

Machine learning models for predicting return to sports after anterior cruciate ligament reconstruction: Physical performance in early rehabilitation

DIGITAL HEALTH
Volume 10: 1–11
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241299065
journals.sagepub.com/home/dhj



Ui-jae Hwang¹ , Jin-seong Kim¹, Keong-yoon Kim² and Kyu-sung Chung² 

Abstract

Objective: Return to sports (RTS) after anterior cruciate ligament reconstruction (ACLR) is a crucial surgical success measure. In this study, we aimed to identify the best-performing machine learning models for predicting RTS at 12 months post-ACLR, based on physical performance variables at 3 months post-ACLR.

Methods: This case-control study included 102 patients who had undergone ACLR. The physical performance variables measured 3 months post-ACLR included the Biodex balance system, Y-balance test, and isokinetic muscle strength test. The RTS outcomes measured at 12 months post-ACLR included the single-leg hop test, single-leg vertical jump test, and Tegner activity score. Six machine learning algorithms were trained and validated using these data.

Results: Random forest models in the test set best predicted the RTS success based on the single-leg hop test (area under the curve [AUC], 0.952) and Tegner activity score (AUC, 0.949). Gradient boosting models in the test set best predicted the RTS based on the single-leg vertical jump test (AUC, 0.868).

Conclusion: Modifiable factors should be considered in the early rehabilitation stage after ACLR to enhance the possibility of a successful RTS.

Keywords

Exercise, machine learning, musculoskeletal, personalized medicine, rehabilitation, risk factors

Submission date: 20 March 2024; Acceptance date: 23 October 2024

Introduction

Return to sports (RTS) after anterior cruciate ligament (ACL) reconstruction (ACLR) is an indicator of surgical success.^{1–3} However, a successful ACLR does not assure a patient's RTS, as studies indicate that only 63% of patients regain their preinjury activity levels, with reported return rates ranging from 39% to 74%.^{2–6} Following ACLR, patients experience reduced knee muscle strength due to hamstring graft harvesting and quadriceps inhibition,⁷ along with postural stability compromise attributed to ACL mechanoreceptor injury.⁸ Collectively, these factors contribute to a decline in physical function,^{8–10} thus posing a challenge for individuals who have undergone

ACLR. Consequently, muscle strength, balance, and overall function recovery are essential to ensure a successful RTS post-surgery.^{9,11,12}

Accurate physical function and specific criteria assessment are particularly important for successful RTS.

¹Department of Physical Therapy, College of Health Science, Laboratory of KEMA AI Research (KAIR), Yonsei University, Wonju, South Korea

²Department of Orthopedic Surgery, Hanyang University Guri Hospital, Guri-si, Republic of Korea

Corresponding author:

Kyu-sung Chung, Department of orthopaedic surgery, Hanyang University Guri hospital, 153, Gyeongchun-ro, Guri-si, Gyeonggi-do, Republic of Korea. Email: drokokboy@hanmail.net



Generally, the criteria for RTS (specifically pivoting sports) are a limb symmetric index (LSI) of the single-leg hop and vertical jump tests < 10% and Tegner activity score > 6 points.^{11,13} Research has identified crucial factors influencing RTS,^{2,14} and most have focused on non-modifiable rehabilitation factors such as age, graft diameter, medial meniscal resection, knee laxity, and posterior tibial slope related to surgery or a joint condition.^{4,14,15} During the early rehabilitation stage after ACLR, muscle weakness and reduced balance occur for various reasons.^{12,16–19} These include quadriceps inhibition resulting from joint effusion and pain,¹⁶ hamstring weakness following graft harvesting in cases of hamstring autografts,¹⁷ and proprioceptive deficits due to the loss of mechanoreceptors in the native ACL.¹⁸ Additionally, the period of reduced weight-bearing and limited mobility immediately post-surgery can lead to general deconditioning.¹² These neuromuscular and proprioceptive impairments collectively contribute to decreased physical function, making it crucial to focus on strength, balance, and overall functional recovery to ensure successful rehabilitation and eventual RTS.¹⁹ Physical performance factors, such as strength, balance, and abnormal biomechanical patterns, can be modified through rehabilitation. Successful RTS rehabilitation strategies can be established by predicting RTS based on initial physical performance data after ACLR.

Recently, integrating machine learning into clinical settings and research has emerged as an advanced methodology for predicting ACLR outcomes.^{15,20–22} The machine learning approach entails amalgamating pertinent indicators while mitigating potential confounding factors, thereby enhancing the precision of predictions. In contrast to conventional statistical techniques, such as regression analysis, machine learning excels in its ability to scrutinize multiple predictive variables concurrently, encompassing their intricate combinations and interactions rather than being restricted by predefined relationships between those variables.²³ Notably, machine learning primarily focuses on delivering reproducible and accurate predictions, relegating interpretations to a secondary position and constantly refining and self-correcting its processes as the dataset expands over time.^{24,25} Considerable attention has been paid to machine learning models in predicting preoperative^{26,27} or 6 months follow-up outcomes^{28,29} for RTS at 12 months post-ACLR, yet less emphasis has been placed on predictors at the early stages of rehabilitation.

Therefore, this study aimed to identify the best-performing machine learning models for predicting RTS at 12 months post-ACLR based on physical performance variables at 3 months post-ACLR and determine the most critical predictors of RTS outcomes. We hypothesized that the best-performing machine learning models provide reliable predictions of RTS at 12 months post-ACLR.

Methods

Patients

The Institutional Review Board Inje University Seoul Paik Hospital approval (PAIK 2023-02-009) approved this study, and informed consent was waived. The medical records of 102 patients who had undergone single-bundle anatomical ACLR using the outside-in technique with a flip-cutter (Arthrex, Naples, FL, USA) between June 2016 and April 2022 were retrospectively reviewed to obtain their demographic and clinical characteristics. A single surgeon performed all the operations. The inclusion criteria for this study were patients who had undergone single-bundle ACLR, aged between 18 and 45 years, and complied with all the required tests at 3 and 12 months post-surgery. The exclusion criteria were as follows: concomitant multiple ligament injury, fracture, meniscal root repair, cartilage repair, osteotomy to correct mechanical alignment, subtotal or total meniscectomy, revision ACLR, and history of knee surgery on the involved and uninvolved sides. The patient characteristics and flowchart are shown in Table 1 and Figure 1, respectively.

