Integrating Kinematic and Kinetic Features for Automated Skill Classification in Robotic Suturing

Noah Drakes[3], A. Michael West Jr.[1], Sergio Machaca[1],
Gina Adrales[2], Peter Kazanides[3], Swaroop Vedula[4], Jeremy D. Brown[1]

[1]*Dept. of Mechanical Engineering, Johns Hopkins University, Baltimore, MD, USA*
[2]*Johns Hopkins School of Medicine, Baltimore, MD, USA*
[3]*Dept. of Computer Science, Johns Hopkins University, Baltimore, MD, USA*
[4]*Malone Center for Engineering in Healthcare, Johns Hopkins University, Baltimore, MD, USA*
ndrakes1@jh.edu

INTRODUCTION

Automated surgical skill assessment has predominantly relied on kinematic descriptors—such as position, velocity, and motion smoothness—to quantify task performance [1], [2]. However, force modulation during tool–tissue interaction reflects essential aspects of motor control not captured by motion alone [3]. Integrating kinetic metrics, including interaction force and jerk magnitude, may therefore provide a more comprehensive representation of surgical expertise and support the development of adaptive haptic feedback strategies for surgical training.

We evaluate whether combining kinematic and kinetic features improves automated classification of surgical expertise using a multilayer perceptron (MLP) and leave-one-user-out (LOUO) validation. Model performance was quantified via pooled area under the ROC curve (AUC) and class-wise sensitivity and specificity averaged across folds, with feature importance assessed through permutation analysis.

MATERIALS AND METHODS

Participants and Task. Twenty-three participants—17 novice surgical residents, 3 intermediate surgical educators, and 3 expert surgeons ($\geq$ 50 robotic cases)—performed a standardized intracorporeal suturing task on a silicone phantom using the da Vinci Si system. Each subject completed a warm-up test in addition to six repetitions of the task. The study was IRB-approved (#IRB00205668) with informed consent obtained from all participants.

Data Acquisition. Kinematic and kinetic data were collected using the MOCHA framework [4], capturing PSM motion, interaction forces, platform reactions, and tool accelerations. As shown in Fig. 1, the PSMs were instrumented with an accelerometer and ForceN sensor (a), and the phantom was mounted on a sensorized platform to measure tool–tissue forces (b). Only recorded motion and force data were analyzed.

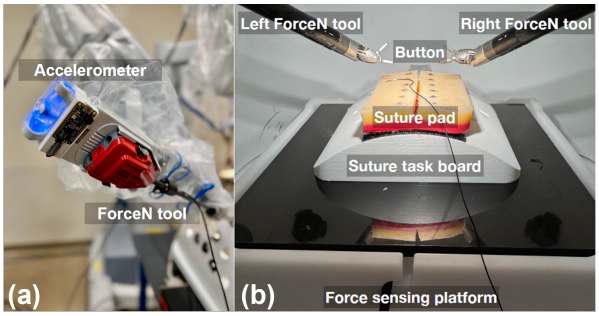

Fig. 1 (a) Patient-side manipulator (PSM) instrumented with an external accelerometer and ForceN sensor to capture tool dynamics and interaction forces. (b) Suturing task environment consisting of a silicone tissue phantom mounted on a sensorized platform for measuring tool–tissue forces.

Feature Extraction and Statistical Analysis. Raw motion and force data were partitioned into individual trials and denoised prior to normalization. For each repetition, 36 quantitative features were derived—13 describing instrument motion and 23 reflecting interaction forces (for a full list of metrics, see Fig. 2). Each metric was evaluated across expertise levels using Welch's one-way ANOVA to account for unequal variances between groups. Given the 36 comparisons performed, a Bonferroni correction was applied to control the family-wise error rate ($\alpha = 0.05/36 \approx 0.0014$). Metrics with $p < \alpha$ were considered statistically significant indicators of skill-level differences.

Supervised Learning. An MLP classifier was trained to predict expertise level (novice, intermediate, expert) using z-score normalized metrics as input features. Generalization was assessed using leave-one-user-out validation, a subject-independent scheme commonly used in surgical skill assessment [2]. Performance was quantified by pooled AUC and by sensitivity and specificity averaged across folds (mean ± std). Permutation importance

TABLE I Classification performance across feature sets. Reported are macro and weighted AUC, with class-specific sensitivity (Sens.) and specificity (Spec.) across LOUO folds, with the mean ± std across folds.

