

# Personalized Risk Stratification for Total Shoulder Arthroplasty with Machine Learning

Jibreel Hussain MBA<sup>1</sup>, Jayalakshmi Jain MS<sup>1</sup>, Nezar Abunnur BS<sup>1</sup>, Apurva Choubey MD<sup>2</sup>, Benjamin A. Goldberg MD<sup>2</sup>, <sup>1</sup>University Illinois College of Medicine, Chicago, IL, <sup>2</sup>University of Illinois at Chicago, Department of Orthopaedics, Chicago IL mhussa45@uic.edu

**Disclosures:** Jibreel Hussain MBA (N), Jayalakshmi Jain MS (N), Nezar Abunnur BS (N), Apurva Choubey MD (N), Benjamin A. Goldberg MD (N)

**Introduction:** Total shoulder arthroplasty (TSA) is an increasingly utilized surgical treatment for glenohumeral arthritis and rotator cuff arthropathy, offering significant improvements in pain and function. However, despite growing surgical volumes and advancements in technique, TSA continues to carry risk for short-term adverse outcomes such as complications, hospital readmission, and discharge to non-home settings. Traditional risk stratification tools often lack precision and fail to capture the nuances of individual patient profiles. This study applied machine learning (ML) models to predict short-term outcomes following TSA and identify the most influential preoperative variables driving each outcome. The goal was to develop an interpretable, outcome-specific tool to support personalized perioperative decision-making.

**Methods:** Patients undergoing TSA between 2019 and 2023 were identified in the National Surgical Quality Improvement Program (NSQIP) database using CPT code 23472. After applying standard exclusions, 27,525 patients were included for analysis. Four supervised ML algorithms—XGBoost, LightGBM, Random Forest, and Elastic Net Logistic Regression—were trained to predict four key 30-day outcomes: readmission, major complications, prolonged length of stay (LOS ≥3 days), and non-home discharge. An ensemble stacking model was also constructed by combining the predictions of all four base models. Data were randomly split into 75% training and 25% testing sets. Model performance was assessed using area under the receiver operating characteristic curve (AUROC). Feature importance was determined using mean SHapley Additive exPlanations (SHAP) values from the top-performing model for each outcome. For major complications, SHAP dependence plots were generated for the four most influential features to assess their specific impact on predicted risk.

**Results:** The ensemble model demonstrated the highest predictive performance for major complications (AUROC 0.76), prolonged LOS (AUROC 0.74), and non-home discharge (AUROC 0.85). For 30-day readmission, LightGBM outperformed all other models, including the ensemble, achieving an AUROC of 0.79. Key predictive factors varied across outcomes. For 30-day readmission, the most impactful variables were ASA classification, sex, platelet count, and age. Prolonged LOS was primarily influenced by race, ASA classification, hematocrit, and age. Predictions for non-home discharge were mainly driven by age, ASA classification, hematocrit, and race.

For major complications, the four most influential features were ASA classification, platelet count, age, and sex. SHAP analysis showed that patients classified as ASA III had significantly higher predicted complication risk compared to ASA I or II. Platelet counts below 170,000 were associated with markedly elevated risk, and patients over 50 exhibited progressively increasing SHAP values for complication risk. Female sex was also associated with a modest increase in predicted risk.

**Conclusion:** Machine learning models effectively predicted short-term adverse outcomes following total shoulder arthroplasty, with ensemble models performing best for most endpoints and LightGBM excelling in readmission prediction. Feature importance analysis revealed that different sets of preoperative factors influenced each outcome, emphasizing the need for outcome-specific risk modeling. Key risk indicators—including ASA classification, platelet count, age, and hematocrit—consistently emerged across models. SHAP analysis provided interpretable insights, confirming associations between higher risk and factors such as advanced age, lower platelet count, and higher ASA class. These findings support the use of ML-driven tools to guide individualized surgical planning and preoperative counseling. A clinical risk calculator was developed based on the ensemble model to provide personalized risk estimates for TSA patients.

**Significance/Clinical Relevance:** TSA is increasingly used to treat degenerative shoulder conditions, but risk profiles vary widely between patients. This study supports clinical decision-making through predictive modeling of post-surgical risk and recovery trajectories.