According to the heuristic rule of maintaining one variable per 10 events, a sample size of at least 80 participants was required for a comprehensive statistical evaluation of RTS contribution based on eight physical performance variables.³⁰

Procedure

One of the authors (X.X.X) performed all the tests at 3 and 12 months post-ACLR. The physical performance variables (as feature or independent variables) measured at 3 months post-ACLR included the Biodex balance system (BBS) test, Y-balance test (YBT), and isokinetic muscle strength test. The RTS variables (as target or dependent variables) measured at 12 months post-ACLR include the single-leg hop test, single-leg vertical jump test, and Tegner activity score. All the tests were performed randomly at 3 and 12 months post-ACLR.

Biodex balance system test

Postural stability was measured using the BBS (Biodex Medical Systems Inc., Shirley, NY, USA).³¹ The BBS is a mobile platform with 12 difficulty levels.³² In this study, postural stability was measured using level eight resistance. The BBS electronically generates the anteroposterior (sagittal plane), mediolateral (frontal plane), and overall indexes. This study's overall index included the anteroposterior (sagittal plane) and mediolateral (frontal plane) indexes in detail.³² During the assessment, the patients were instructed to stand on one leg with both hands on their chest and maintain their balance at the

Table 1. Mean \pm standard deviation of baseline characteristics at 3 months post-ACLR and outcomes of return to sports at 12 months post-ACLR.

Baseline Characteristics at 3 months after ACLR	
Sex (M/F)	75/27
Age (yr)	30.04 \pm 9.3
Height (cm)	171.51 \pm 8.89
Weight (kg)	74.87 \pm 12.05
BMI	25.34 \pm 2.84
Involved side (Rt/Lt)	53/49
Mean follow-up period	3.12 \pm 0.36
Graft type (Hamstring autograft/ Quadriceps tendon autograft/ Hybrid graft/ Allograft)	40/22/22/18
Tenger activity score at 3 months after ACLR	4.69 \pm 1.14
YBT total score	78.87 \pm 6.58
BBS overall index	1.34 \pm 0.68
60°/s knee extensor PT	147.41 \pm 54.86
60°/s knee flexor PT	91.00 \pm 33.63
180°/s knee extensor AP	106.74 \pm 37.62
180°/s knee flexor AP	68.03 \pm 24.1
RTS outcomes at 12 months after ACLR	
Mean follow-up period	12.55 \pm 1.8
Single leg hop test	120.24 \pm 33.53
RTS-single leg hop test (0/1)	45/57
Single leg vertical jump test	13.76 \pm 4.32
RTS-single leg vertical jump test (0/1)	65/37
Tegner activity score at 12 months after ACLR	5.45 \pm 1.20
RTS-Tegner activity score (0/1)	54/48

ACLR: anterior cruciate ligament reconstruction; RTS: return to sports; BMI: body mass index; YBT: Y-balance test; BBS: Biodex balance system; PT: peak torque; AP: average power.

center of the BBS monitor. Each assessment lasted for 30 s, and three trials, with 10 s rest periods between trials, were conducted to ensure reliable measurements. Values from all three trials were automatically recorded.

Y-balance test

Dynamic balance was measured using the YBT (Move2Perform; Evansville, IN, USA).³³ The patients were instructed to stand on their weight-bearing leg in a box and gently push the side of the box using the unsupported leg, reaching as far as possible in the anterior, posteromedial, and posterolateral directions.³³ The patients underwent six practice trials to familiarize themselves with the task, followed by three YBT measurement trials in cm. The average distances achieved in the three directions were analyzed to assess dynamic balance.³³

Isokinetic muscle strength test

Isokinetic muscle strength was measured using the HUMAC-NORM isokinetic extremity system (Computer Sports Medicine Inc., Stoughton, MA, USA).⁹ We chose to assess concentric isokinetic strength as it provides a safe, reliable, and comfortable method for patients in early rehabilitation,^{34,35} while also showing strong correlations with functional outcomes.³⁶ The measurements were taken with the patients seated at angular velocities of 60°/s and 180°/s. The peak torque (PT) and the average power (AP) of the knee flexors and extensors were assessed at the 60°/s and 180°/s angular velocities, respectively. To measure PT, the patients performed four repetitions of concentric quadriceps and hamstring contractions at 60°/s after two practice sessions within 90°–0° knee flexion.⁹ To measure AP, they repeated the contractions 10 times at 180°/s after two practice sessions within similar range of knee flexion.⁹ The system automatically recorded the highest recorded data among the four repetitions for PT and AP calculated from the 10 repetitions for AP measurements.

Single-leg hop and single-leg vertical jump tests

For the single-leg hop test, the patients were instructed to stand on the test leg and hop forward as far as possible, landing on the same leg.³⁷ The hop distance was measured in cm, from the toe at the point of push-off to the heel at the landing point. This test was conducted three times, and the longest distance achieved during the single-leg hop test was analyzed. The single-leg vertical jump test was performed using InBody u-Town (InBody Corp., Seoul, Korea).³⁷ The patients were instructed to stand on one leg in the middle of a pressure-sensitive mat (90×60 cm) and

