ACS 2026 Surgeons and Engineers: A Dialogue on Surgical Simulation

O-13

Research In-Progress

Scalpel and Code An AI-Driven Framework for Automated Surgical Skill Assessment Using Real-World Robotic Data

Han Grezenko; Josh Zhe; Kayla Gray; Mikaela G. Mahrer; Angela Mankin; Rylee McGuire, DO; James A. Mankin, MD; and Hahn Soe-Lin, MD, FACS

CommonSpirit, Phoenix, AZ, University of Arizona, Phoenix, AZ; CommonSpirit, Scottsdale, AZ; Creighton University, Phoenix, AZ

Introduction: The complexity of surgical training highlights the urgent need for objective and scalable assessment methods. Traditional evaluations rely heavily on subjective expert ratings, which are resource-intensive and inconsistent. To address this gap, we are developing Scalpel and Code: a research initiative integrating artificial intelligence (AI) and motion-tracking technologies to create a standardized framework for automated surgical skill evaluation.

Methods: We designed a two-stage approach: 1. A retrospective analysis of de-identified robotic surgery videos collected under institutional IRB approval; and 2. Validation of AI-derived performance metrics against expert surgeons' ratings. Videos are processed through custom computer-vision pipelines, allowing detection of key steps, tasks, instrument usage, and anatomical structures. Annotation protocols are developed with American Board-certified surgeons and guided by validated frameworks such as GEARS. AI model development is conducted in collaboration with data scientists and computer engineering partners, with iterative validation against blinded expert ratings.

Preliminary Results: Currently, over 20 robotic procedure videos have been processed and curated for analysis. The initial focus is on robotic transabdominal preperitoneal inguinal hernia repairs performed between 2020 and 2025. Procedural videos have been segmented into 10 key steps defined by expert consensus, including medial dissection across the pubic tubercle and Cooper's ligament, evaluation of Hesselbach's triangle, and dissection of the hernia sac. Early feasibility testing using open convolutional neural networks combined with publicly available human posture-tracking frameworks has demonstrated the potential to automatically track motion and distinguish robotic arm instruments within the testing dataset.

Next Steps: Next steps include expanding annotations across larger video datasets, refining AI models for higher accuracy, and integrating assessment outputs into a structured feedback platform for surgical trainees. The long-term goal is to embed the validated AI-driven system into surgical training scenarios, enabling objective and reproducible assessments that enhance skill acquisition and bridge engineering innovation with surgical education.

Figure 1: End-to-end workflow for developing, validating, and integrating the *Scalpel and Code* framework. Retrospective robotic training datasets are processed through custom computer-vision pipelines and annotated into critical key steps defined by expert surgeons. AI models are iteratively trained and benchmarked against blinded GEARS-based faculty evaluations. Validated outputs are then integrated to provide objective, real-time feedback for surgical trainees and inform curriculum design.