Predicting Elective Orthopedic Surgical Appointment Cancellations Using a National Machine Learning Model

Hady S Salama, Alison K Klika, Marc C Jacofsky, David Joseph Jacofsky, Wael K Barsoum

INTRODUCTION:

Cancellations of elective orthopedic procedures waste operating room capacity, delay care, and raise costs. We hypothesized that a machine learning model trained on large-scale scheduling data could accurately predict cancellations and outperform existing approaches that report median area under the curve (AUC) of 0.8. We developed and externally expanded a machine learning cancellation risk model and compared performance between a single hospital prototype and a national deployment.

METHODS:

Scheduling data for 337,200 elective orthopedic cases (Jan 2021 – Nov 2024) from 15 markets across the United States were extracted from a HIPAA and HITRUST compliant Snowflake ™ cloud data warehouse. Forty three core variables (demographics, clinical, temporal, utilization) from electronic health records (EHR) were engineered; categorical fields were one hot encoded. AutoGluon AutoML screened candidate algorithms, selecting gradient boosted trees (XGBoost). Bayesian hyperparameter optimization was performed on an 80% training set (n = 269,760). A 5 % “live evaluation” set (n = 16,860) monitored generalization during training, and an unseen 15% held out test set (n = 50,850) provided final metrics. 42,000 cases from a single orthopaedic specialty hospital were modeled and compared. Receiver operating characteristic curves and the AUC were used to evaluate the discriminatory performance of the model. SHapley Additive exPlanations (SHAP) quantified feature impact.

RESULTS:

The nationally representative all markets model had an AUC = 0.862 and accuracy = 0.766 (Table 1). Top SHAP drivers were prior cancellations, total future appointments, booking lead time, facility/business unit, patient BMI, workers comp financial class, travel distance, and calendar day of year (Figure 2). The single orthopaedic specialty hospital model had an AUC = 0.833 and accuracy = 0.751 (Table 1). Expansion to the national cohort therefore improved discrimination (+0.032 AUC) and maintained balanced performance.

DISCUSSION AND CONCLUSION:

A gradient boosted machine learning model trained on routine EHR scheduling data accurately predicts elective orthopedic surgery cancellations across diverse markets in the United States. Identifying high risk cases enables proactive outreach, tighter block time management, and reduced late cancellations. Deployment could enable proactive outreach to high risk patients, reduce day of surgery cancellations, and improve operating room utilization. Prospective implementation studies and integration into scheduling workflows are warranted.

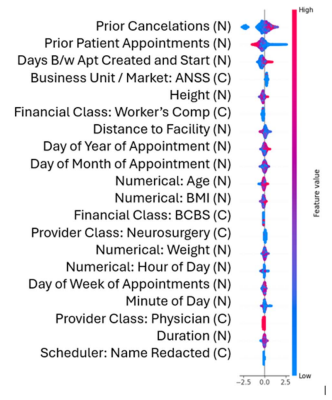

Figure 1. Main SHapley Additive exPlanations (SHAP) summary. The cluster of dots represents the polarity of the feature impact on the final prediction. -2.5 to 2.5 are the SHAP values. Positive SHAP values, to the right, indicate an impact towards the positive class in this case 1 or true for cancelation and negative SHAP values, to the left, indicate impact towards the negative class, in this case 0 or false for non-cancelations. Color indicates the strength of the impact on the final prediction. C indicates categorical variable and N indicates numerical variable.

Model Type	ROC-AUC	Accuracy	Precision	Recall	F1
All Markets National Model	0.862	0.761	0.78	0.77	0.77
Single Orthopaedic Specialty Hospital Model	0.833	0.751	0.76	0.75	0.75