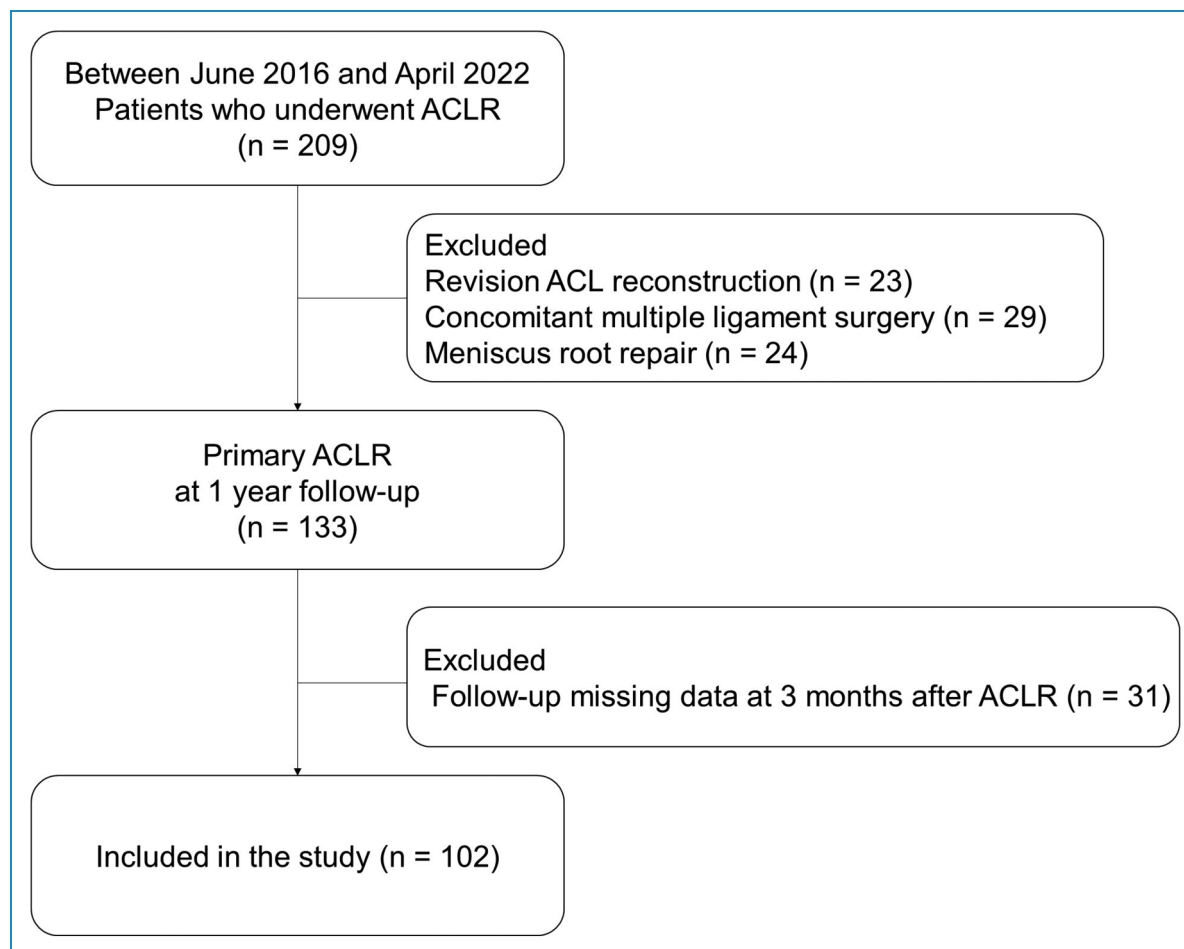


Figure 1. Flowchart of patient selection.

perform three maximal vertical jumps as high as possible, with the knee extended. Height was automatically calculated at the highest point at which the patient jumped. Jump height (cm) = $0.5 \times 9.8 \times (\text{time} \times 0.5)^2 \times 100$. These tests were performed three times, and the longest and highest distances, respectively, were analyzed.

Machine learning modeling

Machine learning analysis used Orange data mining software (Orange 3.3.0, Ljubljana, Slovenia) and Python (Version 3.6.15; Python Software Foundation).

Pre-processing and missing data handling. This study included eight numerical predictors (age, body mass index [BMI], 60°/s knee extensor PT, 60°/s knee flexor PT, 180°/s knee extensor AP, 180°/s knee flexor AP, YBT total score, and BBS overall index). The three RTS targets (single-leg hop test, single-leg vertical jump test, and Tegner activity score) were transformed into dichotomous variables as an LSI of the single-leg hop and vertical

jump test < 10% and a Tegner activity score > 6 points. Missing data were detected using exploratory data analysis. Missing data imputations were performed by eliminating instances with unknown values. The distribution of each variable was confirmed using a boxplot, scatterplot, and linear projection, respectively.

Machine learning algorithm. We split the complete data (n = 102) into a training set (80%, n = 82) for model development and a test set (20%, n = 20) for external validation to predict model performance. Six machine learning algorithms were trained via a five-fold cross-validation, including logistic regression, decision tree, random forest, gradient boosting, support vector machine, and neural network. We selected six common machine learning algorithms based on their widespread use in medical prediction tasks and ability to handle both linear and non-linear relationships in data.

Model validation. The model performance was measured primarily using the area under the curve (AUC) calculated

for the training and test datasets. The secondary measures of model performance were classification accuracy, recall, precision, and F1 score for the training and test datasets. The predictive model performance was classified as excellent (≥ 0.9), good (0.8–0.9), fair (0.7–0.8), or poor (< 0.7) based on the AUC value.¹⁵

The feature permutation importance was calculated using the training data to confirm the critical factors for each predictive variable.^{15,38,39} This analysis involved computing each feature's contribution to the model's performance based on the AUC by measuring the model's prediction error increase. In addition, a Shapley additive explanation summary plot was generated to determine the significance and direction of each predictive variable.^{15,38} Each predictive variable on the y-axis was sorted by relative importance, with the most critical predictors at the top. For each feature (predictive variable), each point (red, meaning higher values or the presence of binary factors) on the x-axis represents the individual participants' contribution to the overall Shapley additive explanation value, with higher positive contributions presented farther to the right.

Results

Patient characteristics

The mean and standard deviation of all variables are provided in Table 1. The 102 patients' data were included in the machine learning analysis. Post-hoc power analysis was performed using Python's statsmodels library. Specifically, we utilized the TTestPower analysis from statsmodels.stats.power module. This analysis confirmed that our sample size of 102 patients provided 80% power to detect an AUC of 0.70 with alpha set at 0.05, assuming a two-sided test and considering the AUC as a measure of effect size for our machine learning models. The mean (\pm standard deviation) age was 30.0 ± 9.3 years, and the proportion of males was 73.5%. The proportions of hamstring autografts, quadriceps tendon autografts, hybrid grafts, and allografts were 39.2%, 21.6%, 21.6%, and 17.6%, respectively. The Tegner activity score's mean (\pm standard deviation) at 3 months post-ACLR was 4.7 ± 1.1 . The success rates of the single-leg hop test, single-leg vertical jump test, and Tegner activity score at 12 months post-ACLR were 44.1%, 63.7%, and 52.9%, respectively.