Feature Set	AUC_{mac}	AUC_{wgt}	Nov. Sens.	Nov. Spec.	Int. Sens.	Int. Spec.	Exp. Sens.	Exp. Spec.
All	0.96	0.98	1.00±0.00	0.78±0.37	0.67±0.47	1.00±0.00	0.67±0.47	0.97±0.15
ANOVA Top16	0.91	0.95	0.98±0.05	0.33±0.42	0.33±0.47	0.99±0.04	0.00±0.00	0.94±0.18
Kinematics	0.94	0.98	0.99±0.04	0.92±0.19	0.39±0.44	0.93±0.24	0.50±0.41	0.93±0.21
Kinetics	0.94	0.98	1.00±0.00	0.64±0.39	0.06±0.08	0.97±0.11	0.61±0.44	0.94±0.19
Perm. Top16	0.94	0.97	0.99±0.04	0.67±0.35	0.33±0.47	0.98±0.05	0.44±0.42	0.93±0.23

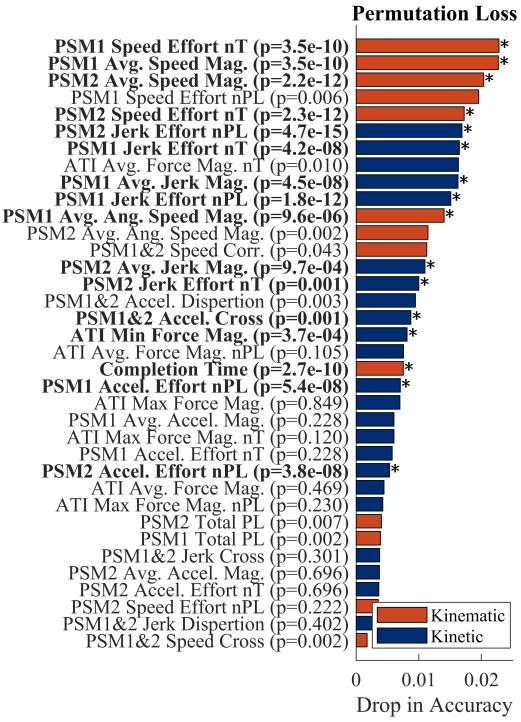

Fig. 2 Permutation importance ranked by mean accuracy drop after feature shuffling. Bolded features were ANOVA-significant ($p < 0.0014$); asterisks denote strong Bonferroni-corrected significance. nT denotes metrics normalized by time (i.e., divided by task completion time), while nPL denotes metrics normalized by path length (i.e., divided by total instrument path length).

measured the mean reduction in classification accuracy following random shuffling of each feature.

RESULTS

Under LOUO validation, the MLP achieved a macro-AUC of 0.96 and weighted AUC of 0.98 when trained on combined kinematic and kinetic metrics, outperforming single-domain models (AUC ≈ 0.94). The integrated model exhibited high novice sensitivity (1.00 ± 0.00), expert specificity (0.97 ± 0.15), and improved intermediate sensitivity (0.67 ± 0.47), indicating balanced classification across expertise levels.

Permutation importance and ANOVA identified multiple significant predictors from both domains, including speed effort, jerk effort, and normalized acceleration effort, highlighting complementary contributions of motion and force-based features.

DISCUSSION

Kinetic features provided complementary and nonredundant information to traditional kinematic descriptors in classifying surgical expertise. While kinematic metrics capture motion smoothness and spatial efficiency, kinetic measures reflect regulation of interaction forces, an equally important component of skilled manipulation. Several kinetic metrics were statistically significant and ranked highly in permutation importance analysis, underscoring their discriminative value.

Although intermediate surgeons are often hard to classify [5], [4], as their performance frequently exhibits characteristics shared with both novice and expert groups, combining kinematic and kinetic features improved differentiation of this cohort. This is shown in Table I, where the model trained on all metrics produced the highest intermediate sensitivity. Future work will incorporate expert GEARS ratings [6]—a validated global rating scale assessing surgical-robotic performance—to relate model outputs to standardized human evaluations. We also plan to explore temporal modeling and shared latent representations between motion and force domains to better capture their underlying coordination structure.

REFERENCES

[1] C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, "Review of methods for objective surgical skill evaluation," *Surgical endoscopy*, vol. 25, no. 2, pp. 356–366, 2011.

[2] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, D. D. Yuh *et al.*, "Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling," in *MICCAI workshop: M2cai*, vol. 3, no. 2014, 2014, p. 3.

[3] C. R. Wagner, R. D. Howe, and N. Stylopoulos, "The role of force feedback in surgery: analysis of blunt dissection," in *Haptic Interfaces for Virtual Environment and Teleoperator Systems, International Symposium on*. IEEE Computer Society, 2002, pp. 73–73.

[4] S. Machaca, Z. Karachiwalla, N. D. Riaziat, and J. D. Brown, "Towards a ros-based modular multi-modality haptic feedback system for robotic minimally invasive surgery training assessments," in *2022 International Symposium on Medical Robotics (ISMR)*. IEEE, 2022, pp. 1–7.

[5] A. Hendricks, M. Panoff, K. Xiao, Z. Wang, S. Wang, and C. Bobda, "Exploring the limitations and implications of the jigsaws dataset for robot-assisted surgery," *IEEE Robotics and Automation Letters*, vol. 9, no. 11, pp. 9534–9541, 2024.

[6] R. Sanchez, O. Rodríguez, J. Rosciano, L. Vegas, V. Bond, A. Rojas, and A. Sanchez, "Robotic surgery training: construct validity of global evaluative assessment of robotic skills (gears)," *Journal of Robotic Surgery*, vol. 10, 09 2016.

[7] Y. Zheng and A. Majewicz-Fey, "Transformer-based automated skill assessment and interpretation in robot-assisted surgery," in *2024 International Symposium on Medical Robotics (ISMR)*, 2024, pp. 1–8.