

Predicting ACL Reconstruction Failure with Machine Learning

Development of Machine Learning Prediction Models

Rafael Krasic Alaiti,^{*†‡} PhD, Caio Sain Vallio,[§] PhD, Andre Giardino Moreira da Silva,^{||} MD, Riccardo Gomes Gobbi,^{||¶} MD, PhD, Prof, José Ricardo Pécora,^{||} MD, PhD, Prof, and Camilo Partezani Helito,^{||#} MD, PhD, Prof

Investigation performed at Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brazil

Background: Anterior cruciate ligament reconstruction (ACLR) is the predominant and widely accepted treatment modality for ACL injury. However, recurrence of ACL rupture or failure of the reconstruction remains a significant challenge. Despite several studies in the literature that have developed prediction models to address this issue by identifying prognostic factors for treatment outcomes using classical statistical methods, the predictive efficacy of these models is frequently suboptimal.

Purpose: To (1) evaluate the predictive performance of different machine learning algorithms for the occurrence of failure in ACLR and (2) identify the most relevant predictors associated with this outcome.

Study Design: Cohort study; Level of evidence, 3.

Methods: A total of 680 patients who underwent ACLR between January 2012 and July 2021 were evaluated. The study outcome was ACLR failure—defined as a complete tear confirmed by magnetic resonance imaging, arthroscopy, or clinical ACL insufficiency—evaluated at a minimum 2-year follow-up. Routinely collected data were used to train 9 machine learning algorithms—including k-nearest neighbors classifier, decision tree classifier, random forest classifier, extra trees classifier, gradient boosting classifier, eXtreme Gradient Boosting, CatBoost classifier, and logistic regression. A random sample of 70% of patients was used to train the algorithms, and 30% were left for performance assessment, simulating new data. The performance of the models was evaluated with the area under the receiver operating characteristic curve (AUC).

Results: The predictive performance of most models was good, with AUCs ranging from 0.71 to 0.85. The models with the best AUC metric were the CatBoost classifier (0.85 [95% CI, 0.81-0.89]) and the random forest classifier (0.84 [95% CI, 0.77-0.90]). Knee hyperextension consistently emerged as the primary predictor for ACLR failure across all models subjected to our analysis.

Conclusion: Machine learning algorithms demonstrated good performance in predicting ACLR failure. Moreover, knee hyperextension consistently emerged as the primary predictor for failure across all models subjected to our analysis.

Clinical Relevance: The findings of this study highlight the potential of machine learning as a valuable clinical tool for decision-making on surgical intervention. By offering nuanced insights, these algorithms may contribute to the evolving landscape of orthopaedic practice. Also, this study confirms knee hyperextension as an important risk factor for ACLR failure.

Keywords: anterior cruciate ligament injury; anterior cruciate ligament reconstruction; artificial intelligence; machine learning

Injury to the anterior cruciate ligament (ACL) stands out as one of the most prevalent and disabling knee joint injuries.¹⁰ Despite evidence supporting the efficacy of nonsurgical interventions for this injury, the predominant and widely accepted treatment modality for a complete ACL

tear is surgical reconstruction.¹⁵ Various techniques, each utilizing different graft types with their respective advantages and disadvantages, are available, and the optimal one is yet to be defined.^{36,38,39}

When assessing surgical success, a commonly employed objective metric is the occurrence of a new ACL rupture or failure of the reconstruction.^{6,28} ACL reconstruction (ACLR) failure is considered when there is documented evidence of a new ligament rupture, assessed through

magnetic resonance imaging or arthroscopy, or when the ligament proves insufficient, resulting in objective knee laxity.³³ This laxity is clinically defined by anterior translation of the tibia of >5 mm compared with the contralateral side that can be typically accompanied by pivot-shift findings of a glide or clunk. Numerous studies have sought to establish risk factors for ACLR failure, encompassing intrinsic factors such as bone morphology or ligament laxity, and surgical factors such as graft types and fixation methods.^{23,45} Traditionally, outcome prediction models for ACLR have predominantly utilized conventional statistical approaches, notably regression models.^{7,30} However, these traditional methods have encountered challenges in accurately predicting post-ACLR outcomes. In response to these limitations, there has been a discernible shift toward the application of machine learning techniques. These advanced computational approaches are increasingly recognized for their potential in developing more robust and reliable prediction models.^{26,46}

Machine learning has been applied to enhance predictive capabilities across various orthopaedic surgeries—including osteochondral transplant,³⁷ hip arthroscopic surgery,³⁴ and rotator cuff repair.² These sophisticated statistical techniques leverage computer algorithms to model complex interactions between variables, potentially resulting in improved predictive accuracy by integrating related indicators and mitigating potential confounding factors. The essence of these techniques lies in their capacity to manage more intricate interactions compared with traditional statistics. They simultaneously analyze multiple predictive variables and their combinations, rather than validating assumed *a priori* relationships between variables. Importantly, machine learning usually prioritizes repeatable and accurate predictions over interpretability, enabling continuous improvement and self-correction. By identifying the most crucial variables for predicting outcomes, this approach can be utilized to develop better predictive algorithms.

