approaches offer greater objectivity, reproducibility, and scalability. Studies have shown that a three-stage machine learning system, integrating instrument detection, motion feature extraction, and skill prediction, can achieve up to 87% accuracy in distinguishing high- and low-quality performance, suggesting that motion trajectories and operational efficiency serve as reliable indicators of surgical skill.

In robotic surgery, AI-based video analysis further expands the scope of skill assessment. Research indicates that models relying solely on two-dimensional video and residual neural networks can classify key skills such as suturing, needle passing, and knot tying with average accuracies exceeding 80%, demonstrating that robust automated assessment can be achieved even without high-fidelity kinematic sensors. Additionally, hybrid models combining convolutional neural networks and long short-term memory networks (CNN-LSTM) can simultaneously recognize surgical actions and assess surgeon expertise from a single video stream, achieving 81% accuracy in experimental settings. These findings indicate that AI video analysis can not only identify “what was performed” but also evaluate “how well it was performed” (Figure 3) (Hashemi et al., 2025).

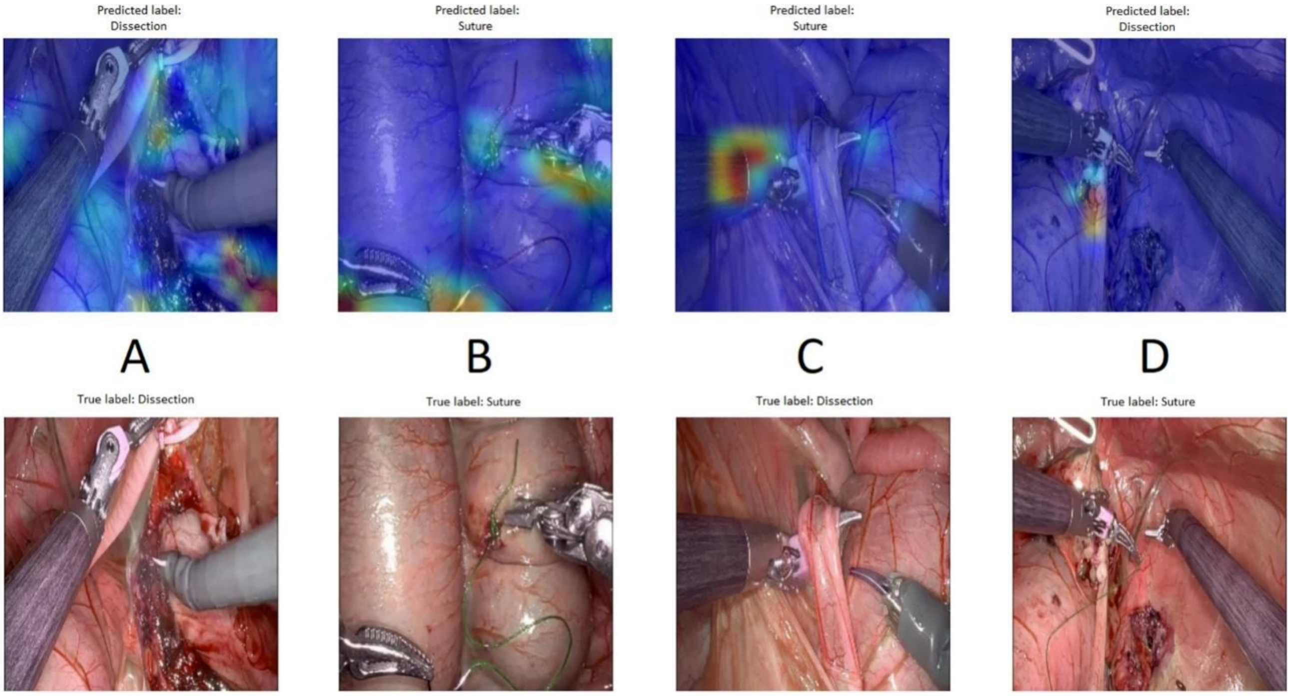

Figure 3 The regions of interest for the algorithm during surgical procedures (Adopted from Hashemi et al., 2025)
Image caption: A and B The areas in focus when classifying dissection and suturing, C shows a misclassification of a dissection sequence, and D shows a misclassification of suturing (Adopted from Hashemi et al., 2025)

Importantly, AI-based video assessment systems are evolving to integrate both summative evaluation and formative feedback. Some deep learning frameworks can generate high-stakes ratings comparable to expert judgments and provide interpretable feedback through heatmaps or key-frame visualization, highlighting areas of suboptimal performance. This enables trainees to identify deficiencies in instrument control, tissue handling, and procedural execution. Systematic reviews suggest that most AI models based on video or motion data achieve accuracies above 80% in simulated environments; however, limitations remain in external validation, standardization of evaluation metrics, and linkage to patient outcomes. In gynecologic laparoscopy and robotic surgery, the customization of such systems for procedures like hysterectomy, salpingectomy, and hysteroscopy could facilitate a shift from case volume-based training to competency-based progression (Tesfai et al., 2024).

5.2 Virtual reality and simulation-based training technologies

The introduction of virtual reality (VR) and augmented reality (AR) technologies has created highly controllable, repeatable, and low-risk training environments for minimally invasive gynecologic surgery. Systematic reviews and meta-analyses demonstrate that both high- and low-fidelity laparoscopic simulators significantly improve operative skills and reduce operative time compared with traditional observation-based or case accumulation