

ACS 2026 Surgeons and Engineers: A Dialogue on Surgical Simulation

O-02

Research Abstracts

Can AI Provide Reliable Feedback for Surgical Skills? A Rubric-Based Comparison with Faculty for Suturing Assessment

Nicholas Roth, MD; Kendall Gross, MD; Traves Crabtree, MD; Pinckney Benedict; Nicole Sommer, MD FACS; Janet Ketchum; Harrison Barr; Jessica Cantrall; and Sowmy Thuppal, PhD

Southern Illinois University School of Medicine, Springfield, IL

Introduction: Surgical residents undergo training and verification of proficiency (VOP) at the beginning of residency. The timeliness and quality of assessment and feedback is crucial, but rarely immediate and often delayed. Artificial intelligence (AI) offers the potential for immediate, detailed feedback that can compare and augment faculty feedback.

Methods: Surgeon driven rubric was developed with AI to assess technical performance and to generate real-time, rubric-based feedback to residents performing simple interrupted and running subcuticular suturing in a simulated environment. This was then compared to historical VOP feedback provided by surgeon video review over a 15-year period. Agreement and inferiority of evaluation between AI and faculty was compared using Cohen's Kappa and McNemar's Test.

Results: High agreement between AI and faculty for simple interrupted in "needle angle" (0.95), "skin approximation (0.9)", and "suture placement" (0.97) and low to moderate agreement in "forceps grasps" (0.48), "square knots"(0.79), and "economy of motion (0.69)" AI feedback for simple interrupted suture was non-inferior to faculty for "needle angle" ($p= 0.48$), "skin approximation" ($p = 0.13$), and "suture placement" ($p = 1$). AI was inferior to faculty for "forceps grasp" ($p = 7.5 E6$), "square knot" ($p = 0.007$), "skin eversion" ($p = 0.001$), and "economy of motion" ($p = 0.0008$). High agreement between AI and faculty for subcuticular in all categories other than "forceps grasps" (moderate, 0.74). AI and faculty were non-inferior in all categories other than "forceps grasps" ($p= 0.002$)

Conclusions: This study demonstrates that AI can be trained to deliver rubric-based feedback for suturing that is immediate, actionable, and largely agrees with and is non-inferior to faculty assessment. Some inferiority seen with AI feedback could be secondary to video lacking sufficient detail for the AI model to fully capture the nuanced details. These findings highlight the promise of augmenting feedback with AI assistance.