Predictive models of machine learning

The six machine learning models' performance in RTS success prediction during model training and testing is shown in Tables 2 to 4. Feature permutation importance, Shapley additive explanation, and receiver operating characteristics curves results are presented in Figure 2.

The random forest algorithm models had the highest AUC in training (AUC, 0.877 [good]; F1, 0.829) and test

(AUC, 0.952 [excellent]; F1, 0.880) sets in predicting the success of RTS based on the single-leg hop test. The gradient boosting algorithm models had the highest AUC in training (AUC, 0.810 [good]; F1, 0.533) and test (AUC, 0.868 [good]; F1, 0.769) sets in predicting the success of RTS based on the single-leg vertical jump test. Regarding RTS success prediction based on Tegner activity score, the random forest algorithm models had the highest AUC in the training (AUC, 0.826 [good]; F1, 0.751) and test (AUC, 0.949 [excellent]; F1, 0.952) sets.

In this study, among the eight variables, the most significant predictors for RTS success at 12 months post-ACLR were identified: RTS-single leg hop test was best predicted by random forest, with YBT total score, BMI, 60°/s knee extensor PT and 180°/s knee extensor AP in feature importance, and high 60°/s knee extensor PT, high YBT total score, high 180°/s knee extensor AP, low BMI in Shapley additive explanation being key factors. RTS Tegner activity score was most accurately forecasted by random forest, primarily influenced by age, 180°/s knee flexor AP, YBT total score and 60°/s knee extensor PT in feature importance, and high 60°/s knee extensor PT, young age, high YBT total score and high 180°/s knee flexor AP in Shapley additive explanation. RTS-single-leg vertical jump test was optimally predicted by gradient boosting, with age, BMI, BBS overall index and YBT total score in feature importance, and young age, low BBS overall index, low BMI and high YBT total score in Shapley Additive Explanation as major determinants.

Discussion

Developing and applying interpretable machine learning models should be considered for musculoskeletal disorder prevention by training, evaluating, and comparing multiple machine learning models.^{25,38,40,41} In this study, we selected eight predictive variables at 3 months post-ACLR and three outcome variables of RTS at 12 months post-ACLR to validate 18 machine learning models, with six machine learning algorithms for each clinical outcome. This study identified that the best-performing machine learning models for the training set in predicting the RTS success based on the single-leg hop test results, single-leg vertical jump test, and Tegner activity score were the random forest, gradient boosting, and random forest algorithm models, with AUC of 0.877 (good), 0.810 (good), and 0.826 (good), respectively. YBT total score, BMI, 60°/s knee extensor PT, and 180°/s knee extensor AP were the top predictors of RTS based on the single-leg hop test results in the random forest model of feature permutation importance. Age, 180°/s knee flexor AP, YBT total score, and 60°/s knee extensor PT were the top predictors of RTS based on the Tegner activity score in the random forest model of feature permutation importance.

Table 2. Performance metrics of six machine learning algorithms in the training and test sets for RTS-single leg hop test.

Performance metrics of six machine learning algorithms in the training set					
Model	AUC	Acc	F1	Precision	Recall
Random Forest	0.877	0.829	0.829	0.829	0.829
Gradient Boosting	0.837	0.817	0.816	0.818	0.817
Neural Network	0.835	0.780	0.780	0.780	0.780
Decision Tree	0.809	0.732	0.732	0.732	0.732
Logistic Regression	0.779	0.659	0.658	0.658	0.659
Support Vector Machine	0.775	0.707	0.707	0.707	0.707
Performance metrics of six machine learning algorithms in the test set					
Model	AUC	Acc	F1	Precision	Recall
Random Forest	0.952	0.850	0.880	1.000	0.786
Support Vector Machine	0.929	0.750	0.783	1.000	0.643
Neural Network	0.905	0.900	0.923	1.000	0.857
Gradient Boosting	0.881	0.850	0.880	1.000	0.786
Logistic Regression	0.857	0.800	0.833	1.000	0.714
Decision Tree	0.744	0.750	0.815	0.846	0.786

AUC: area under curve; Acc: Accuracy.

Age, BMI, BBS overall index, and YBT total score were the top predictors of RTS based on the single-leg vertical jump test results in the gradient boosting model of feature permutation importance. These findings are consistent with previous studies that have demonstrated the potential of machine learning in predicting outcomes after ACLR.^{42,43} The modifiable variables presented in this study can guide exercise or physical therapy in the early rehabilitation stage for successful RTS at 12 months post-ACLR. The clinical implications of our findings include the potential to identify patients at risk of poor RTS outcomes early in the rehabilitation process. This could allow for more targeted interventions and personalized rehabilitation protocols. However, further validation in larger, multi-center cohorts is needed before clinical implementation.

Recent studies have revealed the performance of machine learning algorithms in predicting clinical outcomes after ACLR.^{15,20–22,44} Regarding the outcomes of ACLR, Kunze et al. introduced six machine learning models to predict the achievement of the minimal clinically significant difference of the International Knee Documentation

Committee (IKDC) score based on eight features (age, BMI, preoperative IKDC score, preoperative Lysholm score, medial collateral ligament examination, femoral tunnel fixation, history of contralateral knee surgery, and preoperative degree of knee extension) in a large cohort and obtained the most significant AUC of 0.82 in the elastic-net penalized logistic regression model.²⁰ Martin et al. analyzed data from a national knee ligament registry. They developed an ACLR revision risk calculator using only five predictive variables (age, preoperative Knee injury and Osteoarthritis Outcome Score Quality of Life subscale, graft choice, femoral fixation, and time from injury to surgery) based on the Cox Lasso model, with a concordance of 0.68.²¹ The use of machine learning in orthopedics and sports medicine has been growing, with studies showing its potential in various applications such as injury prediction and rehabilitation planning.⁴⁵ Regarding the performance of machine learning algorithms in predicting RTS after ACLR, Ye et al. reported that based on 15 features, the isotonicly calibrated XGBoost model (AUC, 0.773 [fair]; accuracy, 70.5%) and the random

Table 3. Performance metrics of six machine learning algorithms in the training and test sets for RTS-Tegner activity score.

