Deployment and implementation of a multi-service surgical case length prediction model

Hamed Zaribafzadeh, MS[1,2] Wendy Webster, MA, MBA[1] Christopher Vail, PA-C, MMCi, CSM[1] Tom Daigle, BA[1] Allan Kirk, MD, PhD[1] Peter Allen, MD[1] Ricardo Henao, PhD[2] Daniel Buckland, MD, PhD[3,4]

1. Department of Surgery, School of Medicine, Duke University, USA; 2. Department of Biostatistics and Bioinformatics, School of Medicine, Duke University, USA; 3. Department of Emergency Medicine, School of Medicine, Duke University, USA; 4. Department of Mechanical Engineering and Materials Science, Pratt School of Engineering, Duke University, USA.

Introduction

The operating room (OR) is one of the most expensive resources in the healthcare system; estimated to cost $22-133 per min and generating about 40% of hospital revenue[1-3]. Accurate prediction of surgical case length could improve OR utilization efficiency. Many hospitals use surgeon-estimated, historical median, or Electronic Health Record (EHR) system-generated median case length to schedule surgical cases, which all have shown to be inaccurate[2,4-9]. Many groups have developed machine-learning models and show their potential use for case length prediction[2,4]. To the best of our knowledge, this is the first report on wide implementation of a machine learning model to predict case length at the time of case creation for various services across different hospitals and ambulatory surgery centers.

Approach

It has been shown that surgeons are the most significant contributor to the case length variability[8]. We introduce a method, called *similarity cascade*, to capture case complexity due to primary surgeon and procedural factors, calculate the historical median case length, and make the model implicitly dependent on surgeon-name by replacing 1000+ surgeon names with a single variable. 107,898 elective surgical cases across Duke University Health System (DUHS) from Jan 2021 to Apr 2022 were collected for training and evaluation. Using the *similarity cascade* approach and limited case posting data, we created and evaluated a machine-learning model with a custom loss function to predict case length which outperforms schedulers and EHR system median (data not shown). To create a production pipeline (Figure 1) and seamlessly deploy and implement our model, we collaborated with several groups within DUHS to 1) create an application programming interface (API) providing case data one day after posting; 2) build the framework to run the model on Kubernetes; and 3) integrate an additional field into our EHR system to show the model output for schedulers use (if they choose so) with the process and pipeline were subject to oversight from an internal DUHS governance and evaluation committee. After deployment and several rounds of silent evaluation, the model was generally implemented on August 1st, 2022.

Prospective analysis

33,815 surgical cases from Aug-Dec 2022 were collected. The schedulers were free to adjust the model output. Figure 2A shows the gradual adoption of the model by schedulers as the gap between the schedulers and model predicted case length has been shrinking. The model performed more balanced in terms of under- and over-prediction error and predicted 5.9% more cases within 20% of the actual case length, our defined arbitrary error margin, compared with the schedulers (Figure 2B). Specifically, the model significantly outperformed the schedulers for short cases (*i.e.*, ≤ 60 minutes) showing 18.8% and 13% more of the half-an-hour- and an-hour-long predicted cases within the error margin, respectively (Figure 2C-D).

Conclusion

We created a machine-learning model and a production pipeline that is being leveraged every day to provide the schedulers with the predicted case length in the EHR system at the time of scheduling. This framework could be utilized, and modified if needed, to deploy future models across our institute.

References

1 Bellini, V. *et al.* Artificial Intelligence: A New Tool in Operating Room Management. Role of Machine Learning Models in Operating Room Optimization. *J Med Syst* **44**, 20 (2019). https://doi.org:10.1007/s10916-019-1512-1

2 Jiao, Y., Sharma, A., Ben Abdallah, A., Maddox, T. M. & Kannampallil, T. Probabilistic forecasting of surgical case duration using machine learning: model development and validation. *J Am Med Inform Assoc* **27**, 1885-1893 (2020). https://doi.org:10.1093/jamia/ocaa140

3 Ito, M. *et al.* Does case-mix classification affect predictions? A machine learning algorithm for surgical duration estimation. *Healthcare Analytics* **2** (2022). https://doi.org:10.1016/j.health.2022.100119

4 Zhao, B., Waterman, R. S., Urman, R. D. & Gabriel, R. A. A Machine Learning Approach to Predicting Case Duration for Robot-Assisted Surgery. *J Med Syst* **43**, 32 (2019). https://doi.org:10.1007/s10916-018-1151-y

5 Rozario, N. & Rozario, D. Can machine learning optimize the efficiency of the operating room in the era of COVID-19? *Can J Surg* **63**, E527-E529 (2020). https://doi.org:10.1503/cjs.016520

6 Strömblad, C. T. *et al.* Effect of a Predictive Model on Planned Surgical Duration Accuracy, Patient Wait Time, and Use of Presurgical Resources. *JAMA Surgery* **156** (2021). https://doi.org:10.1001/jamasurg.2020.6361

7 Robertson, A., Kla, K. & Yaghmour, E. Efficiency in the operating room: optimizing patient throughput. *Int Anesthesiol Clin* **59**, 47-52 (2021). https://doi.org:10.1097/AIA.0000000000000333

8 Bartek, M. A. *et al.* Improving Operating Room Efficiency: Machine Learning Approach to Predict Case-Time Duration. *J Am Coll Surg* **229**, 346-354 e343 (2019). https://doi.org:10.1016/j.jamcollsurg.2019.05.029

