ACS 2025 Surgeons and Engineers: A Dialogue on Surgical Simulation Meeting

O-02

Research Abstracts

Machine Learning for Resident Learning: FLS PEG Transfer with Real-Time Evaluation Comparable to Faculty Evaluators

Vin Hudson, MD; Andrew Hu, MD; Joelle Getrajdman, MD, FACS; Nell Maloney Patel, MD, FACS; Issam Koleilat, MD, FACS; Divya Kewalramani; Yunzhe Xue; Usman Roshan; Mayur Narayan, MD, MPH, MBA, MHPE, FACS; and Advaith Bongu, MD, FACS

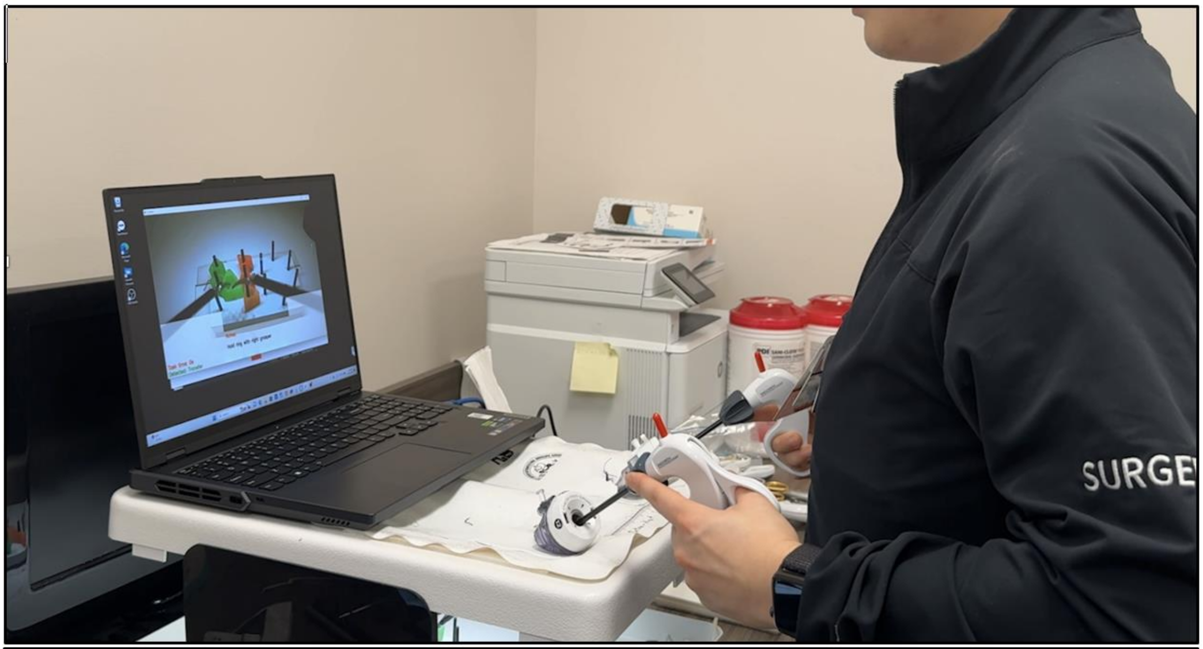
Robert Wood Johnson University Hospital, New Brunswick, NJ; University of Pittsburgh Medical Center, Pittsburgh, PA; New Jersey Institute of Technology, Newark, NJ