Although there are several published studies on the use of machine learning models in the prediction of ACLR outcomes, to our knowledge, there is only 1 other study in the literature that utilized machine learning to predict ACLR objective failure.⁴⁶ However, numerous methodological and reporting transparency issues inhibit the clinical utility of the study's results despite the excellent performance of the models. Therefore, the main goal of this study was to

(1) evaluate the predictive performance of different machine learning algorithms for the occurrence of objective failure in ACLR and (2) identify the most relevant predictors associated with this outcome. We hypothesized that the best-performing machine learning models would provide reliable predictions and reveal the most significant predictors of ACLR objective failure.

METHODS

Data Source

We retrospectively observed a cohort of patients with ACL rupture who underwent arthroscopic ACLR between January 2012 and July 2021. The surgical procedures were performed by 3 surgeons in the same institution. Patients who had single-bundle ACLR with any autograft type were included. Patients who underwent associated extra-articular reconstruction and those with meniscal injuries were also included. Cases in which the allograft was used were not included in this series. Patients who had associated procedures—such as treatment of medial or lateral collateral ligament injuries, posterior cruciate ligament injuries, osteotomies, and major cartilage procedures—were excluded. To ensure adequate assessment of ACLR failure, patients with <2 years of follow-up were excluded. The study protocol received institutional review board approval, and informed consent was obtained. This study followed the guidelines of the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD).³²

Surgery and Rehabilitation

The same level of care was available to all patients. ACLR was performed with a single bundle, aiming to place the femoral tunnel in the topography of the anteromedial bundle of the native ACL. The femoral tunnel was performed by the outside-in technique. The tibial tunnel was performed from the anteromedial plateau, aiming at the footprint of the native ACL close to the medial tibial spine. The graft fixation was performed with interference screws with the knee around 30° of flexion for the hamstrings and in full extension for the patellar tendon grafts. The maximum

*Address correspondence to Rafael Krasic Alaiti, PhD, Universidade de São Paulo, Rua Dr. Ovídio Pires de Campos 333, Third Floor, Cerqueira Cesar, São Paulo, SP 05403-010, Brazil (email: rafael.alaiti@gmail.com).

[†]Research, Technology, and Data Science Office, Grupo Superador, São Paulo, Brazil.

[‡]Universidade de São Paulo, São Paulo, Brazil.

[§]Health Innovation, Data Science, and MLOps, Semantix, São Paulo, Brazil.

^{||}Grupo de Joelho, Instituto de Ortopedia e Traumatologia, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil.

[¶]HCOR, São Paulo, SP, Brazil.

^{¶¶}Hospital Sírio Libanês, São Paulo, SP, Brazil.

Final revision submitted September 9, 2024; accepted October 24, 2024.

The authors have declared that there are no conflicts of interest in the authorship and publication of this contribution. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

Ethical approval for the study was obtained from the Ethics Committee on Human Research of the Clinical Board of the Clinical Hospital of the Medical School of the University of São Paulo (research protocol number 2.472.968).

manual tension was applied for graft fixation. Revision ACLR cases were not included in the analysis.

Anterolateral ligament (ALL) reconstruction was performed using a free soft tissue graft for extra-articular reconstruction. Femoral fixation was performed using an interference screw proximal and posterior to the lateral epicondyle and tibial fixation in a tunnel passing from the Gerdy tubercle and the fibular head to the anteromedial tibia. ALL fixation was always performed in full extension and neutral knee rotation. After femoral fixation, the graft used for reconstruction was passed deep to the iliotibial band (ITB) and superficially to the lateral collateral ligament on its way to the tibia.^{19,21} For lateral extra-articular tenodesis, 1 variation of the modified Lemaire technique was used.¹⁴ First, a 10-mm wide strip approximately 10 cm long from the posterior third of the ITB was dissected, maintaining its insertion in the Gerdy tubercle. This graft was then fixed to the femur in a position posterior and proximal to the lateral epicondyle, in 0° to 30° of flexion, and neutral knee rotation. In its proximal path toward the femur, the graft was passed deeply to the lateral collateral ligament. Fixation was performed with an interference screw or a suture anchor. The indications for performing an associated extra-articular procedure were always at the discretion of each surgeon and changed throughout the study. However, as a rule, patients considered to be at an increased risk of reconstruction failure underwent this type of surgery.^{3,4,13}

All patients followed the same rehabilitation protocol, with weight-bearing and range of motion allowed since surgery, with progression as tolerated. Patients who had meniscal repair were instructed to use knee immobilizer for 4 weeks, with a restricted range of motion from 0° to 90° during this period. After the first 4 weeks, the protocol was similar to the group without meniscal injuries. Patients who wished to return to sports were allowed to do so 8 months after surgery, provided that their knees had adequate muscular control and no joint effusion, as evaluated by the single-leg hop test and the cross-over hop test. For patients who had an associated extra-articular anterolateral reconstruction, the rehabilitation protocol was similar to the protocol applied for isolated ACLR. Patients routinely returned for postoperative follow-ups at approximately 1 week, 3 weeks, 6 weeks, 3 months, 6 months, and 1 year after ACLR, with yearly follow-ups thereafter as part of standard clinical care.

