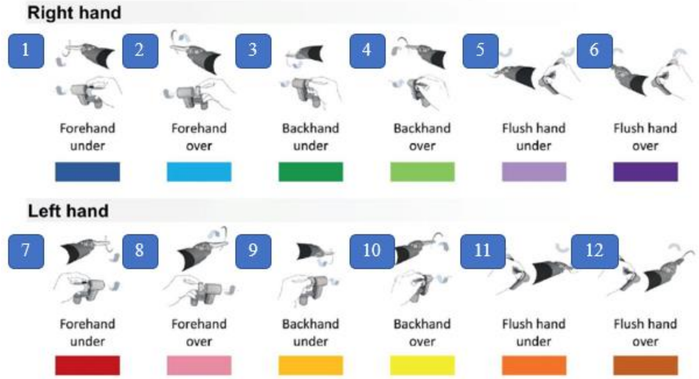
Deep-learning based computer vision to automate identification of suturing gestures

Andrew J. Hung[1], Aastha[2], Jessica H. Nguyen[1], Kartik Aron[1], Vijay Damerla[1], Yan Liu[2]
1. USC institute of Urology; 2. Department of Computer Science, Viterbi, USC

Introductions and objectives

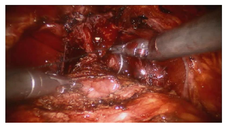
- In our previous work, needle driving gestures during a vesicourethral anastomosis of robotic prostatectomy were classified and associated with tissue tears/patient outcomes



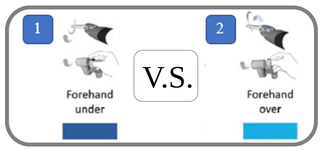
- Herein, we train and validate deep-learning based computer vision (CV) to automate the identification of suturing gestures

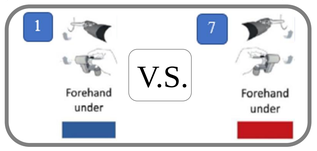
Methods

- Two raters manually identified the "ground truth" gesture labels of sutures during live surgery



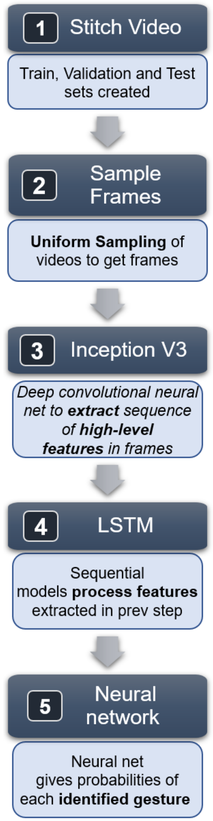
- An AI agent was built to distinguished Gesture 1 (G1) vs Gesture 2 (G2), and G1 vs Gesture 7 (G7)





Methods

- The AI-agent was comprised of computer vision, sequential, and deep learning models



- **Step 1:** Videos were sorted into training, validation, and testing sets.
  G1 vs G2: 122 training, 31 validation, 31 test
  G1 vs G7: 161 training, 41 validations, 40 test
- **Steps 2-3:** Each video was uniformly sampled at 10 frames per second and applied to Inception V3, a 42-layer deep learning network, to extract high-level features from each frame.
- **Step 4:** Frames were processed using three sequential models
  - Long Short Term Memory (LSTM)
  - Gated Recurrent Units (GRU)
  - Recurrent Neural Networks (RNNs)
- **Step 5:** Output was applied to neural network to process an identified gesture. Chi-square test compared the classification performance to a random classifier (50% chance of each gesture).

Results

Classification	Predictive Model	Accuracy (%)	p Value
G1 vs. G2	**LSTM**	**90.32**	**0.002**
	GRU	86.67	0.040
	RNN	90.00	0.040
G1 vs. G7	**LSTM**	**92.50**	**<0.001**
	GRU	82.50	0.040
	RNN	**92.50**	**<0.001**

- G1 vs G2, which differ in needle grasp (over vs. under), was best distinguished by LSTM (accuracy=90.32%, p=0.002) compared to GRU and RNN.
- In order to distinguish G1 vs G7, differing in use of left vs. right instrument, LSTM and RNN were both best performing, achieving a 92.50% accuracy (p<0.001), followed by GRU (82.50%, p=0.04).

Conclusions

- Our results demonstrate CV's ability to recognize features that distinguish suturing gestures.
- Future work includes automatic detection of each classified gesture and automated risk assessment feedback, based on gesture and tissue location (urethra/bladder neck clock position), and likelihood of tissue trauma according to our database of "gestures-to-tissue tear" library.