Performance metrics of six machine learning algorithms in the training set					
Model	AUC	Acc	F1	Precision	Recall
Random Forest	0.826	0.756	0.751	0.762	0.756
Gradient Boosting	0.773	0.683	0.677	0.684	0.683
Neural Network	0.763	0.695	0.695	0.696	0.695
Support Vector Machine	0.735	0.744	0.741	0.745	0.744
Logistic Regression	0.700	0.683	0.677	0.684	0.683
Decision Tree	0.614	0.622	0.620	0.620	0.622
Performance metrics of six machine learning algorithms in the test set					
Model	AUC	Acc	F1	Precision	Recall
Random Forest	0.949	0.950	0.952	1.000	0.909
Gradient Boosting	0.889	0.800	0.818	0.818	0.818
Support Vector Machine	0.838	0.800	0.800	0.889	0.727
Logistic Regression	0.768	0.600	0.556	0.714	0.455
Neural Network	0.758	0.800	0.833	0.769	0.909
Decision Tree	0.707	0.750	0.762	0.8	0.727

AUC: area under curve; Acc: Accuracy.

forest model (AUC, 0.777 [fair]; accuracy, 69.2%) were the best-performing models for return to preinjury sports and pivoting sports, respectively.¹⁵ The present study confirmed that the random forest machine learning algorithm showed high performance (AUC = 0.949 [excellent]; accuracy: 0.950, F1: 0.952) in the test set in predicting the success of RTS based on the Tegner activity score. RTS requires functional or physical performance, such as muscle strength, postural control, and balance. Thus, modifiable physical performance predictors are suitable for predicting the success of RTS after ACLR and may improve the performance of the models presented in this study.

Notably, several studies have investigated the factors associated with RTS following anterior ACLR, such as muscle strength,^{28,46} functional performance^{2,47} and self-reported function,^{28,48} at 6 months follow-up. Lentz et al. highlighted the role of fear of re-injury, quadriceps strength, and self-reported function at 6 months post-ACLR in influencing RTS outcomes at 12 months post-ACLR.²⁸ Sousa et al. conducted a study demonstrating that individuals who exhibited excellent performance in terms of muscle

strength and functional testing at 6 months post-ACLR had higher Tegner activity scores at least 2 years after ACLR than individuals with delayed performance.²⁹ Within the first 3 months after ACLR, the primary focus of rehabilitation is to reduce pain, restore quadriceps and hamstring strength, and incorporate proprioception training.^{49,50} Typically, 3 months post-surgery marks a crucial transition point in the rehabilitation process; it mainly marks an athlete's commencement of impact drills, such as double-leg jumping tasks and running.^{1,12,49,50} During this time, they must possess sufficient strength and coordination to perform these tasks without deficits and in a good form.⁵⁰ Our study developed models to predict RTS success based on variables in the early rehabilitation stage after ACLR. High 60°/s knee extensor PT, high YBT total score, high 180°/s knee extensor and flexor AP, low BBS overall index, young age, and low BMI at 3 months after ACLR was inputted into the machine learning algorithms for the Shapley additive explanation analysis of RTS success based on the single-leg hop and single-leg vertical jump tests and Tegner activity score at 12 months post-

Table 4. Performance metrics of six machine learning algorithms in the training and test sets for RTS-Single leg vertical jump test.

Performance metrics of six machine learning algorithms in the training set					
Model	AUC	Acc	F1	Precision	Recall
Gradient Boosting	0.810	0.659	0.533	0.533	0.533
Random Forest	0.791	0.720	0.566	0.652	0.500
Decision Tree	0.665	0.683	0.536	0.577	0.500
Neural Network	0.643	0.622	0.415	0.478	0.367
Support Vector Machine	0.547	0.622	0.205	0.444	0.133
Logistic Regression	0.506	0.598	0.195	0.364	0.133
Performance metrics of six machine learning algorithms in the test set					
Model	AUC	Acc	F1	Precision	Recall
Gradient Boosting	0.868	0.850	0.769	0.833	0.714
Random Forest	0.824	0.850	0.769	0.833	0.714
Support Vector Machine	0.747	0.750	0.545	0.750	0.429
Decision Tree	0.731	0.750	0.615	0.667	0.571
Neural Network	0.725	0.750	0.615	0.667	0.571
Logistic Regression	0.725	0.600	0.333	0.4	0.286

AUC: area under curve; Acc: Accuracy.

ACLR. Shapley additive explanations analysis reveals the correlation between predictors and clinical outcomes.⁵¹ Nevertheless, a strong correlation between a variable and an outcome does not inherently signify a causal connection.^{51,52} Thus, future studies should investigate whether targeted interventions based on variables prioritized for prediction in machine learning algorithms result in improved clinical outcomes, such as the success of RTS based on the single-leg hop and single-leg vertical jump tests results and Tegner activity score at 12 months post-ACLR. This approach aligns with the growing trend of using machine learning to develop personalized rehabilitation strategies in sports medicine and rehabilitation.⁵³

This study has several limitations that should be addressed in future research. First, the sample size was relatively small for machine learning applications, which may limit the findings' generalizability. An appropriate sample size is critical to reduce the risk of overfitting when developing a prediction model. Future studies should consider a larger sample size and employ a three-layered structure – training sample, test sample, and validation sample – to enhance the model's

generalizability. Second, we did not incorporate controls for variables such as surgical technique or graft type, despite the well-documented knowledge that these factors can exert distinct influences on quadriceps and hamstring strength outcomes. Third, although physical performance outcomes at 3 months post-ACLR were analyzed in the prediction models, other potential predictive variables, such as psychological factors, surgical metrics, and pre-injury activity levels, were omitted. Including these non-performance-based metrics could potentially improve the predictive power of the models. Fourth, the best-performing model may not have been one of the six models we selected, as machine learning algorithms are diverse. Future research should consider testing additional models and potentially implementing cross-validation techniques to identify the optimal predictive model. Lastly, our study was limited to data from a single institution. A multi-centric approach using an independent sample to validate the models could enhance the generalizability of the results in future studies.