Study Outcome (Target)

The main outcome of this study was objective ACL failure—defined as a complete tear confirmed by magnetic resonance imaging or arthroscopy—or clinical ACL insufficiency—defined as an anterior translation of the tibia of >5 mm or a pivot-shift test indicating high-grade rotational instability (clunk or gross). During the study period, different knee fellows and physical therapists were responsible for filling the database. Failure was recorded during regular patient follow-ups. Only patients with at least a minimum 24-month follow-up were included in the analysis.

Predictors

Predictors were selected according to already identified risk factors routinely collected at the institution. Patient characteristics (age and sex) and preoperative predictors (knee hyperextension, time from injury in months, manual maximum side-to-side difference, pivot-shift test, and meniscal injury) were collected by a research assistant in the preoperative period using a questionnaire. Intraoperative predictors—including anterolateral extra-articular augmentation, intra-articular graft size and type, and type of meniscal procedure—were collected from the surgical report.

Passive knee hyperextension was measured preoperatively (at the time of the surgical procedure and under anesthesia) using a goniometer in the contralateral knee to minimize the effects of the ACL injury on the affected knee, assuming both knees had the same degree of mobility before the ACL injury, as previously published by Sobrado et al,⁴⁰ Guimarães et al,¹⁶ and Helito et al.¹⁸ A senior knee surgeon performed all the measurements of knee extension.

Statistical Analysis

A random sampling technique was employed, utilizing 70% of the patient data for algorithm training, while the remaining 30% was reserved for performance assessment, simulating the application of the models to new data. To mitigate any bias, we implemented stratified cross-validation with 10 folds to train the models and fine-tune hyperparameters, taking all necessary precautions to prevent data leakage between the training and testing phases. Numeric predictors underwent transformation using the Yeo-Johnson method,⁴⁷ while categorical predictors were encoded using one-hot encoding. To address the class imbalance, we applied the borderline synthetic minority oversampling technique.¹⁷

Hyperparameter optimization was conducted to enhance model performance, with a particular focus on maximizing the area under the receiver operating characteristic curve (AUC). This optimization was achieved using the Optuna library,¹ utilizing the 3-structured Parzen estimator⁵ as the search algorithm and the asynchronous successive halving algorithm²⁵ as the early stopping mechanism. Consistency in preprocessing techniques was maintained across all algorithms to ensure a fair and unbiased comparison.

The effectiveness of our models was assessed using various performance metrics—including AUC, accuracy, precision, recall, F1 score, Matthews correlation coefficient, and Brier score loss. The general interpretation of these metrics is that higher values indicate superior model performance, except for the Brier score loss, where lower values are desirable. To estimate the variability of these metrics, we calculated 95% bootstrap confidence intervals. The predicting performance was categorized as excellent (>0.9), good (0.8-0.9), fair (0.7-0.8), and poor (<0.7) by the AUC value.⁴⁸ All reported performance metrics were extracted using the test data set.

We acknowledged the critical role of precise probability estimates in clinical decisions. Therefore, we calibrated all models to improve their dependability. This was achieved using Platt Scaling³⁵ and cross-validation for robust calibration. The effectiveness of this calibration process was evaluated using the Brier score loss.^{8,41}

Our study employed the following algorithms: k-nearest neighbors classifier, decision tree classifier, random forest classifier, extra trees classifier, gradient boosting classifier, light gradient boosting machine, eXtreme Gradient Boosting (XGBoost), CatBoost classifier, and logistic regression. To gain insights into the variables' influence, we utilized SHAP (SHapley Additive exPlanations)²⁹ to interpret the final model.

RESULTS

During the period evaluated, 776 surgeries were performed for ACLR, but 96 cases were excluded because they did not meet the study inclusion criteria. The analyzed sample consisted of 680 patients, without any further exclusions from the study after this point. Therefore, 680 patients were included in the study and further divided into train and test datasets. All patients included in the analysis had complete data for each of the variables evaluated.

Considering the whole sample evaluated, 454 (66.8%) patients presented at least 1° of hyperextension, 222 (32.6%) presented >5°, and 66 (9.7%) presented >10°. Considering the 37 failures, 2 (5.4%) failures occurred in patients with no hyperextension, 6 (16.2%) in patients with ≤5°, 16 (43.2%) in patients with 5° to 10°, and 13 (35.1%) in patients with >10°. The predictors and their summarized values are presented in Table 1.

The most common models for prediction with structured data were then fitted. Table 2 presents the performance and variability (ie, 95% bootstrap CIs) of the models in the test dataset, ordered by the AUC metric. The models with the best AUC metric were the CatBoost classifier (0.85 [95% CI, 0.81-0.89]) and the random forest classifier (0.84 [95% CI, 0.77-0.90]).

