Automated Video Evaluation of Orthopedic Surgical Skills Using Pose Estimation Algorithms

Thomas Nyul, BS[1], Aritra Chakraborty, BS[2], Sean Sibley, DPT[3], Kyle Zelenski, BS[4], Daniel Rohde, BS[5], Meir T. Marmor, MD[5]
[1]Dartmouth University Geisel School of Medicine, [2]Loyola University Chicago Stritch School of Medicine, [3]University of Colorado School of Medicine,
[4]Des Moines University College of Osteopathic Medicine, [5]University of California, San Franciso
Email of Presenting Author: thomas.e.nyul.med@dartmouth.edu

Disclosures: Thomas Nyul (N), Aritra Chakraborty (N), Sean Sibley (N), Kyle Zelenski (N), Daniel Rohde (N), Meir T. Marmor (N)

INTRODUCTION: Assessing surgical skills in orthopedic surgery training is vital for skill acquisition and high-quality patient care. However, traditional evaluation methods like global rating scales (GRS) and operation-specific checklists have limitations including the need for highly trained evaluators and evaluator bias. The aim of this study was to identify a method for surgical skill evaluation using basic 2-dimensional video and a pose estimation algorithm. We hypothesized that single-camera video capture of surgical tasks processed with a pose estimation algorithm (OpenPose) could identify patterns in upper extremity movements that are strongly associated with GRS score rated by experienced surgeon evaluators. Ultimately, we aim to develop an unbiased, easily applied motion analysis-based evaluation protocol for surgical skill development.

METHODS: Institutional review board approval was obtained and informed consent was given by each participant. There were 24 total subjects included (20 male, 4 female). Medical students (n=7), residents (n=12), fellows (n=2) and attending surgeons (n=3) were filmed performing standardized surgical tasks in a cadaveric lab. Tasks included surgical incision, suturing, one-handed and two-handed ties, clamp reduction of a fracture, lag drilling and screwing, and plate fixation with screws. Standard 2D videos were captured with an iPhone camera at a standardized distance and position from the operating table **(Figure 1**). Video was analyzed using OpenPose, an open-source pose-estimation algorithm that tracks human movement in 2D video without need for specialized markers or devices. The "BODY_25 model" with the "hand keypoint module" (21 landmarks per hand) was applied to extract upper extremity and fine motor motion features. Time to task completion, an established positive correlate of greater surgical skill, was also measured for each task.

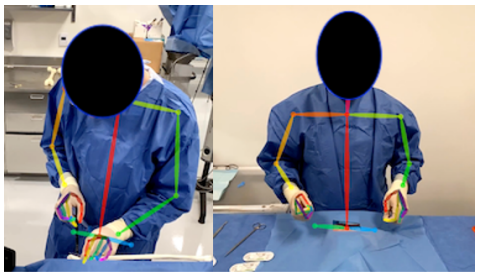


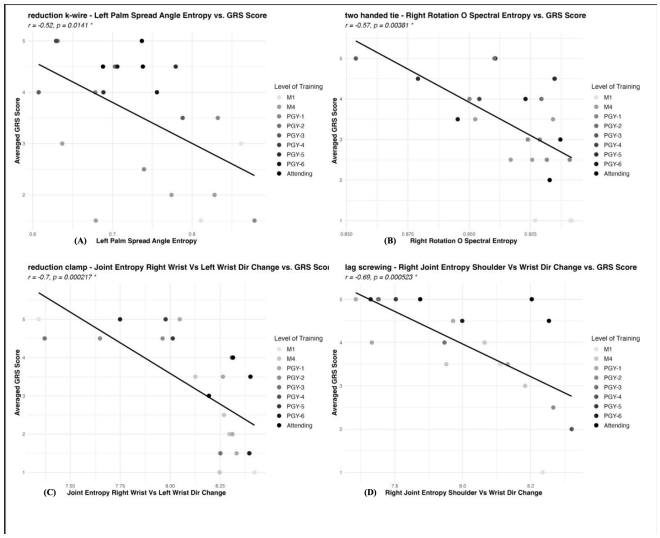
Figure 1: Subjects performing surgical tasks, processed with OpenPose.
(Left: plate screwing, Right: two-handed tie)

