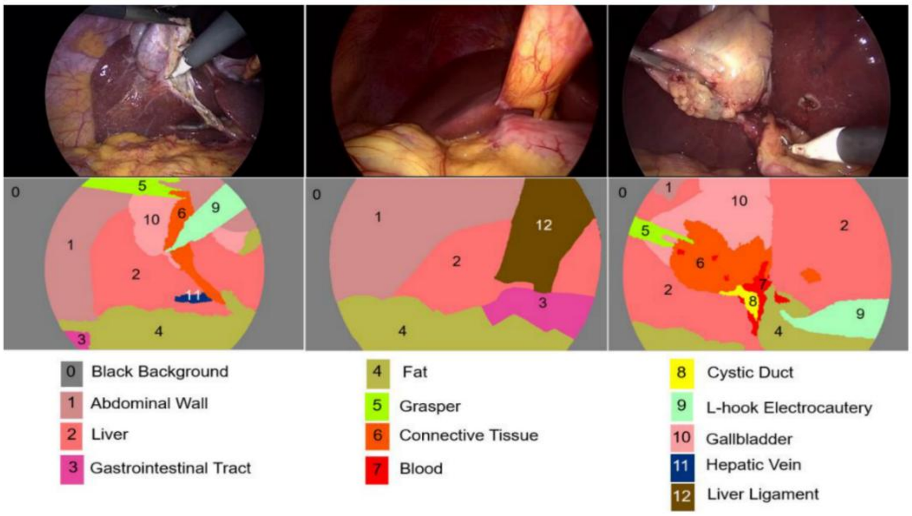


Fig. 7. Multimodal interaction in robotic surgery. The AI system provides comprehensive anatomical mapping and semantic labeling, laying the groundwork for future Vision-Language Models (VLM) like "SurgicalGPT".

4 Conclusion

The integration of computer vision (CV) into laparoscopic surgery represents a profound interdisciplinary shift that augments traditional surgical intuition with unprecedented engineering precision. This transformation is converting surgery into a digital, quantifiable, and inherently safer profession by bridging the gap between raw intraoperative data and real-time decision support. The core of this progress lies in the synergy between cutting-edge engineering architectures and robust clinical validation. 3D Convolutional Neural Networks (3DCNN), leveraged through the Laparoscopic Surgical Performance Dataset (LSPD), now allow for the automated assessment of surgical skill by analyzing spatiotemporal instrument coordination with an F1 score of 0.91. High-speed detection models such as Faster R-CNN and YOLOv9 further enhance this environment by localizing instruments and identifying complex lesions with a test precision of 0.98, effectively mitigating the visual perception errors that lead to preventable complications. Clinical relevance is demonstrated through pioneering initiatives like the "ImNavi" trial, where semantic segmentation algorithms (UreterNet and NerveNet) provide real-time navigation to protect vital autonomic nerves. Furthermore, by automating the Critical View of Safety (CVS) criteria, AI systems provide navigational masks that guide surgeons toward safer dissection strategies during cholecystectomy. This digital revolution also extends to the surgeon's physical well-being; CV-based ergonomic monitoring calculates craniovertebral angles in real-time to detect forward head posture (FHP) and prevent long-term musculoskeletal disorders. As the field transitions toward Vision-Language Models (VLM) like SurgicalGPT, the historical "black box" problem is addressed through Explainable AI (XAI), fostering a future of transparent, bidirectional communication between the surgeon and the machine. Ultimately, this convergence ensures a higher standard of care where technical innovation and clinical expertise work in a seamless, validated partnership.

5 Acknowledgment

The authors would like to express their sincere gratitude to the Central Customs Hospital, Odlar Yurdu University and Azerbaijan Technological University for providing the necessary clinical environment and academic resources required to conduct this interdisciplinary research. We are particularly grateful to the surgical team and the technical department staff for their invaluable support during the data collection and clinical validation phases of this study. This collaboration between the fields of general surgery and computer engineering has been fundamental to the synthesis of the clinical and algorithmic approaches presented in this article.

Funding: The authors declare that they have not received funding.

References

1. Badgery, H., Zhou, Y., Siderellis, A., Read, M., Davey, C.: Machine learning in laparoscopic surgery. In: Artificial Intelligence in Medicine, pp. 175–192. Springer Nature Singapore (2022)
2. Bondarenko, A., Jumutc, V., Netter, A., et al.: Object detection in laparoscopic surgery: A comparative study of deep learning models on a custom endometriosis dataset. Diagnostics (Basel) 15(10), 1254 (2025)
3. Chavez-Saavedra, G., et al.: Surgeon posture: Monitor and table position determined body work angles. ResearchGate. https://www.researchgate.net/publication/376408025_Surgeon_posture_Monitor_and_table_position_determined_body_work_angles (2023)
4. Corallino, D., Balla, A., Coletta, D., et al.: Systematic review on the use of artificial intelligence to identify anatomical structures during laparoscopic cholecystectomy: A tool towards the future. Langenbeck's Arch. Surg. 410, 101 (2025)