The SHAP was used to interpret the predictors' relationship with the model outcome that presented the best performance (Figure 1). The most relevant predictor was knee hyperextension, in which patients with higher degrees of knee hyperextension were more likely to undergo a rerupture of the graft. The second most relevant predictor was the execution of a medial meniscus meniscectomy, with patients who had this procedure being more susceptible to graft rerupture. With regard to variable importance, knee hyperextension consistently emerged as the primary predictor of significance across all models subjected to our analysis. It is particularly notable that when employing the CatBoost classifier model, knee hyperextension was identified as the most important variable for constructing new decision trees, resulting in the highest information gain for classifying outcomes in nearly 70% of the observed cases. This was followed by a medial

TABLE 1
Predictors Included in the Model (N = 680)^a

Predictors	Values
Age, mean (SD), y	30.93 (8.43)
Sex, n (%)	
Female	171 (25.15)
Male	509 (74.85)
Hyperextension, mean (SD), deg	4.05 (4.33)
Lateral extra-articular procedure (ALLR or LET), n (%)	223 (32.79)
No lateral extra-articular procedure, n (%)	457 (67.21)
Time from Injury to Surgery, mean (SD), mo	6.79 (7.59)
Intraarticular graft size, mean (SD), mm	8.25 (0.84)
Graft type, n (%)	
Patellar tendon autograft	105 (15.44)
Hamstring tendon autograft	575 (84.56)
Manual maximum side-to-side difference preop, mean (SD)	7.55 (1.18)
Pivot-shift preop grade, n (%)	
1	143 (21.03)
2	337 (49.56)
3	200 (29.41)
Meniscus, n (%)	
No lesion	463 (68.09)
Medial suture	80 (11.76)
Medial meniscectomy	64 (9.41)
Lateral meniscectomy	40 (5.88)
Lateral suture	23 (3.38)
Both suture	8 (1.18)
Both meniscectomy	2 (0.29)
Follow-up duration, mean (SD), mo	34.7 (11.45)
Failure, n (%)	
Yes	37 (5.44)
No	643 (94.56)

^aALLR, anterolateral ligament reconstruction; LET, lateral extra-articular tenodesis; preop, preoperative.

meniscus meniscectomy, which played a prominent role in the construction of new decision trees in almost 20% of cases (Figure 2). Additional details regarding the interpretation of other models can be found in the supplemental material, which is available online separately.

DISCUSSION

In this study, we developed 9 machine learning algorithms and identified the best-performing machine learning model to predict ACLR objective failure. The model with the best AUC metric was the CatBoost classifier (0.85 [95% CI, 0.81-0.89]), and the predictive performance of most models was good, with AUCs ranging from 0.71 to 0.85. Knee hyperextension consistently emerged as the primary predictor across all models subjected to our analysis. It is particularly notable that when employing the CatBoost classifier model, knee hyperextension was identified as the most important variable for constructing new decision trees, resulting in the highest information gain for classifying outcomes in nearly 70% of the observed cases. This was

TABLE 2
Performance Measures of Failure Prediction Models^a

Model	AUC	Accuracy	Precision	Recall	F1	MCC	Brier
CatBoost classifier	0.85 (0.81-0.89)	0.78 (0.76-0.81)	0.14 (0.10-0.19)	0.63 (0.49-0.77)	0.24 (0.17-0.30)	0.23 (0.14-0.30)	0.11 (0.10-0.12)
Random forest classifier	0.84 (0.77-0.90)	0.89 (0.87-0.91)	0.23 (0.15-0.32)	0.45 (0.31-0.58)	0.31 (0.20-0.49)	0.27 (0.16-0.37)	0.08 (0.06-0.09)
Decision tree classifier	0.84 (0.77-0.89)	0.86 (0.84-0.88)	0.22 (0.16-0.29)	0.63 (0.50-0.76)	0.33 (0.24-0.41)	0.32 (0.23-0.40)	0.09 (0.08-0.11)
Extra trees classifier	0.81 (0.74-0.87)	0.86 (0.84-0.88)	0.25 (0.19-0.32)	0.81 (0.70-0.91)	0.39 (0.30-0.47)	0.40 (0.33-0.48)	0.11 (0.09-0.12)
k-nearest neighbor classifier	0.80 (0.76-0.83)	0.79 (0.76-0.81)	0.16 (0.12-0.21)	0.72 (0.60-0.84)	0.27 (0.20-0.33)	0.27 (0.19-0.34)	0.14 (0.13-0.15)
Gradient boosting classifier	0.79 (0.71-0.85)	0.73 (0.71-0.76)	0.13 (0.09-0.17)	0.72 (0.60-0.84)	0.22 (0.17-0.28)	0.23 (0.16-0.29)	0.13 (0.12-0.15)
Light gradient boosting machine	0.77 (0.73-0.81)	0.72 (0.70-0.75)	0.13 (0.09-0.16)	0.72 (0.60-0.84)	0.22 (0.16-0.27)	0.22 (0.15-0.28)	0.15 (0.14-0.16)
Logistic regression	0.75 (0.71-0.79)	0.74 (0.71-0.76)	0.13 (0.09-0.18)	0.72 (0.60-0.84)	0.23 (0.17-0.29)	0.23 (0.16-0.30)	0.17 (0.16-0.19)
Extreme gradient boosting	0.71 (0.64-0.78)	0.80 (0.77-0.82)	0.12 (0.07-0.17)	0.45 (0.31-0.59)	0.19 (0.13-0.27)	0.16 (0.07-0.23)	0.14 (0.12-0.15)

^aData in parentheses are 95% bootstrap CIs. AUC, area under the receiver operating characteristic curve; MCC, Matheus correlation coefficient.

