Advances in Informatics, Management and Technology in Healthcare
J. Mantas et al. (Eds.)

© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI220789

A Machine Learning Solution to Predict Elective Orthopedic Surgery Case Duration

Divya SAHADEV[a,1] Thomas LOVEGROVE[b] and Holger KUNZ[a]
[a] *University College London, Institute of Health Informatics*
[b] *East Kent Hospitals University NHS Foundation Trust*

Abstract. We used surgery durations, patient demographic and personnel data taken from the East Kent Hospitals University NHS Foundation Trust (EKHUFT) over a period of 10 years (2010-2019) for a total of 25,352 patients that underwent 15 highest volume elective orthopedic surgeries, to predict future surgery durations for the subset of elective surgeries under consideration. As part of this study, we compared two different ensemble machine learning methods random forest regression (RF) and XGBoost (eXtreme Gradient Boosting) regression. The two models were approximately 5% superior to the existing model used by the hospital scheduling system.

Keywords. Machine learning; Predictive Modelling; Surgery Case Duration

1. Introduction

Covid-19 has amplified the weaknesses in health care systems across the world. It has become swiftly apparent that regardless of the National Health Service (NHS) being a respected health care system, it was severely underprepared to tackle the challenges posed by the pandemic (1). UK has seen a growing decline over the years in the performance for elective surgeries (1). Despite the number of patients being treated, there is an increase in patients who have already waited for over a year for treatment (2).

The aim of this study is to use surgery durations, patient demographic and personnel data taken from the 'East Kent Hospitals University NHS Foundation Trust' (EKHUFT) over a period of 10 years (2010-2019) for a total of 25,352 patients that underwent 15 highest volume elective orthopedic surgeries, to predict future surgery durations for the subset of elective surgeries.

2. Methods

We used surgery durations, patient demographic and personnel data taken from the 'East Kent Hospitals University NHS Foundation Trust' over a period of 10 years (2010-2019) for a total of 25,352 patients that underwent 15 highest volume elective orthopedic surgeries, to predict future surgery durations for the subset of elective surgeries under consideration. Each record included the age band that the patient belonged to, ethnicity, sex, code for the anesthetic type used, theatre code, intended management (planned or

[1] Corresponding Author, Holger Kunz <h.kunz@ucl.ac.uk>

day case), multiple deprivation index score (IMD), day of the week, pseudonymized codes for the surgeon, anesthetist and consultant, primary procedure code associated with the present case and history of long-term illnesses such as cardiovascular disease, diabetes, hypertension, obesity, cancer, and chronic kidney disease. The study dataset also included temporal variables such as expected length and actual length of the procedure. The expected length is calculated by the current software system using historical surgery times by surgeon and procedure, which could be overwritten by the surgeon.

Ensemble tree-based algorithms were chosen due to their ease in interpretation, speed of construction, ability to deal with categorical and continuous variables, robust to outliers and missing data and their ability to model non-linear relationships (3). As part of this study, we compared two different tree-based ensemble machine learning methods random forest regression and XGBoost (eXtreme Gradient Boosting) regression. The cost function was customized to penalize underestimating of surgical times. For the XGB model a piecewise asymmetric cost function that penalizes predicted time less than actual time (underestimation) more than an overestimation was applied. The intended cost function will penalize both over and underestimation. After 30 minutes, an overestimation is treated the same as an underestimation. The training setup and hyperparameter tuning was conducted for the Scikit-learn Random Forest (RF) regressor package and the XGBoost package. Grid search was applied to identify the best parameters for both models.

3. Results

Both models performed similarly in comparison to the current hospital model. Both RF and XGB models do not have corresponding predictions for case durations over 160 minutes, which can be seen with the expected length. The density plot in Figure 1 shows that the XGB model is slightly biased towards the positive, as compared to the RF model. This is due to the asymmetric loss function and shows that the model is performing as expected. RF model shows a higher peak than XGB and has a slightly lesser variance, therefore more centered at 0, which could be explained by the smaller grid space explored for XGB.

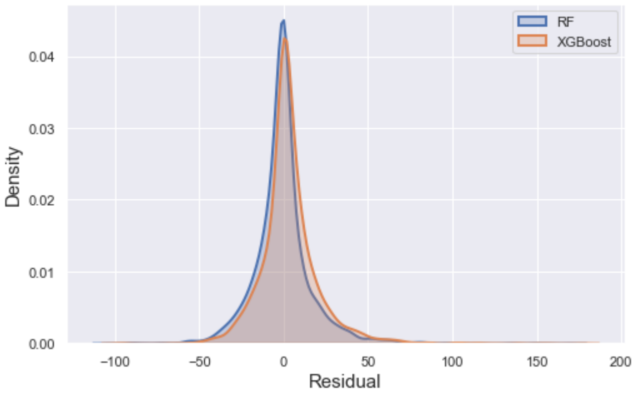

Figure 1. Kernel density of residuals for (A) RF model and (B) XGB model

4. Discussion

The study uses real-world data over a period of 10 years from a hospital trust of patients. The strength of the study is the inclusion of a customized penalization for under-estimation vs. over-estimation. Usually, evaluation metrics for regression ignore the direction of the prediction. However, in clinical practice it is of vital importance to take the direction of the prediction into account as another surgical procedure is waiting in the working pipeline of an operating theatre. Both RF and XGB resulted in similar prediction accuracy. Even though both models are based on decision trees and performed similarly, the XGB model is more computationally efficient and therefore better suited for more widespread, real-time implementation (4).

The weakness of the study is a focus only on the ten most frequent surgeries as the sample size for highly specialized surgeries is too small to justify an inclusion in the prediction. Also, revision surgeries performed at the exact same anatomical location was not considered as part of this study. This could result in certain elective procedures having an extended duration. Hence, for future studies it would be beneficial to exclude revision surgeries from the data to model. Overall, the improved predictive power demonstrated in this study could lead to decreases in under and over utilization of operating room resources. Future work of this study will look on the optimization of operating theatre planning and how the developed prediction algorithms can be extended to other surgeries in general surgery and urology. Relevant data including but not limited to personal identifiable information was anonymized. The NHS Trust internal research ethics approved this study.

5. Conclusion

In summary, our study demonstrates that predictive modelling using ML (machine learning) is more accurate than the current method to predict case duration, and that implementation of a ML algorithm can lead to an increase in the number of accurately scheduled case durations.

References

[1] Macdonald N, Clements C, Sobti A, Rossiter D, Unnithan A, Bosanquet N. Tackling the elective case backlog generated by Covid-19: the scale of the problem and solutions. Journal of Public Health. 2020 Nov 23;42(4):712–6.

[2] Carr A, Smith JA, Camaradou J, Prieto-Alhambra D. Growing backlog of planned surgery due to covid-19. BMJ. 2021 Feb 9;372:n339..

[3] Murphy KP. Machine Learning: A Probabilistic Perspective. The MIT Press. 2012.

[4] Bartek MA, Saxena RC, Solomon S, Fong CT, Behara LD, Venigandla R, et al. Improving Operating Room Efficiency: Machine Learning Approach to Predict Case-Time Duration. In: Journal of the American College of Surgeons. 2019. p. 346-354.e3.