Machine Learning Models Effectively Predict Extended Length of Stay Following Total Shoulder Arthroplasty

Steven Haochen Liu, Christian J Leonardo, Rachel Anna Loyst, David E Komatsu[1], Edward D Wang[2]
[1]Stony Brook University, [2]Stony Brook Univ Schl of Med

INTRODUCTION: An effective predictive tool to identify patients at risk of extended length of stay (LOS) following total shoulder arthroplasty (TSA) may aid physicians in improving surgical outcomes, mitigate adverse hospital-associated complications, and cut costs of care. The objective of this study was to develop and internally validate 5 different machine learning (ML) models for the prediction of short versus extended LOS following TSA.

METHODS: Current Procedural Terminology (CPT) codes were used to identify a total of 14,298 patients from the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database who underwent TSA: 12,020 who had short LOS and 2,278 who had extended LOS. Variables collected in this study included patient demographics, comorbidities, preoperative laboratory studies, and surgical characteristics. Five supervised machine learning models were developed: artificial neural network (ANN), K-nearest neighbor (KNN), elastic-net penalized logistic regression (NEPLR), histogram-based boosting (HGB), and random forest (RF). Models were trained using patient data and evaluated based on discrimination measured using the area under the receiver operating curve (AUC).

RESULTS: Four of 5 ML models had an AUC > 0.8 and therefore showed excellent discrimination (Figure 1, Table 1):): ANN (AUC [95% CI] = 0.818 [0.812, 0.824]), NEPLR (AUC [95% CI] = 0.802 [0.797, 0.807]), HGB (AUC [95% CI] = 0.808 [0.803, 0.814]), and RF (AUC [95% CI] = 0.815 [0.808, 0.821]). The KNN model (AUC [95% CI] = 0.740 [0.734, 0.746]) showed acceptable discrimination, but significantly worse discrimination compared to the other models. For the ANN model, which had the highest AUC, the most important features in predicting extended LOS following TSA were a longer operation time, lower preoperative hematocrit, older age, and TSA performed due to a fracture (Figure 2).

For the ANN model, individual patient-level explanations were generated (Figure 3).

Figures 3a and 3b illustrate individual patients for whom the ANN model predicted a 4.58% chance and 87.0% chance of experiencing extended LOS, respectively, based on preoperative characteristics. Blue and red bars illustrate the relative weights of protective and risk factors as determined by the ANN model for the patient.

DISCUSSION AND CONCLUSION: This study developed and validated 5 ML models capable of predicting preoperative factors that are associated with extended LOS following TSA. Four of five models showed excellent model performance based on discrimination. Models also demonstrated an ability to explain an individual patient’s risk of having an extended LOS by identifying risk factors and protective factors as well as the weights of these factors. These results are promising for the potential use of artificial intelligence in assisting with identifying strong surgical candidates for TSA.