Machine Learning–Based Prediction of 1-Year Harris Hip Score after Total Hip Arthroplasty: Analysis of 8,715 Cases Using SHAP

Katsuyama Y, Mihara M, Takayanagi S, Hirano M, Kumahara Y, Hirakawa K.
Department of Orthopaedic Surgery, Shonan Kamakura Joint Reconstruction Center, Kamakura, Japan

Introduction

Total hip arthroplasty (THA) is an established and effective treatment for various hip disorders. However, postoperative functional recovery and patient satisfaction vary widely, and even with identical surgical techniques and implants, 1-year postoperative outcomes are not uniform. Such variability is thought to arise from the complex interplay of multiple factors, including preoperative functional status, physical characteristics, joint range of motion, and disease pathology. Predicting postoperative outcomes based on preoperative factors is valuable for risk communication, planning postoperative rehabilitation strategies, and optimizing perioperative management. Recent advances in machine learning have made it possible to construct high-precision prognostic models that can extract multivariate and nonlinear relationships from large datasets(1-3). In this study, we analyzed 8,715 primary THA cases at our institution to develop an XGBoost model—a gradient boosting decision tree algorithm that builds an ensemble of weak learners to improve predictive accuracy—for predicting the 1-year postoperative Harris Hip Score (HHS). We evaluated the model's performance and analyzed feature contributions using SHAP （SHapley Additive exPlanations) values. Python (Python Software Foundation, America) was used for all analyses.

Methods

We retrospectively analyzed 8,715 primary THA cases performed at our institution between January 2011 and December 2023, for which 1-year postoperative HHS data were available. Exclusion criteria included revision THA, infection, and early interruption of rehabilitation due to major perioperative complications. Predictive features included:
・Patient demographics: age, sex, BMI, height, weight
・Preoperative functional assessment: preoperative HHS, preoperative hip range of motion (flexion, extension, abduction, adduction, external rotation, internal rotation)
・Disease characteristics: disease pattern (unilateral disease operated side (uni), bilateral disease operated side (bi-1), bilateral disease both sides completed(bi-2)), diagnosis (OA, RA, ON, RDC, others)
We employed XGBoost, a gradient boosting method, as the machine learning model. The target variable classified 1-year postoperative HHS into three categories: Good (≥80), Excellent (≥90), and Poor (<70). Model performance was evaluated using the area under the receiver operating characteristic curve (AUC) with 10-fold cross-validation. Feature importance was assessed via SHAP analysis to visualize influential factors for each prediction class.

Results

・Good group (HHS ≥80): AUC = 0.787. Top SHAP contributors were age, preoperative HHS, height, BMI, flexion range of motion, and external rotation range of motion. Younger age, higher preoperative HHS, taller stature, lower BMI, and greater ROM were associated with favorable outcomes (figure1, 2).
・Excellent group (HHS ≥90): AUC = 0.753. Major contributors included preoperative HHS, age, disease pattern, height, and flexion ROM. Particularly, unilateral disease cases demonstrated higher rates of excellent outcomes.
・Poor group (HHS <70): AUC = 0.731. Key predictors were preoperative HHS, height, age, and BMI. Lower preoperative HHS, shorter stature, older age, and higher BMI were strong predictors of poor outcomes (figure3).
In all prediction models, preoperative HHS showed the highest contribution, with ROM, physical characteristics, disease pattern, and diagnosis category also enhancing model accuracy.

Discussion

Using a large dataset of 8,715 cases, we developed a machine learning model that predicted 1-year postoperative HHS from preoperative factors with clinically applicable accuracy (AUC 0.73–0.79). Preoperative HHS was the strongest predictor across all classifications, reaffirming its direct link to prognosis. ROM, physical characteristics, and disease pattern also contributed, underscoring the comprehensive importance of preoperative evaluation. These findings can provide individualized prognostic information during preoperative counseling and guide risk stratification and rehabilitation planning(1-3).

Conclusion

An XGBoost-based machine learning model can classify patients into good, excellent, and poor outcome groups at 1 year post-THA using only preoperative clinical and physical factors, achieving high predictive accuracy. Future work will include external validation and multi-institutional comparisons to improve generalizability, aiming for real-time clinical implementation as a decision-support tool.
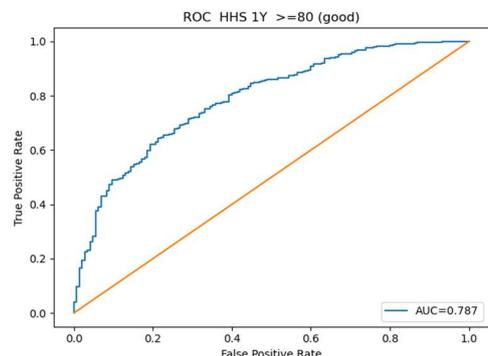

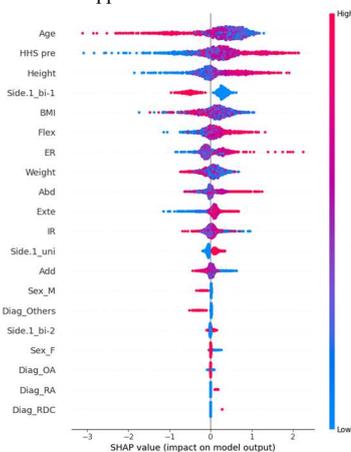Figure 1: ROC curve for predicting HHS ≥80 at 1 year (AUC=0.787).

References
1. Kunze KN, et al. The Journal of Arthroplasty. 2020;35:2119–2123.
2. Ramkumar PN, et al. The Journal of Arthroplasty. 2019;34:632-637.
3. Shin W, et al. IEEE J Biomed Health Inform. 2023;27(12):6074-6087.

Figure 2: Good group, HHS ≥80 : SHAP summary plot showing predictors of good outcome.
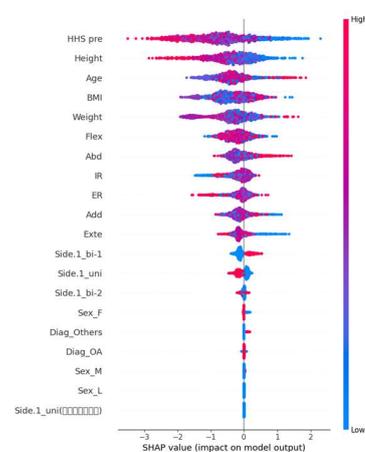
Figure 3: Poor group, HHS <70 : SHAP summary plot showing predictors of poor outcome.