5. Frontiers in Pediatrics: AI-driven evaluation of surgical technique and intraoperative guidance.https://www.frontiersin.org/journals/pediatrics/articles/10.3389/fped.2025.1584628/full (2025)
6. Ganimat, I. B., & Qiyas, H. E. (2021). Analysis and selection performance indicators multiservice communication networks based on the concept NGN and FN. Computer and information systems and technologies.
7. Hashimoto, D.A., Rosman, G., Witkowski, E.R., et al.: Computer vision analysis of intraoperative video: Automated recognition of operative steps in laparoscopic sleeve gastrectomy. Ann. Surg. 270(3), 414–421 (2019)
8. Hashimov, E., Ibrahimov, B., Islamov, I., Talibov, A., & Akhundov, R. (2026). Research of the efficiency multiservice networks using MIMO technology. Advanced Information Systems, 10(1), 66-71.
9. Hirides, S., Hirides, P., Kouloufakou, K., Hirides, C.: Artificial intelligence and computer vision during surgery: Discussing laparoscopic images with ChatGPT4—Preliminary results. Surg. Sci. 15, 169–181 (2024)
10. Hong, W.Y., et al.: CholecSeg8K: A semantic segmentation dataset for laparoscopic cholecystectomy. ResearchGate. https://www.researchgate.net/figure/Images-from-the-CholecSeg8K-dataset_fig2_381401566 (2024)
11. Ibrahimov, B. G. (2025, November). Research Quality of Communication in Multiservice Telecommunication Networks. In 2025 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED) (pp. 1-5). IEEE.
12. Ibrahimov, B. G., Aliyev, A., & Rafizade, U. R. (2024, April). Research on the Effectiveness of Wireless Networks Cellular Communications Based on Digital Technologies. In International Conference on Information Technologies and Their Applications (pp. 55-65). Cham: Springer Nature Switzerland.
13. Ibrahimov, B., Alieva, A., & Rafizade, U. (2025, August). Research of the Spectral Efficiency Multiservice Networks and Radiotechnical Systems Using Multiantenna Transmission Technology. In 2025 6th International Conference on Problems of Cybernetics and Informatics (PCI) (pp. 1-4). IEEE.
14. Ibrahimov, B., Hasanov, A., & Hashimov, E. (2024). Research and analysis of efficiency indicators of critical infrastructures in the communication system. Advanced Information Systems, 8(2), 58-64.
15. Ibrahimov, B., Tagiyev, A., & Rafizade, U. (2024). Research on the Effectiveness Energy Saving Mechanisms in Mobile Networks when Using the Wi-Fi Radio Interface. Engineering Headway, 7, 157-163.
16. Kitaguchi, D., Fuse, N., Wakabayashi, M., et al.: Exploring the impact of an artificial intelligence-based intraoperative image navigation system in laparoscopic surgery on clinical outcomes: A protocol for a multicenter randomized controlled trial. Eur. Surg. Res. 67(1), 38–45 (2026)
17. Madani, A., Namazi, B., Altieri, M.S., et al.: Artificial intelligence for intraoperative guidance: Using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy. Ann. Surg. 276(2), 363–369 (2022)
18. Mascagni, P., Alapatt, D., Sestini, L., et al.: Computer vision in surgery: From potential to clinical value. npj Digit. Med. 5(1), 163 (2022)
19. Naghiyeva, A. (2025). A NEW METHOD FOR HIDING CONFIDENTIAL INFORMATION IN COLOR IMAGES. Research in: Technical & Aqrarian Sciences.
20. Osmanli, T. (2025). AI-Enhanced Predictive Modelling of Virtual Laboratory Microlearning in Online Distance Education. Journal homepage: http://iieta. org/journals/isi, 30(9), 2461-2471.

21. Power, D., Burke, C., Madden, M.G., et al.: Automated assessment of simulated laparoscopic surgical skill performance using deep learning. Sci. Rep. 15, 13591 (2025)
22. Prem Kumar, A., Sindhu, S., Manangi, M., et al.: An adaptation of computer vision of artificial intelligence for the assessment of postural ergonomics in laparoscopic surgery. World J. Laparosc. Surg. 16(3), 119–124 (2023)
23. Rafizade, U. (2025). VECTOR-STATISTICAL METHODOLOGY FOR EVALUATING RELIABILITY INDICATORS UNDER NOISY DATA CONDITIONS. Collection of scientific papers «ΛΌΓΟΣ», (September 5, 2025; Boston, USA), 124-133.
24. Rajaraman, S., et al.: Surgical tool detection in open surgery video. ResearchGate. https://www.researchgate.net/figure/An-example-of-surgical-tool-detection-in-an-open-surgery-video_fig1_364459703 (2022)
25. ResearchGate: Grad-CAM attention heat maps overlaid on the original image for surgical analysis. https://www.researchgate.net/figure/Grad-CAM-attention-heat-maps-overlaid-on-the-original-image-The-heatmap-coloring_fig4_391664540 (2025)
26. Scientific Data: High-quality clinical annotation for surgical datasets. Nature Portfolio. https://www.nature.com/articles/s41597-025-04642-4 (2025)
27. Springer Nature: Object detection and orientation in surgery. https://link.springer.com/chapter/10.1007/978-3-032-00267-9_8 (2022)
28. Tokuyasu, T., Iwashita, Y., Matsunobu, Y., et al.: Development of an artificial intelligence system using deep learning to indicate anatomical landmarks during laparoscopic cholecystectomy. Surg. Endosc. 35, 1651–1658 (2021)
29. Wauben, L.S., et al.: Critical view of safety (CVS) for laparoscopic cholecystectomy. ResearchGate.https://www.researchgate.net/publication/5628285_Critical_view_of_safety_CVS_for_laparoscopic_cholecystectomy (2011)
30. Zhou, R., Wang, D., Zhang, H., et al.: Vision techniques for anatomical structures in laparoscopic surgery: A comprehensive review. Front. Surg. 12, 1557153 (2025)
31. Нагиева, А. Ф. Г. (2022). Методы скрытия данных на основе интерполяции изображений. E-Scio, (3 (66)), 51-58.