

Predicting 12-Week Knee Flexion After Total Knee Arthroplasty: An Interpretable and Parsimonious Machine Learning Model

Purpose/Hypothesis: We developed a machine learning model to predict 12-week knee flexion angle (KFA) using routinely collected preoperative and intraoperative measurements. We hypothesized that a parsimonious model would identify an optimal feature set balancing predictive accuracy with clinical interpretability.

Methods: We retrospectively analyzed 192 primary TKA patients. We used data-driven optimization including Shapley Additive exPlanations, for feature selection among 21 candidate variables. A Ridge regression model ($\alpha=10.0$, optimized via 10-fold cross-validation) was developed and internally validated using 1,000 bootstrap iterations. Model performance was evaluated using regression metrics (R^2 , root mean squared error [RMSE], mean absolute error [MAE]) and classification performance (area under the curve [AUC]) for clinically meaningful thresholds (unrestricted KFA $\geq 115^\circ$, stiffness $< 100^\circ$). A random 80/20 train-test split was maintained across all analyses.

Results: We identified 4 optimal predictors: intraoperative post-arthrotomy KFA (SHAP = 2.14), preoperative KFA (SHAP = 2.05), body mass index (BMI), and age. The 4-feature model achieved test $R^2 = 0.283$ (95% CI: 0.13-0.35). RMSE was 7.8° (95% CI: 6.9° - 8.9°), and MAE was 6.3° (95% CI: 5.4° - 7.2°) for the KFA. Classification performance was robust, with AUC = 0.844 (95% CI: 0.782-0.895) for predicting unrestricted KFA and AUC = 0.935 (95% CI: 0.878-0.973) for identifying stiffness risk. Cross-validation demonstrated superior stability compared to more complex models (CV $R^2 = 0.187 \pm 0.167$ vs. 0.087 ± 0.372 for 8-feature model).

Conclusions: This parsimonious 4-variable model reliably predicts 12-week KFA and identifies patients at risk for suboptimal outcomes. The model's reliance on routinely collected measurements and strong classification performance (AUC > 0.84) supports its potential clinical utility for preoperative counseling and surgical planning. External validation in independent cohorts is warranted.