Future studies should address key areas to enhance the application of machine learning models in predicting RTS

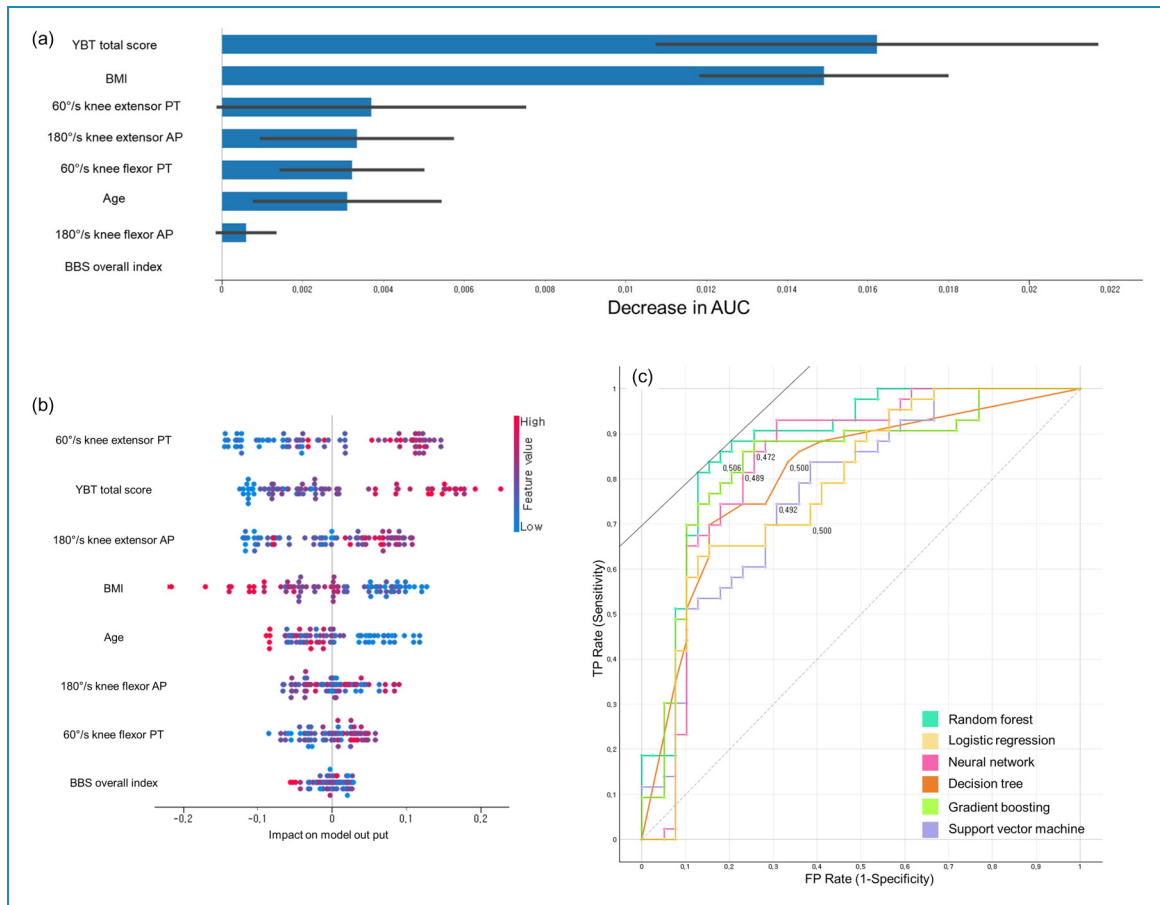


Figure 2. (a) Random forest model's feature permutation importance in the training set for predicting RTS-single leg hop test; (b) Shapley additive explanation analyses of random forest model in the training set for predicting RTS-single leg hop test; (c) ROC curves of six machine learning algorithms in the training set for predicting RTS-single leg hop test.

after ACLR. Research should include a broader age range to better generalize findings across different age groups and activity levels. Incorporating psychological factors, such as readiness and fear of reinjury, could provide more comprehensive predictions of RTS outcomes. To maximize clinical utility, future studies should explore how stakeholders can leverage these predictions to tailor rehabilitation protocols, potentially developing decision support tools that integrate model predictions with clinical expertise. This could enhance the practical application of machine learning in ACLR rehabilitation and RTS decision-making. Furthermore, investigating whether targeted interventions based on variables prioritized by machine learning algorithms improve clinical outcomes could validate the efficacy of this approach in ACLR rehabilitation.

Conclusion

In conclusion, using machine learning models, this study predicted RTS outcomes at 12 months post-ACLR based on physical performance measures at 3 months

post-ACLR. The findings highlighted modifiable factors such as 60°/s knee extensor PT, YBT total score, 180°/s knee extensor and flexor AP, BBS overall index, and BMI in predicting successful RTS. These factors can guide the development of tailored rehabilitation strategies during the early rehabilitation stages after ACLR. Machine learning proved effective in our study, but further research is required to validate the practical impact of these modifiable factors on RTS outcomes.

Acknowledgments: We want to thank Sports Medical Center team in Seoul Paik Hospital for their active participation and cooperation.

Author's contributions: UJH, JSK, KSC were responsible for the investigation, conceptualization and methodology. KYK and KSC provided supervision and resources for analysis. UJH, JSK, KSC contributed to formal analysis and methodology. KYK is responsible for data curation and clinical validation. UJH and KSC drafted the manuscript. All authors reviewed and conducted the final approval of the version to be published. All authors agreed to submit the report for publication.

Author's Note: Jin-seong Kim is also affiliated with the Department of Physical Therapy, Ilsan Paik Hospital, College of Medicine, Inje University, Goyang-si, Gyeonggi-do, Republic of Korea.

