UT Southwestern
Medical Center

**Dear Endourological Society,**

Thank you so much for affording me this wonderful opportunity to participate in urological research over the summer. I am humbled and so grateful to have been selected as one of the 2020 Summer Student Scholarship recipients. I had an enriching, challenging, and truly meaningful research experience.

This project has helped me grow both as a student and as a member of the academic community at large. The opportunity to explore urology so early in my medical school career has been instrumental in my professional development. Thank you for making it possible.

I would also like to thank my mentor, Dr. Alaina Garbens, for her guidance and patience throughout this project and her support and advice in pursuing a career in urology.

Sincerely,

Avery Wolfe

MD Candidate, Class of 2022

**UT Southwestern Medical Center**

UT Southwestern
Medical Center

# USING A MULTI-TASK CONVOLUTIONAL NEURAL NETWORK TO PREDICT SURGEON SKILL AND PATIENT OUTCOMES IN ROBOTIC PARTIAL NEPHRECTOMY

*Avery Wolfe BS, Alaina Garbens MD PhD*

*UT Southwestern Medical Center, Department of Urology, Dallas TX*

## INTRODUCTION:

Surgeons who have higher levels of technical skill have significantly lower rates of complications, patient mortality and hospital costs.[1, 2] Therefore, looking at ways to rate, track, and improve surgeon performance is important as it could improve patients' safety and hospital efficiency. There are two main scoring systems to evaluate surgeon technical skill level that have been extensively validated in the literature - the Objective Structured Assessment of Technical Skill grading scale (OSATS, Appendix A)[3] and the Global Evaluative Assessment of Robotic Skills grading scale (GEARS, Appendix B).[3, 4]

Currently, the only method to assess and score surgeon skill level is through video review using human raters. This process is costly, and both time and labor intensive. Therefore, video rating of surgeon performance has not become widespread standard of practice, despite its many benefits.

One solution is the use of a multi-task convolutional neural network (mtCNN). A mtCNN can predict multiple, independent outcomes (ex. predicting 6 GEARS domain scores). Our pilot study using a prostatectomy anastomotic model found that assessment of surgeon skill using GEARS can be reproduced with high accuracy using human expert review as the gold standard. We wish to broaden the use of mtCNN to "real" surgeries and attempt to correlate surgeon skill to patient outcomes. The objective of this project is to train and evaluate a mtCNN to predict surgeon technical skill and patient outcomes in robotic partial nephrectomy (RPN).

## METHODS:

RPN cases between January 2016 and December 2019 are automatically recorded at our institution onto a secure cloud network and are downloadable for educational/research purposes.

We reviewed these recorded cases and identified all T1 tumors for inclusion in the project. We excluded primarily cystic tumors, as they proved to be too technically challenging to rate reproducibly according to the OSATS and GEARS criteria. Video analysis and rating was focused on the tumor resection and renorraphy steps given they are the most difficult steps and directly affect oncological outcomes.

Each patient case was separated into three separate sections during which surgeon skill levels were rated separately – the tumor resection, the renorraphy, and the bolster step. Many of these cases included more than one surgeon operating at different stages of the cases, so surgeon switching was noted and documented to score each surgeon separately. OSATS and GEARS scores were assigned for each of the three case sections, and if more than one surgeon operated during that section, individual OSATS and GEARS score were assigned for each participating surgeon.

Information on patient demographics, clinical tumor stage, RENAL nephrometry score, tumor location, and pre- and post- operative blood work was also collected, along with perioperative data, final pathological stage, histology, positive margin status, biochemical recurrence and date of last follow up. Data on the operating surgeon such as surgeon type (faculty, fellow or resident), fellowship trained and years in practice was also recorded.

**PRELIMINARY RESULTS:**

To date, 34 RPN videos have been rated. The mean OSATS score for the tumor resection step was 30.8/35 (standard deviation 2.8, range 24-35), and the mean GEARS score was 26.7/30 (standard deviation 2.5, range 21-30). For the renorraphy step, the mean OSATS score was 31.6/35 (standard deviation 2.4, range 26-35) and the mean GEARS score was 26.6/30 (standard deviation 2.2, range 22-30). Finally, the mean OSATS score for the bolster step was 31.6/35 (standard deviation 2.7, range 26-35), and the mean GEARS score was 26.4/30 (standard deviation 2.5, range 21-30). The vast majority of the cases (30/34, 88%) did have more than one surgeon operating throughout various portions the patient case; the scores for each of these surgeons were counted separately in the analysis. Preliminary data has also been shared with our computer science colleagues at Southern Methodist University (SMU) in Dallas, TX who will oversee the machine learning aspect of the project and training of the mtCNN.

**PRELIMINARY DISCUSSION AND CONCLUSION:**

Currently, we are completing the human rating portion of building the mtCNN for rating technical skill in RPN. Preliminary data has been promising for building our mtCNN. Our primary aim is that our mtCNN will be able to reliably and accurately score surgeon technical skill in RPN, generating similar OSATS and GEARS scores to those that would be assigned by an appropriately trained human rater. Once built and validated, the final goal will be to create a surgeon report card to help improve surgeon technical skill.