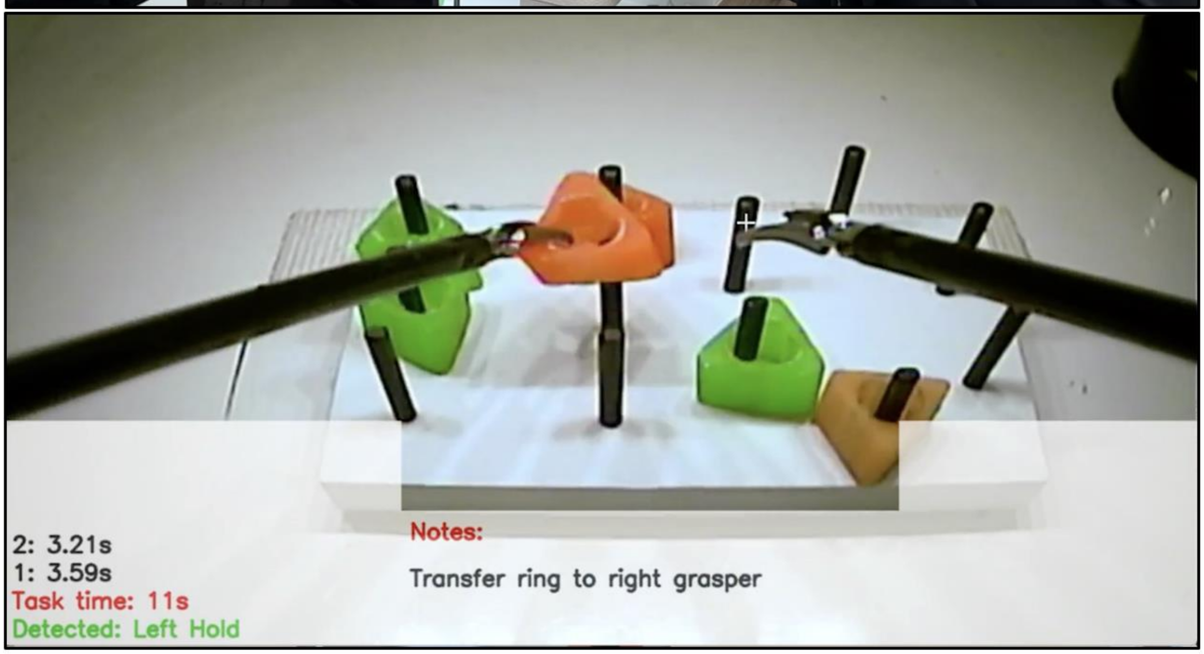
Introduction: While machine learning (ML) has advanced the capabilities of FLS trainers, real-time ML coaching remains limited. Currently available systems can provide detailed post-task skills assessments. This study aimed to assess if an ML-based system was capable of providing real-time FLS coaching and could be comparable to faculty assessors.

Methods: To determine the number of videos required for testing, sample size calculations were performed using PASS 2020. Trainees were recorded performing peg transfers with competency (PTC) defined by FLS guidelines as pass/fail. Task failures and their time points were recorded. A training set of 60 videos using a “you only look once” (YOLO) framework for object detection and logic for rule adherence was employed. The model’s performance was then validated with 24 additional videos. Faculty ground truth (FGT) on PTC was established a priori by the consensus of five educational faculty members. ML model performance compared to FGT was evaluated using a threshold for Cohen’s κ;; ≥ 0.70.

Results: The ML model correctly identified outcomes in 92% of cases, achieving a predictive performance (F1) score of 0.93 and Cohen’s κ;; of 0.83. This suggests strong agreement with FGT. Total analysis time for all videos was 6 minutes, compared to 56 minutes per faculty member. In 70% (31/44) of failure instances, the ML model detected errors concurrently with FGT, with identical error classification. Of the 13 remaining cases, 62% (8/13) involved earlier error detection by the ML model. Basic real-time coaching was successfully integrated into a heads-up user-interface (Figure 1).

Conclusions: The integration of ML for real-time coaching into the FLS simulation environment was successfully achieved, demonstrating efficiency and accuracy better than expert faculty. This approach both reduces time required for assessments and enhances consistent error detection.


SURGE


2: 3.21s
1: 3.59s
Task time: 11s
Detected: Left Hold
Notes:
Transfer ring to right grasper