Availability of data and materials: The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics approval statement: The Institutional Review Board Inje University Seoul Paik Hospital approval (PAIK 2023-02-009) approved this study, and informed consent was waived.

Funding: The authors received no financial support for the research, authorship, and/or publication of this article.

Guarantor: Kyu-sung Chung.

ORCID iDs: Ui-jae Hwang  <https://orcid.org/0000-0002-2050-5503>
Kyu-sung Chung  <https://orcid.org/0000-0002-2378-0359>

References

1. Kvist J. Rehabilitation following anterior cruciate ligament injury: current recommendations for sports participation. *Sports Med* 2004; 34: 269–280.
2. Ardern CL, Taylor NF, Feller JA, et al. Fifty-five per cent return to competitive sport following anterior cruciate ligament reconstruction surgery: an updated systematic review and meta-analysis including aspects of physical functioning and contextual factors. *Br J Sports Med* 2014; 48: 1543–1552.
3. Suzuki M, Ishida T, Matsumoto H, et al. Association of psychological readiness to return to sports with subjective level of return at 12 months after ACL reconstruction. *Orthop J Sports Med* 2023; 11: 23259671231195030.
4. Joreitz R, Lynch A, Rabuck S, et al. Patient-specific and surgery-specific factors that affect return to sport after ACL reconstruction. *Int J Sports Phys Ther* 2016; 11: 264.
5. Lindanger L, Strand T, Mølster AO, et al. Effect of early residual laxity after anterior cruciate ligament reconstruction on long-term laxity, graft failure, return to sports, and subjective outcome at 25 years. *Am J Sports Med* 2021; 49: 1227–1235.
6. Ortiz E, Zicaro JP, Mansilla IG, et al. Revision anterior cruciate ligament reconstruction: return to sports at a minimum 5-year follow-up. *World J Orthop* 2022; 13: 12.
7. de Jong SN, van Caspel DR, van Haeff MJ, et al. Functional assessment and muscle strength before and after reconstruction of chronic anterior cruciate ligament lesions. *Arthroscopy* 2007; 23: 21. e21–221. 11.
8. Paterno MV, Schmitt LC, Ford KR, et al. Biomechanical measures during landing and postural stability predict second anterior cruciate ligament injury after anterior cruciate ligament reconstruction and return to sport. *Am J Sports Med* 2010; 38: 1968–1978.
9. Chung KS, Ha JK, Yeom CH, et al. Are muscle strength and function of the uninjured lower limb weakened after anterior cruciate ligament injury? Two-year follow-up after reconstruction. *Am J Sports Med* 2015; 43: 3013–3021.
10. Chona D, Eriksson K, Young SW, et al. Return to sport following anterior cruciate ligament reconstruction: the argument for a multimodal approach to optimise decision-making: current concepts. *J ISAKOS* 2021; 6: 344–348.
11. Barber-Westin SD and Noyes FR. Factors used to determine return to unrestricted sports activities after anterior cruciate ligament reconstruction. *Arthroscopy* 2011; 27: 1697–1705.
12. Myer GD, Paterno MV, Ford KR, et al. Rehabilitation after anterior cruciate ligament reconstruction: criteria-based progression through the return-to-sport phase. *J Orthop Sports Phys Ther* 2006; 36: 385–402.
13. Øiestad BE, Holm I and Risberg MA. Return to pivoting sport after ACL reconstruction: association with osteoarthritis and knee function at the 15-year follow-up. *Br J Sports Med* 2018; 52: 1199–1204.
14. Muller B, Yabroudi MA, Lynch A, et al. Return to preinjury sports after anterior cruciate ligament reconstruction is predicted by five independent factors. *Knee Surg Sports Traumatol Arthrosc* 2022; 30: 84–92.
15. 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: 3786–3795.
16. Rice DA and McNair PJ. Quadriceps arthrogenic muscle inhibition: neural mechanisms and treatment perspectives. *Semin Arthritis Rheum* 2010; 250–266. DOI: 10.1016/j.semarthrit.2009.10.001.
17. Konrath JM, Vertullo CJ, Kennedy BA, et al. Morphologic characteristics and strength of the hamstring muscles remain altered at 2 years after use of a hamstring tendon graft in anterior cruciate ligament reconstruction. *Am J Sports Med* 2016; 44: 2589–2598.
18. Gokeler A, Benjaminse A, Hewett TE, et al. Proprioceptive deficits after ACL injury: are they clinically relevant? *Br J Sports Med* 2012; 46: 180–192.
19. Lepley LK. Deficits in quadriceps strength and patient-oriented outcomes at return to activity after ACL reconstruction: a review of the current literature. *Sports Health* 2015; 7: 231–238.
20. Kunze KN, Polce EM, Ranawat AS, et al. Application of machine learning algorithms to predict clinically meaningful improvement after arthroscopic anterior cruciate ligament reconstruction. *Orthop J Sports Med* 2021; 9: 23259671211046575.
21. Martin RK, Wastvedt S, Pareek A, et al. Machine learning algorithm to predict anterior cruciate ligament revision demonstrates external validity. *Knee Surg Sports Traumatol Arthrosc* 2022; 30: 368–375.
22. Martin RK, Wastvedt S, Pareek A, et al. Predicting anterior cruciate ligament reconstruction revision: a machine learning analysis utilizing the Norwegian Knee Ligament Register. *JBJ* 2022; 104: 145–153.
23. Rajula HSR, Verlato G, Manchia M, et al. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina (B Aires)* 2020; 56: 55.