Joint coordinates were processed in Python to compute kinematic and entropy-based metrics, including position, velocity, acceleration, and joint/finger entropy. Expert evaluators scored each task recording using the validated Global Rating Scale (GRS) while blinded to subject identity through face occlusion. Exploratory analysis identified features with moderate-to-strong correlation to GRS. To further assess predictive performance, linear regression models and a random forest regression model with leave-one-out cross-validation (LOOCV) were applied. Feature importance from the random forest was used to highlight motion variables most strongly associated with GRS scores, with model performance reported as LOOCV R^2.

RESULTS SECTION: Across all tasks, multiple upper extremity and hand metrics demonstrated moderate-to-strong correlations with GRS scores (Pearson r = 0.40-0.65, p < 0.05) (**Figure 2**). Velocity and entropy-based features, particularly wrist and finger-spread entropy, were consistently significant. Time to task completion had strong to very strong correlation with GRS scores (Pearson r = -0.62 to -0.84, p < 0.05), other than plate screwing (Pearson r = -0.40, p= 0.07) which did not reach statistical significance. The random forest model achieved an overall LOOCV R^2 of 0.42, with best predictive performance for the plate screwing task ($R^2 \approx 0.60$) and two-handed tie ($R^2 \approx 0.45$). Feature importance highlighted fine-motor entropy measures, especially right and left hand spread metrics, as the strongest predictors of surgical skill **(Figure 3)**.

DISCUSSION: The findings of this study demonstrate that features extracted from single-camera 2D video were strongly correlated with expert-evaluated surgical skill, despite the lack of a single uniform metric correlated across all surgical tasks. Still, a random forest regression model explained 42% of the variance in expert GRS scores under LOOCV, indicating that 2D video–derived motion features capture meaningful but incomplete aspects of surgical performance. This moderate predictive ability supports their use as objective adjuncts to expert assessment, while highlighting the need for additional features or multimodal integration to achieve robust skill evaluation. Key limitations include single-institution, convenience sample design, and reliance on a uniform instructional setting. Task specificity, participant resistance to standardized methods, and the fixed task sequence further constrain applicability to real-world surgical settings.

SIGNIFICANCE/CLINICAL RELEVANCE: This pilot study leverages single-camera motion capture and advanced computer vision algorithms to create an objective, unbiased framework for assessing orthopedic surgical skills. By standardizing evaluation and enabling real-time intraoperative feedback, it aims to reduce evaluator burden, standardize assessments, and ultimately improve surgical training in an effort to improve patient outcomes.



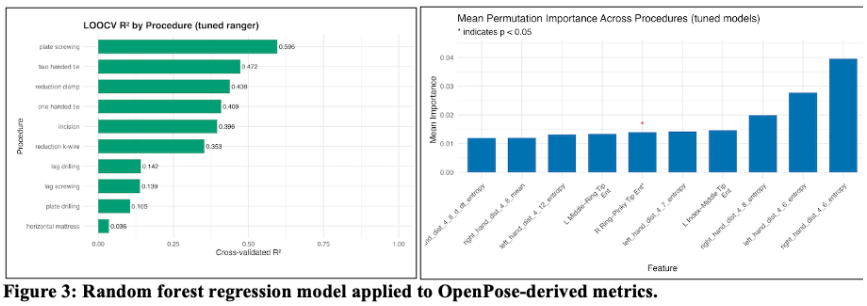


Figure 3: Random forest regression model applied to OpenPose-derived metrics.
Model achieved an overall LOOCV R^2=0.42, with best predictive performance for the plate screwing task (R2 ≈ 0.60) and two-handed tie (R2 ≈ 0.45). Model explained 42% of the variance in expert GRS across all tasks, indicating moderate predictive ability.