9 Tuwatananurak, J. P. *et al.* Machine Learning Can Improve Estimation of Surgical Case Duration: A Pilot Study. *J Med Syst* **43**, 44 (2019). https://doi.org:10.1007/s10916-019-1160-5

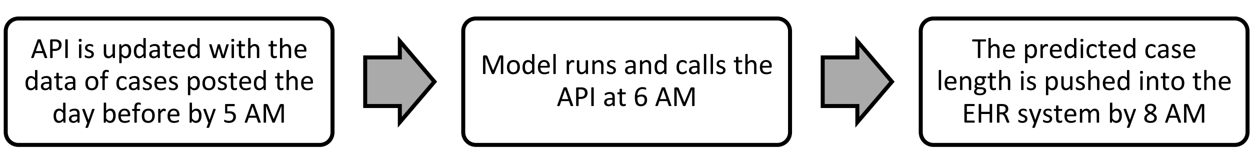

Figure 1. General workflow of the production pipeline running the case length model everyday

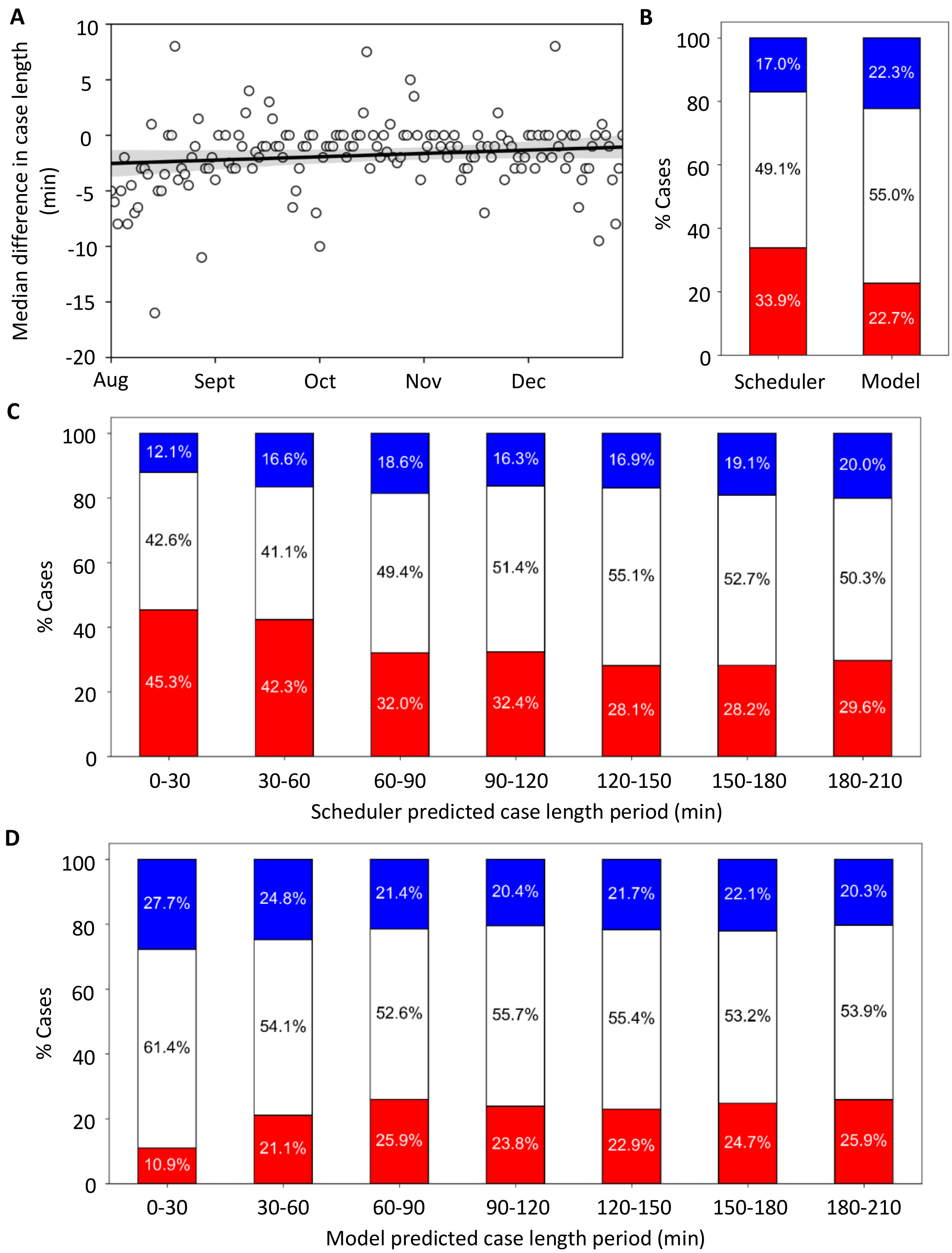

Figure 2. Prospective evaluation of the case length model during Aug-Dec 2022. A) Median difference in predicted case length by the scheduler and model in minutes, B) Overall performance comparison between scheduler and model, C-D) scheduler and model performance at different predicted case

periods, respectively. Predicted case lengths within, under, and over 20% of the actual case length are depicted in white, red, and blue, respectively.