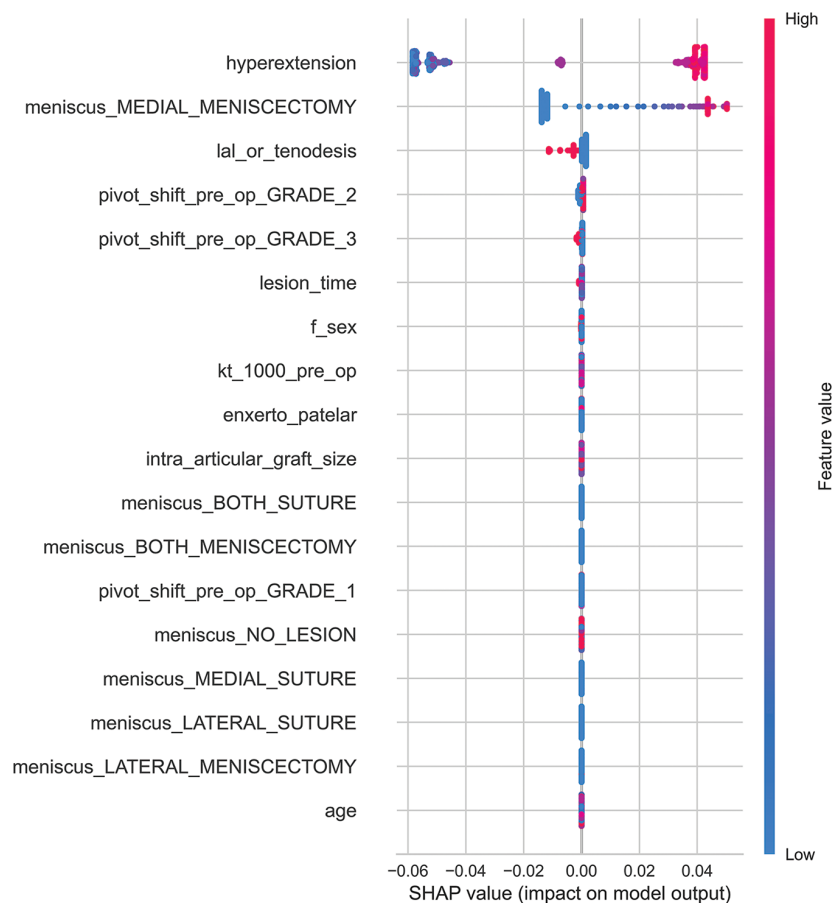


Figure 1. SHAP values of the gradient boosting classifier model. This figure provides other relevant information for model interpretation. (1) The predictors are ordered from top to bottom according to their relevance. (2) The more to the right the points of a variable are, the greater the influence of the variable in predicting the outcome (ie, failure). (3) The redder the point, the higher the predictor value, and the bluer the point, the lower the predictor value, Preop, preoperative; SHAP, SHAP SHapley Additive exPlanations.

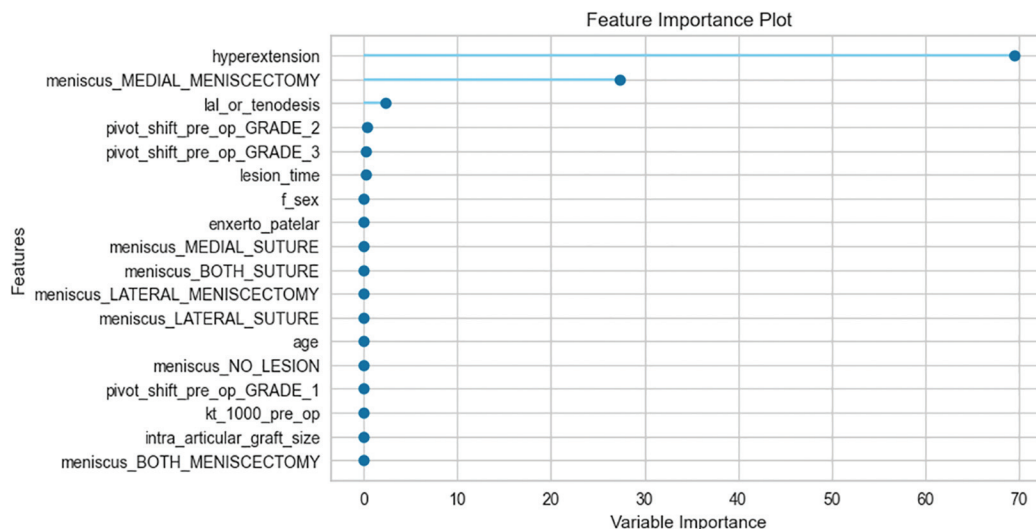


Figure 2. Variable importance plot for the gradient boosting classifier model. Variable importance is a metric that quantifies the significance of each predictor variable in our model. It denotes the proportion of times each variable was selected for building new decision trees, reflecting its impact on the model's predictive accuracy and overall performance. Higher proportions indicate greater importance in the classification of outcomes. Preop, preoperative.

followed by the execution of a medial meniscus meniscectomy, which played a prominent role in the construction of new decision trees in almost 20% of cases. These findings are significant, as they not only demonstrate the reliability of machine learning algorithms in predicting ACLR objective failure but also establish increased knee hyperextension values as one of the most relevant variables for predicting this outcome.

