

Validation of Machine Learning Models in Predicting Blood Transfusion Following Simultaneous Bilateral Primary Total Knee Arthroplasty

Muhammad Hamza Ilyas, MD¹; Oh-Joon Kwon¹; Pengwei Xiao, PhD¹; William Sampson, BA¹; Isaiah Freeman, BSc¹; Hahn Kang, BSc¹; Michelle Shimizu, MD¹; Carlo Mannina, BSc¹; Oh-Jak Kwon¹; Sina Afzal, MD¹; Zhijun Li, MD, PhD¹; Margaret Arthur¹; Young-Min Kwon, MD, PhD¹
¹Bioengineering Laboratory, Massachusetts General Hospital/Harvard Medical School, Boston, MA
 ymkwon@mgh.harvard.edu

Disclosures: Muhammad Hamza Ilyas (N), Oh-Joon Kwon (N), William Sampson (N), Isaiah Freeman (N), Hahn Kang (N), Michelle Shimizu (N), Carlo Mannina (N), Oh-Jak Kwon (N), Sina Afzal (N), Zhijun Li (N), Margaret Arthur (N), Young-Min Kwon (5- MicroPort; 5- DePuy; 5-Smith & Nephew; 5- Stryker; 5- Zimmer Biomet; 5- Medacta)

INTRODUCTION: Approximately 6% of total knee arthroplasties (TKAs) performed in the United States are simultaneous bilateral procedures, which offer potential perioperative efficiencies but are associated with a high incidence of blood transfusion, reported to be as high as 29%. Blood transfusion, in turn, is associated with increased risks of complications, prolonged hospitalization, and mortality, underscoring the need for accurate preoperative risk stratification. Therefore, this study aims to develop machine learning models to predict postoperative transfusion risk following simultaneous bilateral TKA using a large national dataset.

METHODS: 12,977 patients who underwent simultaneous primary TKA between 2010 and 2023 were identified by the Current Procedural Terminology (CPT) code 27447 within the American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) database. The ACS-NSQIP is a national database containing preoperative and 30-day postoperative outcome data from participating academic and community institutions with a high interrater reliability of >95%. The outcome of interest was the occurrence of blood transfusion within 72 hours after surgery. Demographic, preoperative, intraoperative, and immediate postoperative outcomes were collected. Patients under 18 years old and those with missing data were excluded. Five machine learning models, artificial neural network (ANN), k-nearest neighbor (KNN), histogram gradient boosting (HGB), and random forest (RF), were developed and their performance was assessed and compared using discrimination, calibration, and decision curve analysis.

RESULTS SECTION: In the final study cohort, 1,547 of 12,977 (11.9%) patients required a postoperative blood transfusion. All ML models demonstrated good performance on the testing dataset when predicting the occurrence of blood transfusion within 72 hours postoperatively, with AUCs ranging from 0.85 to 0.68, calibration slopes between -0.03 and 0.21, and Brier scores between 0.02 and 0.05 (Table 1). Random forest yielded the best prediction accuracy for transfusion, achieving an AUC of 0.85, a calibration slope of -0.03, a calibration intercept of 0.11, and a Brier score of 0.02. The strongest predictors were pre-operative hematocrit, pre-operative blood urea nitrogen (BUN), and pre-operative international normalized ratio (INR) (Figure 1).

DISCUSSION: This study developed machine learning models with strong predictive capability for postoperative blood transfusion following simultaneous bilateral TKA, while also identifying key preoperative risk factors. The models demonstrated good predictive accuracy across established performance metrics. Early and accurate identification of high-risk patients can facilitate targeted preventive strategies, including optimized perioperative management, tailored monitoring, and individualized postoperative care plans. Integration of these models into clinical workflows may enhance patient safety, reduce transfusion-related complications, and improve overall surgical outcomes.

SIGNIFICANCE/CLINICAL RELEVANCE: Machine learning models were able to successfully predict postoperative blood transfusion in patients that underwent simultaneous bilateral primary total knee arthroplasty, demonstrating potential clinical utility in perioperative risk assessment.

Metric	RF	ANN	HGB	KNN
AUC	0.85	0.7	0.78	0.68
Slope	-0.03	0.08	0.21	0.08
Intercept	0.11	0.04	0.04	0.01
Brier Score	0.02	0.03	0.02	0.05

Table 1. Discrimination and calibration of all machine learning models in predicting post-operative occurrence of blood transfusion following simultaneous bilateral knee arthroplasty.

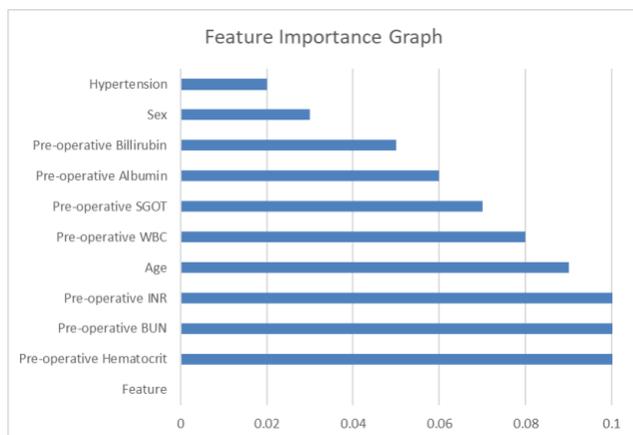


Figure 1. Relative global variable importance plot for the prediction of blood transfusion occurrence following simultaneous bilateral knee arthroplasty.