18-6348

Surgeon technical skill assessment using computer vision-based analysis

Yetong Zhang, Hei Law, Tae-Kyung Kim, David Miller, James Montie, Jia Deng, Khurshid Ghani, for the Michigan Urological Surgery Improvement Collaborative, Ann Arbor, MI*

INTRODUCTION AND OBJECTIVES: There is evidence that assessments of surgical skill correlates with patient outcomes. We developed a computer vision based system to analyze skill, leveraging crowdsourcing to compile training data, convolutional neural networks (ConvNets), and machine learning algorithms to differentiate skill level of surgeons performing robotic prostatectomy. The aim of this study is to demonstrate its accuracy and establish computer vision metrics for performance.

METHODS: Videos of the urethrovesical anastomosis from 29 surgeons in the Michigan Urological Surgery Improvement Collaborative underwent blinded review by 56 peer surgeons using the Global Evaluation Assessment of Robotic Skills (GEARS). Movement of the right and left robotic instruments were annotated using crowdworkers on Amazon Mechanical Turk. We utilized the annotations to train a robotic instrument tracker using a ConvNets algorithm designed for human pose estimation. Computer vision metrics were correlated with peer ratings for surgical skill. We then assessed the ability of the system to predict skill movement of the robotic instruments and classify the skill of a surgeon with a linear support vector machine (SVM) classifier, using peer evaluations of skill as reference standard.

RESULTS: Peer GEARS scores ranged from 14.2 to 24.9. In total, 364904 video frames were annotated by 3242 crowdworkers. Accuracy (area under curve) for detecting the right and left robotic instruments in surgical videos with the ConvNets algorithm was 0.80 and 0.84, respectively. Computer vision derived features that correlated with peer ratings of skill included average speed, acceleration, and smoothness of instrument movement; Peak velocity and acceleration; Duration; Idle time; Dexterity (% time both tools are moving); and changes in distance between instrument tips (Figure 1). Using these features, a linear classifier could predict skill in 29 surgeons based on overall GEARS or domain scores for efficiency, bimanual dexterity, depth perception, or force sensitivity ranging from 83% to 100%.

CONCLUSIONS: Our system uses crowdsourcing and deep neural networks to track instruments and assess the skill level of surgeons performing robotic surgery. Since computer vision methods rely only on video data, it has the potential to be transferred to other minimally invasive procedures.

Source of Funding: Blue Cross Blue Shield of Michigan; Intuitive Surgical

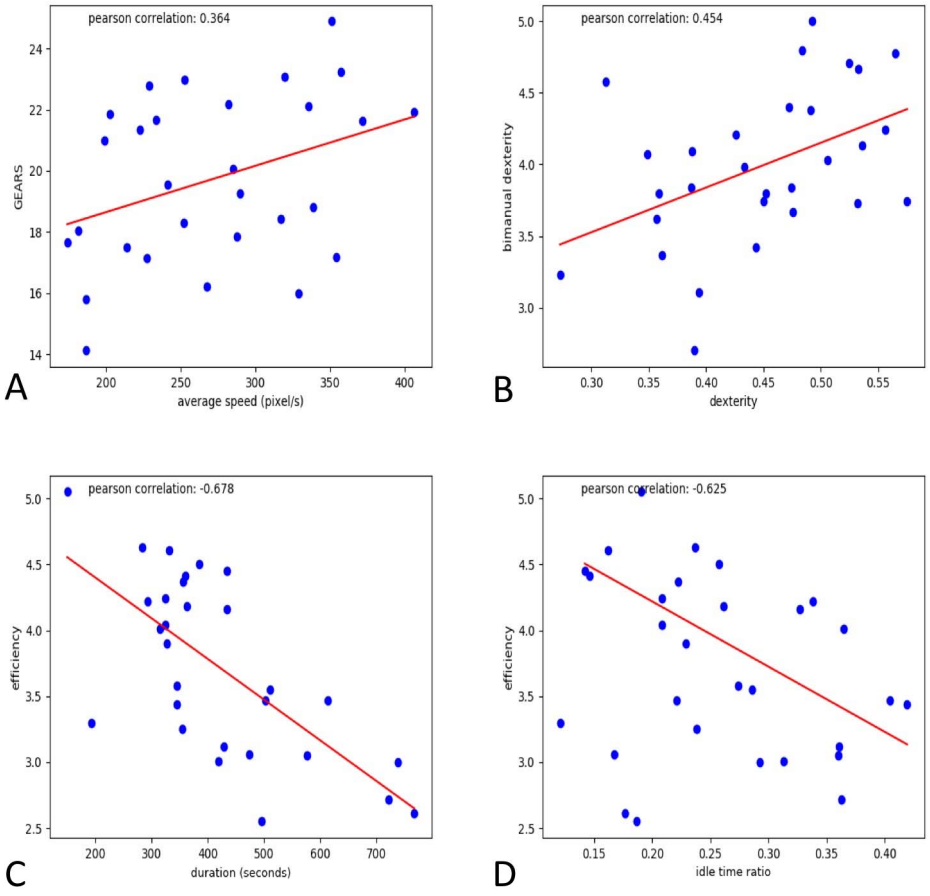

Figure 1. Correlation between GEARS peer ratings of skill (y axis) and computer-vision derived features of skill (x axis) for surgeons performing robotic prostatectomy. Selected features: A. Average speed. B. Dexterity C. Duration D. Idle time