To our knowledge, only the study conducted by Ye et al⁴⁶ has employed machine learning algorithms to predict ACLR objective failure. In their study, 15 predictive variables and 6 outcome variables of ACLR were selected to validate a total of 36 machine models, with 6 models dedicated to each clinical outcome. Graft failure was among the considered outcomes, and the XGBoost model demonstrated superior performance (AUC, 0.944). Medial meniscal resection, participation in competitive sports, and a steep posterior tibial slope were the most important predictors of graft failure. However, despite the excellent performance of their models, numerous methodological and reporting transparency issues inhibit the clinical utility of the study results.⁹ These issues include the lack of transparency in model development—including not using the TRIPOD; dichotomizing continuous measurements for the predictors and predicted outcomes; oversampling the data; and not providing 95% CIs.

Despite the absence of other studies attempting to utilize machine learning algorithms to predict ACLR objective failure, numerous studies have sought to investigate the most important associated factors.^{22,27,31} Patient demographic factors, intrinsic patient factors (eg, bone morphology and ligament laxity), surgery-related factors (eg, graft type and diameter), and associated injuries have all been considered significant for new injuries after ACLR. A

recent study by Helito et al¹⁸ revealed that knee hyperextension of $>6.5^\circ$ was the sole factor associated with ACLR failure among a range of variables. Patients with $>6.5^\circ$ of hyperextension exhibited a 14.65 times higher chance of a new injury compared with those with less hyperextension. This study reinforces these findings by demonstrating a consistent linear relationship between knee hyperextension and failure probability. Specifically, our analysis revealed a monotonic increase in failure probability corresponding to greater degrees of knee hyperextension. Numerous other authors also regard knee hyperextension as a prognostic factor for poor outcomes in reconstruction.^{16,20,23,24} In a systematic review, Sundemo et al⁴² concluded that patients with knee hyperextension and ligament hyperlaxity demonstrated worse outcomes compared with those without these findings.

The second factor associated with failure identified in our study was medial meniscectomy, which is also relevant because the absence of the meniscus increases anterior tibial translation forces and stress on the ACL graft.^{43,44} Fithian et al¹² showed that medial meniscectomy at the time of ACLR led to higher postoperative anterior tibial translation compared with isolated ACLR or ACLR associated with meniscal repair. Dejour et al¹¹ also showed that both static and dynamic anterior tibial translation after ACLR increased in knees that had partial medial meniscectomy. All other variables exhibited minimal predictive significance in the model.

While the models developed in this study demonstrated good performance, it is essential to acknowledge a few limitations. First, the analyses conducted were not preplanned. Therefore, some relevant preoperative predictors of ACLR objective failure—such as sports participation, tibial slope, and body mass index—might not have been included in

the analysis. Also, the mean time from injury to surgery of 7 months may be considered long and might have impacted the failure mechanisms.

In addition, although the sample size used in this study is, to our knowledge, one of the largest analyzing ACLR objective failures in the literature, it is important to note that training our machine learning models with a dataset of only 680 patients might limit their utility. Machine learning models typically benefit from larger training data sets, and their performance would likely improve significantly with an increase in training data in the future. Therefore, future studies should also use larger sample sizes and include other relevant preoperative predictors for this population to develop better prediction models.

Moreover, the study relied solely on data available from our institution, and we did not conduct follow-ups with patients beyond what was documented in the medical records with a minimum 2-year follow-up. This approach, while reflective of real-world clinical practice, does carry the limitation that some patients who experienced failures after the minimum 2-year evaluation may have sought treatment elsewhere, potentially impacting the comprehensiveness of our dataset.

Finally, the difference in performance metrics suggests that the model's output is better utilized as a probability estimate rather than a binary classification. Clinicians should interpret the predicted probability as an indication of how closely a patient resembles those who have experienced graft failure rather than as a definitive prediction of failure.

Supervised machine learning algorithms are primarily designed to learn from hidden patterns in available data about maximizing outcome prediction rather than explaining causal relationships between a prediction and the outcome. This is because they adjust the weight of each variable based on the hyperparameters set and treat each category of categorical variables as a variable by itself, which makes it difficult to interpret the individual role of each predictor variable. This is also one of the reasons why machine learning algorithms may outperform conventional statistical methods and linear thinking. Nevertheless, the results of this study suggest that the use of machine learning algorithms might be a promising new tool that can assist clinicians during clinical decision-making to decide when to prescribe surgical treatment.

CONCLUSION

Most of the machine learning algorithms demonstrated good performance in predicting ACLR failure. In addition, knee hyperextension consistently emerged as the primary predictor across all models subjected to our analysis.

