Virtual ACS 2021 Surgeons and Engineers: A Dialogue on Surgical Simulation Meeting

Research

Segment-level Assessment of Surgical Technical Skill using Machine Learning for Automated Surgical Coaching and Deliberate Practice

Anand Malpani, PhD; S. Swaroop Vedula, MBBS PhD; Chi Chiung Grace Chen, MD MHS; and Gregory D. Hager, PhD.
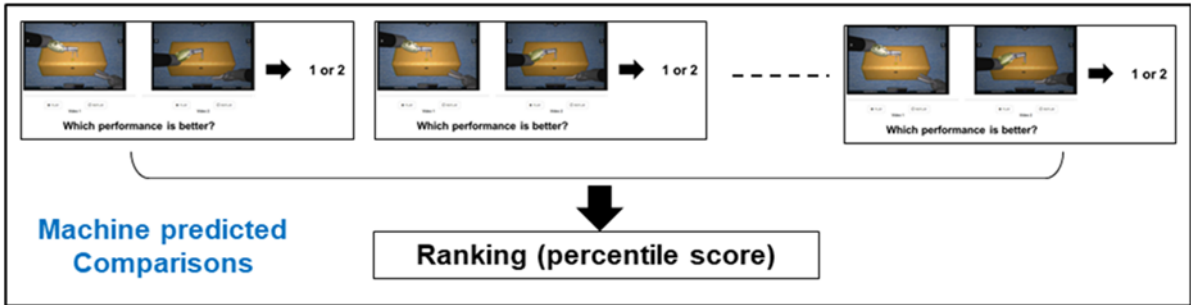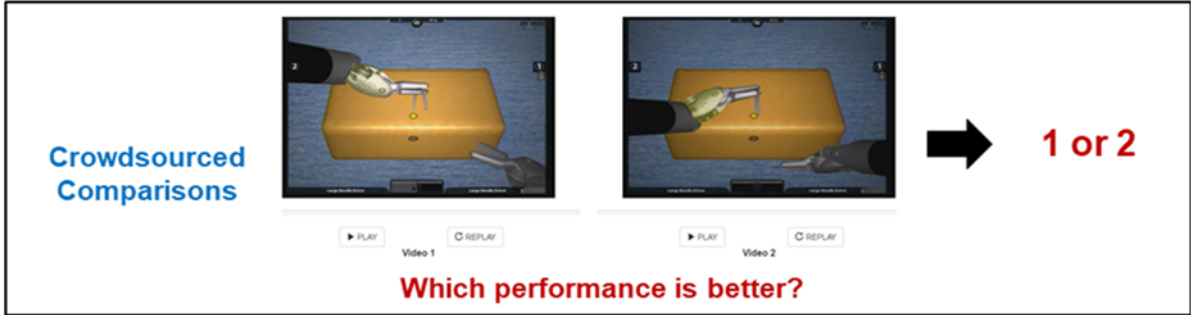
Johns Hopkins University, Baltimore, MD

Introduction: Technical skills coaching is important for improving patient outcomes in surgery. However, expert one-on-one coaching is not scalable for routine assessment and feedback. Our work is towards augmenting a human surgical coach with an automated virtual coach. Routine and targeted assessment is needed to enable deliberate practice which leads to efficient and effective learning. In this work, we present an approach that can generate ranking scores for a given performance at segment-level.

Methods: We used a dataset of 30 performances of the "Suture Sponge I" task available on the da Vinci Skills Simulator, a virtual reality simulation training platform for the da Vinci system. This dataset contained video, instrument motion, and endoscope motion recordings. We labeled start and end of each constituent needle passing segment resulting in 360 such segments. We obtained pairwise comparisons-based skill ratings for 100 pairs of performances generated by random selection of segments. This involved a rater to view a pair of performances side-by-side on a web page and select their "preference" indicating the better skilled performance. The rater indicated their level of confidence in choosing the preference on a 3 choice question. We recruited 5 raters per pair and chose the majority rating as the preference for the pair. We computed 7 metrics using motion data, e.g., completion time, instrument path length, instrument shaft area swept, and instrument velocity peaks. We used the "support vector machine" algorithm, a machine learning technique, to predict preferences by using the metrics for the given pair of performances. We performed 5-fold cross validation to estimate the accuracy of the algorithm.

Results: The inter-rater agreement in preference ratings was moderate (Fleiss kappa of 0.51, percent agreement of 85.5%). The support vector machine algorithm predicted preferences with 85.48% accuracy (standard error: 0.35).

Conclusions: A support vector machine algorithm can predict pairwise comparisons of needle passing segments with accuracy that is similar to the inter-rater agreement within human ratings of such comparisons.

Crowdsourced Comparisons

PLAY   REPLAY   PLAY   REPLAY
Video 1   Video 2

→ **1 or 2**

Which performance is better?

Machine predicted Comparisons

Which performance is better? → 1 or 2

Which performance is better? → 1 or 2

Which performance is better? → 1 or 2

Ranking (percentile score)

Segment-level assessment using pairwise comparisons-based ranking. The percentile rank scores provide targeted feedback to the learner on which segment to focus on enabling deliberate practice.