Personalized Risk Stratification for Total Knee Arthroplasty: Insights from Machine Learning Models

Jibreel Hussain, Jayalakshmi Jain, Apurva Swapnil Choubey, Nezar Abunnur, Saif Khan, Yasser Farid, Mark H Gonzalez

INTRODUCTION: As the demand for total knee arthroplasty (TKA) continues to grow, so does the need for precise tools to predict which patients are at risk for poor outcomes. Traditional risk calculators often fall short in capturing the complexity of surgical patients, but machine learning (ML) algorithms offer a powerful alternative. This study aimed to evaluate multiple ML algorithms for their ability to predict short-term outcomes after primary TKA, and to identify the most influential preoperative factors associated with each outcome. The ultimate goal is to develop an interpretable, clinically useful tool to support risk stratification and optimization in TKA patients.

METHODS: Patients undergoing primary TKA were identified using the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) database from 2019 to 2023 using CPT code 27447. After applying standard exclusions, 298,828 patients were included. Four supervised ML models—XGBoost, LightGBM, Random Forest, and Elastic Net Logistic Regression (EN Logistic Regression)—were trained to predict four outcomes: 30-day readmission, major complications, prolonged length of stay (LOS ≥3 days), and non-home discharge. An ensemble stacking model combining these four was also constructed. Data were split 75:25 for training and testing. Model performance was assessed using area under the receiver operating characteristic curve (AUROC). Feature importance was determined using mean SHAP (SHapley Additive exPlanations) values derived from the most predictive model for each outcome. For major complications, further analysis was performed on the top four SHAP-ranked features using dependence plots to evaluate trends in model output based on specific patient characteristics.

RESULTS:

The ensemble model outperformed the individual algorithms for all outcomes, achieving AUROC values of 0.85 for 30-day readmission, 0.79 for major complications, 0.73 for prolonged LOS, and 0.79 for non-home discharge. The most influential patient characteristics differed among outcomes. For 30-day readmission, the top predictors were age, ASA classification, sex, hematocrit, and white blood cell count, respectively. Prolonged LOS was predominantly influenced by race, followed by age, sex, hematocrit, and anesthetic technique. Non-home discharge predictions were primarily driven by age, ASA classification, ethnicity, and sex.

The top four predictors for major complications, which included hematocrit, age, ASA classification, and sex, were further examined using SHAP dependence plots. Hematocrit levels below 35% were associated with increased model-predicted complication risk, while levels of at least 40% had a negative SHAP contribution. Patients on the extremes of age, those younger than 40 or older than 80, exhibited elevated SHAP values indicating increased risk, while intermediate age ranges showed lower, protective SHAP values. Additionally, an ASA classification of III corresponded to greater predicted risk compared to ASA classes I or II. Female patients showed modestly increased SHAP values, indicating slightly elevated risk relative to male counterparts.

DISCUSSION AND CONCLUSION: An ensemble ML model trained on NSQIP data accurately predicted short-term adverse outcomes after TKA, with highest performance observed for non-home discharge and readmission. Feature importance analysis showed that different sets of preoperative factors were predictive for each outcome, underscoring the importance of outcome-specific risk modeling. Hematocrit, age, and ASA classification consistently emerged as key predictors across multiple postoperative outcomes, highlighting their central role in patient risk profiling. For major complications, detailed SHAP analysis revealed clear directional associations, such as increased risk with lower hematocrit levels, advanced age, and higher ASA classification, offering clinically interpretable insights into risk. These findings not only validate the use of machine learning for outcome-specific modeling but also provide a framework for tailoring preoperative optimization strategies. To facilitate real-world application, we developed a clinical calculator based on our ensemble model that allows providers to input individual patient data and receive personalized risk estimates for each outcome.