REFERENCES

1. Akiba T, Sano S, Yanase T, et al. Optuna: a next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD 2019. Association for Computing Machinery. 2019; 2623-2631.*
2. Alaiti RK, Vallio CS, Assunção JH, et al. Using machine learning to predict nonachievement of clinically significant outcomes after rotator cuff repair. *Orthop J Sports Med. 2023;11(10):23259671231206180.*
3. Ariel de Lima D, Helito CP, Lima FRA, Leite JAD. Surgical indications for anterior cruciate ligament reconstruction combined with extra-articular lateral tenodesis or anterolateral ligament reconstruction. *Rev Bras Ortop. 2018;19;53(6):661-667.*
4. Barahona M, Mosquera M, De Padua V, et al. Latin American formal consensus on the appropriate indications of extra-articular lateral procedures in primary anterior cruciate ligament reconstruction. *J ISAKOS. 2023;8(3):177-183.*
5. Bergstra J, Bardenet R, Bengio Y, et al. Algorithms for Hyperparameter Optimization. *Proceedings of the 24th International Conference on Neural Information Processing Systems, Granada, Spain: Curran Associates Inc; 2011;2546-2554.*
6. Boots N, van Melick N, Bogie R. A hamstring autograft diameter ≤ 8 mm is a safe option for smaller, lighter and female athletes who want to return to pivoting sports after ACL reconstruction: results from a retrospective evaluation of a local ACL register. *Knee Surg Sports Traumatol Arthrosc. 2023;31(12):5830-5836.*
7. Borque KA, Jones M, Laughlin MS, et al. Effect of lateral extra-articular tenodesis on the rate of revision anterior cruciate ligament reconstruction in elite athletes. *Am J Sports Med. 2022;50(13):3487-3492.*
8. Brier GW. Verification of forecasts expressed in terms of probability. *Monthly weather review. 1950;78(1):1-3.*
9. Bullock GS, Ward P, Losciale J, et al. Predicting the objective and subjective clinical outcomes of anterior cruciate ligament reconstruction: a machine learning analysis of 432 patients: letter to the editor. *Am J Sports Med. 2023;51(5):NP15-NP16.*
10. Chia L, De Oliveira Silva D, Whalan M, et al. Non-contact anterior cruciate ligament injury epidemiology in team-ball sports: a systematic review with meta-analysis by sex, age, sport, participation level, and exposure type. *Sports Med. 2022;52(10):2447-2467.*
11. Dejour D, Pungitore M, Valluy J, Nover L, Saffarini M, Demey G. Tibial slope and medial meniscectomy significantly influence short-term knee laxity following ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc. 2019;27(11):3481-3489. doi: 10.1007/s00167-019-05435-0*
12. Fithian A, Manoharan A, Chapek M, et al. Medial meniscectomy at the time of ACL reconstruction is associated with postoperative anterior tibial translation: a retrospective analysis. *Orthop J Sports Med. 2024;12(8):23259671241263096. doi: 10.1177/23259671241263096*
13. Getgood A, Brown C, Lording T, et al. The anterolateral complex of the knee: results from the International ALC Consensus Group Meeting. *Knee Surg Sports Traumatol Arthrosc. 2019;27(1):166-176.*
14. Getgood AMJ, Bryant DM, Litchfield R, et al. Lateral extra-articular tenodesis reduces failure of hamstring tendon autograft anterior cruciate ligament reconstruction: 2-year outcomes from the STABILITY study randomized clinical trial. *Am J Sports Med. 2020;48(2):285-297.*
15. Gföller P, Abermann E, Runer A, et al. Non-operative treatment of ACL injury is associated with opposing subjective and objective outcomes over 20 years of follow-up. *Knee Surg Sports Traumatol Arthrosc. 2019;27(8):2665-2671.*
16. Guimarães TM, Giglio PN, Sobrado MF, et al. Knee hyperextension greater than 5° is a risk factor for failure in ACL reconstruction using hamstring graft. *Orthop J Sports Med. 2021;9(11):23259671211056325.*