24. Ramkumar PN, Luu BC, Haeberle HS, et al. Sports medicine and artificial intelligence: a primer. *Am J Sports Med* 2022; 50: 1166–1174.
25. Ley C, Martin RK, Pareek A, et al. Machine learning and conventional statistics: making sense of the differences. *Knee Surg Sports Traumatol Arthrosc* 2022; 30: 753–757.
26. Ueda Y, Matsushita T, Araki D, et al. Factors affecting quadriceps strength recovery after anterior cruciate ligament reconstruction with hamstring autografts in athletes. *Knee Surg Sports Traumatol Arthrosc* 2017; 25: 3213–3219.
27. Kitaguchi T, Tanaka Y, Takeshita S, et al. Preoperative quadriceps strength as a predictor of return to sports after anterior cruciate ligament reconstruction in competitive athletes. *Phys Ther Sport* 2020; 45: 7–13.
28. Lentz TA, Zeppieri G Jr, George Set al. Comparison of physical impairment, functional, and psychosocial measures based on fear of reinjury/lack of confidence and return-to-sport status after ACL reconstruction. *Am J Sports Med* 2015; 43: 345–353.
29. Sousa PL, Krych AJ, Cates RA, et al. Return to sport: does excellent 6-month strength and function following ACL reconstruction predict midterm outcomes? *Knee Surg Sports Traumatol Arthrosc* 2017; 25: 1356–1363.
30. Liew BXW, Kovacs FM, Rugamer D, et al. Machine learning versus logistic regression for prognostic modelling in individuals with non-specific neck pain. *Eur Spine J* 2022; 31: 2082–2091.
31. Wikstrom EA, Tillman MD, Chmielewski TL, et al. Measurement and evaluation of dynamic joint stability of the knee and ankle after injury. *Sports Med* 2006; 36: 393–410.
32. Arifin N, Osman NAA and Abas WABW. Intrarater test-retest reliability of static and dynamic stability indexes measurement using the Biodex Stability System during unilateral stance. *J Appl Biomech* 2014; 30: 300–304.
33. Linek P, Sikora D, Wolny T, et al. Reliability and number of trials of Y Balance Test in adolescent athletes. *Musculoskeletal Sci Pract* 2017; 31: 72–75.
34. Hsiao S-F, Chou P-H, Hsu H-C, et al. Changes of muscle mechanics associated with anterior cruciate ligament deficiency and reconstruction. *J Strength Cond Res* 2014; 28: 390–400.
35. Wilk KE, Romaniello WT, Soscia SM, et al. The relationship between subjective knee scores, isokinetic testing, and functional testing in the ACL-reconstructed knee. *J Orthop Sports Phys Ther* 1994; 20: 60–73.
36. Zwolski C, Schmitt LC, Quatman-Yates C, et al. The influence of quadriceps strength asymmetry on patient-reported function at time of return to sport after anterior cruciate ligament reconstruction. *Am J Sports Med* 2015; 43: 2242–2249.
37. Kim JS, Choi MY, Kong DH, et al. Does a lower limb balance test after anterior cruciate ligament reconstruction have a significant correlation with postoperative clinical score, stability, and functional performance test? *Clin Orthop Surg* 2023; 15: 02.
38. Hwang U-J, Kwon O-Y, Kim J-H, et al. Classification of chronic ankle instability using machine learning technique based on ankle kinematics during heel rise in delivery workers. *Digital Health* 2024; 10: 20552076241235116.
39. Hwang U-J and Gwak G-T. Machine learning vs. logistic regression for classifying pressure pain hypersensitivity based on postural analysis data in food service workers with nonspecific neck/shoulder myofascial pain. *J Musculoskeletal Sci Technol* 2023; 7: 71–79.
40. Chan VC, Ross GB, Clouthier AL, et al. The role of machine learning in the primary prevention of work-related musculoskeletal disorders: a scoping review. *Appl Ergon* 2022; 98: 103574.
41. Edouard P, Verhagen E and Navarro L. Machine learning analyses can be of interest to estimate the risk of injury in sports injury and rehabilitation. *Ann Phys Rehabil Med* 2022; 65: 101431.
42. Wiggins AJ, Grandhi RK, Schneider DK, et al. Risk of secondary injury in younger athletes after anterior cruciate ligament reconstruction: a systematic review and meta-analysis. *Am J Sports Med* 2016; 44: 1861–1876.
43. Christino MA, Fleming BC, Machan JT, et al. Psychological factors associated with anterior cruciate ligament reconstruction recovery. *Orthop J Sports Med* 2016; 4: 2325967116638341.
44. Chu Y, Knell G, Brayton RP, et al. Machine learning to predict sports-related concussion recovery using clinical data. *Ann Phys Rehabil Med* 2022; 65: 101626.
45. Halilaj E, Rajagopal A, Fiterau M, et al. Machine learning in human movement biomechanics: best practices, common pitfalls, and new opportunities. *J Biomech* 2018; 81: 1–11.
46. Petersen W, Taheri P, Forkel P, et al. Return to play following ACL reconstruction: a systematic review about strength deficits. *Arch Orthop Trauma Surg* 2014; 134: 1417–1428.
47. Cristiani R, Mikkelsen C, Forssblad M, et al. Only one patient out of five achieves symmetrical knee function 6 months after primary anterior cruciate ligament reconstruction. *Knee Surg Sports Traumatol Arthrosc* 2019; 27: 3461–3470.
48. Ardern CL, Taylor NF, Feller JA, et al. Sports participation 2 years after anterior cruciate ligament reconstruction in athletes who had not returned to sport at 1 year: a prospective follow-up of physical function and psychological factors in 122 athletes. *Am J Sports Med* 2015; 43: 848–856.
49. Erickson LN, Jacobs CA, Johnson DL, et al. Psychosocial factors 3-months after anterior cruciate ligament reconstruction predict 6-month subjective and objective knee outcomes. *J Orthop Res* 2022; 40: 231–238.
50. Kline PW, Johnson DL, Ireland ML, et al. Clinical predictors of knee mechanics at return to sport following ACL reconstruction. *Med Sci Sports Exercise* 2016; 48: 90.
51. Heskes T, Sijben E, Bucur IG, et al. Causal shapley values: exploiting causal knowledge to explain individual predictions of complex models. *Adv Neural Inf Process Syst* 2020; 33: 4778–4789.
52. Bullock GS, Hughes T, Sergeant JC, et al. Clinical prediction models in sports medicine: a guide for clinicians and researchers. *J Orthop Sports Phys Ther* 2021; 51: 517–525.
53. Shull PB, Jirattigalachote W, Hunt MA, et al. Quantified self and human movement: a review on the clinical impact of wearable sensing and feedback for gait analysis and intervention. *Gait Posture* 2014; 40: 11–19.