**FUTURE DIRECTIONS:**

The next step in this project is the training of the mtCNN, using the human rated videos as ground truth and then validating the mtCNN with additional cases. This portion of the project will be carried out by out computer science/machine learning colleagues at SMU. Preliminary data has been shared with them, and we will continue to share data with them as more videos are rated. My work in the project will be ongoing, as the 34 videos rated thus far are only a small fraction of the several hundred cases we have available. We have also applied for funding to the Intuitive Surgical Technology Grant. The results of this grant will be known later this year.

**REFERENCES:**

1. Fecso, A. B., Szasz, P., Kerezov, G. et al.: The Effect of Technical Performance on Patient Outcomes in Surgery: A Systematic Review. Ann Surg, **265:** 492, 2017
2. Birkmeyer, J. D., Finks, J. F., O'Reilly, A. et al.: Surgical skill and complication rates after bariatric surgery. N Engl J Med, **369:** 1434, 2013
3. Niitsu, H., Hirabayashi, N., Yoshimitsu, M. et al.: Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. Surg Today, **43:** 271, 2013
4. Goh, A. C., Goldfarb, D. W., Sander, J. C. et al.: Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. J Urol, **187:** 247, 2012

UTSouthwestern
Medical Center

**APPENDIX A:** Objective Structured Assessment of Technical Skills (OSATS) Rating Scale [3]

## GLOBAL RATING SCALE OF OPERATIVE PERFORMANCE

Please circle the number corresponding to the candidate's performance in each category, irrespective of training level

| | | | | |
|---|---|---|---|---|
| **Respect for Tissue :** | | | | |
| 1<br>Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments | 2 | 3<br>Careful handling of tissue but occasionally caused inadvertent damage | 4 | 5<br>Consistently handled tissue appropriately with minimal damage |
| **Time and Motion :** | | | | |
| 1<br>Many unnecessary moves | 2 | 3<br>Efficient time/motion but some unneccessary moves | 4 | 5<br>Clear economy of movement and maximum efficiency |
| **Instrument Handling :** | | | | |
| 1<br>Repeatedly makes tentative or awkward moves with instruments by inappropriate use of instruments | 2 | 3<br>Competent use of instruments but occasionally appeard stiff or awkward | 4 | 5<br>Fluid moves with instruments and no awkwardness |
| **Knowledge of Instruments :** | | | | |
| 1<br>Frequently asked for wrong instrument or used inappropriate instrument | 2 | 3<br>Knew names of most instruments and used appropriate instrument | 4 | 5<br>Obviously familiar with the instruments and their names |
| **Flow of Operation :** | | | | |
| 1<br>Frequently stopped operating and seemed unsure of next move | 2 | 3<br>Demonstrated some forward planning with reasonable progression of procedure | 4 | 5<br>Obviously planned course of operation with effortless flow from one move to the next |
| **Use of Assistants :** | | | | |
| 1<br>Consistently placed assistants poorly or failed to use assistants | 2 | 3<br>Appropriate use of assistants most of time | 4 | 5<br>Strategically used assistants to the best advantage at all time |
| **Knowledge of Specific Procedure :** | | | | |
| 1<br>Deficient knowledge. Needed specific instruction at most steps | 2 | 3<br>Knew all important steps of operation | 4 | 5<br>Demonstrated familiarity with all aspects of operation |

UTSouthwestern
Medical Center

**APPENDIX B:** Global Evaluative Assessment of Robotic Skills (GEARS) Rating Scale [4]

| | | | | |
|---|---|---|---|---|
| **Depth perception** | | | | |
| 1 | 2 | 3 | 4 | 5 |
| Constantly overshoots target, wide swings, slow to correct | | Some overshooting or missing of target, but quick to correct | | Accurately directs instruments in the correct plane to target |
| **Bimanual dexterity** | | | | |
| 1 | 2 | 3 | 4 | 5 |
| Uses only one hand, ignores nondominant hand, poor coordination | | Uses both hands, but does not optimize interaction between hands | | Expertly uses both hands in a complementary way to provide best exposure |
| **Efficiency** | | | | |
| 1 | 2 | 3 | 4 | 5 |
| Inefficient efforts; many uncertain movements; constantly changing focus or persisting without progress | | Slow, but planned movements are reasonably organized | | Confident, efficient and safe conduct, maintains focus on task, fluid progression |
| **Force sensitivity** | | | | |
| 1 | 2 | 3 | 4 | 5 |
| Rough moves, tears tissue, injures nearby structures, poor control, frequent suture breakage | | Handles tissues reasonably well, minor trauma to adjacent tissue, rare suture breakage | | Applies appropriate tension, negligible injury to adjacent structures, no suture breakage |
| **Autonomy** | | | | |
| 1 | 2 | 3 | 4 | 5 |
| Unable to complete entire task, even with verbal guidance | | Able to complete task safely with moderate guidance | | Able to complete task independently without prompting |
| **Robotic control** | | | | |
| 1 | 2 | 3 | 4 | 5 |
| Consistently does not optimize view, hand position, or repeated collisions even with guidance | | View is sometimes not optimal. Occasionally needs to relocate arms. Occasional collisions and obstruction of assistant. | | Controls camera and hand position optimally and independently. Minimal collisions or obstruction of assistant |