17. Han H, Wang WY, Mao BH. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. *Advances in Intelligent Computing. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg.* 2005; 3644:878-887.
18. Helito CP, da Silva AGM, Sobrado MF, et al. Patients with more than 6.5 degrees of knee hyperextension are 14.6 times more likely to have anterior cruciate ligament hamstring graft rupture and worse knee stability and functional outcomes. *Arthroscopy.* 2024;40(3): 898-907.
19. Helito CP, da Silva AGM, Sobrado MF, et al. Small hamstring tendon graft for anterior cruciate ligament reconstruction combined with anterolateral ligament reconstruction results in the same failure rate as larger hamstring tendon graft reconstruction alone. *Arthroscopy.* 2023;39(7):1671-1679.
20. Helito CP, Sobrado MF, Giglio PN, et al. Combined reconstruction of the anterolateral ligament in patients with anterior cruciate ligament injury and ligamentous hyperlaxity leads to better clinical stability and a lower failure rate than isolated anterior cruciate ligament reconstruction. *Arthroscopy.* 2019;35(9):2648-2654.
21. Helito CP, Sobrado MF, Moreira da Silva AG, et al. The addition of either an anterolateral ligament reconstruction or an iliotibial band tenodesis is associated with a lower failure rate after revision anterior cruciate ligament reconstruction: a retrospective comparative trial. *Arthroscopy.* 2023;39(2):308-319.
22. Johnson QJ, Jabal MS, Arguello AM, et al. Machine learning can accurately predict risk factors for all-cause reoperation after ACLR: creating a clinical tool to improve patient counseling and outcomes. *Knee Surg Sports Traumatol Arthrosc.* 2023;31(10):4099-4108.
23. Kim SJ, Choi CH, Kim SH, et al. Bone-patellar tendon-bone autograft could be recommended as a superior graft to hamstring autograft for ACL reconstruction in patients with generalized joint laxity: 2- and 5-year follow-up study. *Knee Surg Sports Traumatol Arthrosc.* 2018; 26(9):2568-2579.
24. Kim SJ, Choi CH, Lee SK, et al. Minimum two-year follow-up of anterior cruciate ligament reconstruction in patients with generalized joint laxity. *J Bone Joint Surg Am.* 2018;100(4):278-287.
25. Li L, Jamieson K, Rostamizadeh A, et al. A system for massively parallel hyperparameter tuning. *Proceedings of Machine Learning and Systems.* 2020;2:230-246.
26. Lopez CD, Gazgalis A, Peterson JR, et al. Machine learning can accurately predict overnight stay, readmission, and 30-day complications following anterior cruciate ligament reconstruction. *Arthroscopy.* 2023;39(3):777-786.e5.
27. Lu Y, Reinholz AK, Till SE, et al. Predicting the risk of posttraumatic osteoarthritis after primary anterior cruciate ligament reconstruction: a machine learning time-to-event analysis. *Am J Sports Med.* 2023;51(7):1673-1685.
28. Lubowitz JH. Editorial commentary: ACL reconstruction: single-bundle versus double-bundle. *Arthroscopy.* 2015;31(6):1197-1198.
29. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30:4765-4774.
30. MARS Group; Wright RW, Huston LJ, Haas AK, et al. Association between graft choice and 6-year outcomes of revision anterior cruciate ligament reconstruction in the MARS cohort. *Am J Sports Med.* 2021;49(10):2589-2598.
31. Martin RK, Wastvedt S, Pareek A, et al. Predicting subjective failure of ACL reconstruction: a machine learning analysis of the Norwegian Knee Ligament Register and patient reported outcomes. *J ISAKOS.* 2022;7(3):1-9.
32. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015; 162(1):W1-W73.
33. Pache S, Del Castillo J, Moatshe G, et al. Anterior cruciate ligament reconstruction failure and revision surgery: current concepts. *J ISA-KOS.* 2020;5:351-358.
34. Pettit MH, Hickman SHM, Malviya A, et al. Development of machine learning algorithms to predict attainment of minimal clinically important difference after hip arthroscopy for femoroacetabular impingement yield fair performance and limited clinical utility. *Arthroscopy.* 2024;40(4):1153-1163.e2. doi: 10.1016/j.arthro.2023.09.023
35. Platt JC. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers.* 2000; 10(3):61-74.
36. Quinn M, Byrne RA, Albright JA, et al. Peroneus longus tendon autograft may present a viable alternative for anterior cruciate ligament reconstruction: a systematic review. *Arthroscopy.* 2024;40(4):1366-1376.e1. doi: 10.1016/j.arthro.2023.10.016
37. Ramkumar PN, Karnuta JM, Haeberle HS, et al. Effect of preoperative imaging and patient factors on clinically meaningful outcomes and quality of life after osteochondral allograft transplantation: a machine learning analysis of cartilage defects of the knee. *Am J Sports Med.* 2021;49(8):2177-2186.
38. Renfree SP, Brinkman JC, Tummala SV, et al. ACL reconstruction with quadriceps soft tissue autograft versus bone-patellar tendon-bone autograft in cutting and pivoting athletes: outcomes at minimum 2-year follow-up. *Orthop J Sports Med.* 2023;11(9):23259671 231197400.
39. Ripoll T, Moreira da Silva AG, Saoudi S, et al. Comparison between continuous and separate grafts for ALL reconstruction when combined with ACL reconstruction: a retrospective cohort study from the SANTI study group. *Am J Sports Med.* 2023;51(12):3163-3170.
40. Sobrado MF, Giglio PN, Bonadio MB, et al. Outcomes after isolated acute anterior cruciate ligament reconstruction are inferior in patients with an associated anterolateral ligament injury. *Am J Sports Med.* 2020;48(13):3177-3182.
41. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128-138.
42. Sundemo D, Hamrin Senorski E, Karlsson L, et al. Generalised joint hypermobility increases ACL injury risk and is associated with inferior outcome after ACL reconstruction: a systematic review. *BMJ Open Sport Exerc Med.* 2019;5(1):e000620.
43. Syam K, Chouhan DK, Dhillon MS. Outcome of ACL reconstruction for chronic ACL injury in knees without the posterior horn of the medial meniscus: comparison with ACL reconstructed knees with an intact medial meniscus. *Knee Surg Relat Res.* 2017;29(1):39-44.
44. van der Wal WA, Meijer DT, Hoogslag RAG, et al. Meniscal tears, posterolateral and posteromedial corner injuries, increased coronal plane, and increased sagittal plane tibial slope all influence anterior cruciate ligament-related knee kinematics and increase forces on the native and reconstructed anterior cruciate ligament: a systematic review of cadaveric studies. *Arthroscopy.* 2022;38(5):1664-1688.e1.
45. Yang CP, Hung SF, Lin KY, et al. The increased lateral tibial slope may result in inferior long-term clinical outcome after DB-ACL reconstruction. *Arch Orthop Trauma Surg.* 2024; 144(2):619-626.
46. Ye Z, Zhang T, Wu C, et al. Predicting the objective and subjective clinical outcomes of anterior cruciate ligament reconstruction: a machine learning analysis of 432 patients. *Am J Sports Med.* 2022;50(14):3786-3795.
47. Yeo IK, Johnson RA. A new family of power transformations to improve normality or symmetry. *Biometrika.* 2000;87(4):954-959.
48. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation.* 2007;115(5):654-657.