Precision Risk Modeling in Total Hip Arthroplasty: A Machine Learning Framework for Personalized Outcome Prediction

Jayalakshmi Jain, Jibreel Hussain, Nezar Abunnur, Apurva Swapnil Choubey, Esa Syed, Malyika F Hussain, Yasser Farid, Mark H Gonzalez

INTRODUCTION: As total hip arthroplasty (THA) volumes continue to rise, accurately identifying patients at risk for adverse postoperative outcomes is increasingly important. Traditional risk calculators often lack the nuance to capture complex, multifactorial relationships. Machine learning (ML) models offer a promising alternative by leveraging large datasets to identify predictive patterns. This study evaluates multiple ML algorithms for forecasting short-term outcomes following primary THA and identifies the most impactful preoperative variables contributing to each. The goal is to build an interpretable and clinically applicable decision-support tool to aid in perioperative planning and optimization.

METHODS: We queried the National Surgical Quality Improvement Program (NSQIP) database from 2019–2023 to identify patients who underwent primary THA (CPT 27130). After applying standard exclusion criteria, a cohort of 199,243 patients was analyzed. Four supervised ML models—XGBoost, LightGBM, Random Forest, and Elastic Net Logistic Regression—were developed to predict four 30-day outcomes: hospital readmission, major complications, prolonged length of stay (LOS ≥3 days), and discharge to a non-home facility. A stacked ensemble model combining predictions from these base learners was also constructed. The dataset was split 75% for training and 25% for testing. Predictive performance was assessed using the area under the receiver operating characteristic curve (AUROC). Feature importance was assessed using mean SHapley Additive exPlanations (SHAP) values from the best-performing model for each endpoint. For major complications, additional SHAP dependence plots were generated for the top four features to illustrate how individual patient factors influenced model predictions.

RESULTS:

The ensemble model consistently achieved the highest AUROC across all outcomes: 0.80 for readmission, 0.76 for major complications, 0.77 for prolonged LOS, and 0.82 for non-home discharge. Predictive factors varied by outcome. Readmission was most strongly associated with age, ASA classification, BMI, hematocrit, and anesthesia technique. For prolonged LOS, hematocrit was the top contributor, followed by anesthesia type, age, race, and ASA status. Non-home discharge was best predicted by hematocrit, platelet count, anesthesia technique, ethnicity, and BMI.

For major complications, the leading features were hematocrit, age, ASA classification, and anesthesia method. SHAP dependence plots demonstrated that hematocrit levels below 40% were associated with increased predicted complication risk. Advanced age (≥70) and ASA Class III also contributed to elevated risk. Similarly, general or epidural anesthesia was linked to higher SHAP values, suggesting these techniques may correlate with greater complication risk compared to other approaches.

DISCUSSION AND CONCLUSION: This study demonstrates that ensemble machine learning models can effectively predict a range of short-term postoperative outcomes after primary THA. Hematocrit, age, and ASA classification consistently emerged as critical predictors, though their relative influence varied by outcome. SHAP analysis offered interpretable insights, revealing that lower hematocrit, older age, and higher ASA status increase risk for major complications. These results support the value of outcome-specific modeling and underscore the clinical utility of ML-based tools for preoperative risk stratification. To support integration into practice, we developed a user-friendly risk calculator using the ensemble model, enabling personalized outcome predictions based on individual patient profiles.