International Journal of Research in Orthopaedics
Case TR et al. Int J Res Orthop. 2023 Sep;9(5):888-895
http://www.ijoro.org

Original Research Article

DOI: https://dx.doi.org/10.18203/issn.2455-4510.IntJResOrthop20232605

Employing machine learning to predict adverse acute post-surgical outcomes following elective ulnar collateral ligament reconstruction

Trevor R. Case[1], Richard C. Rice[2]*, Andrew J. Cabrera[2], Mikayla Kricfalusi[1], Prathyusha Dasari[3], Jose Jesurajan[2], David E. Ruckle[2], Adam LaFleur[4], Christopher M. Jobe[2], Hasan M. Syed[2]

[1]California University of Science and Medicine, Colton, California, USA
[2]Loma Linda University Orthopedics, Loma Linda, California, USA
[3]University of California Riverside School of Medicine, Riverside, California, USA
[4]Riverside University Health System, Moreno Valley, California, USA

Received: 13 July 2023
Revised: 14 August 2023
Accepted: 17 August 2023

***Correspondence:**
Dr. Richard C. Rice,
E-mail: rcaseyrice@gmail.com

ABSTRACT

Background: Ulnar collateral ligament reconstruction ameliorates valgus elbow instability in various patient populations, including overhead athletes, patients with acute UCL rupture following high energy trauma, and those with chronic, subclinical elbow laxity. This study aims to explore machine learning algorithms to identify risk factors in patients undergoing elective UCL reconstruction in the ambulatory setting to predict postoperative outcomes.
Methods: RStudio was used to create a filtering code to identify adult patients who underwent elective UCL reconstruction from 2008 to 2018 in the American college of surgeons national surgical quality improvement program database. Patients were analyzed using six ML algorithms, which were trained to predict outcomes such as extended length of stay, non-home discharge, and adverse events based on various patient characteristics and surgical variables. Algorithmic performance was then assessed and top performing algorithms underwent further analysis to determine relative feature importance using a permutation feature importance method.
Results: ML exhibited excellent performance in predicting LOS, with an average AUC of 0.953, similar to that of logistic regression. Regarding NHD, ML demonstrated a 60.8% increase in AUC compared to LR. In predicting AAE, ML achieved an average AUC that was 12.7% higher than LR.
Conclusions: The highly predictive capability of ML indicates the possibility to represent a procedure-specific complementary tool for the preoperative risk stratification process. This study provides a comprehensive analysis of UCL reconstruction in the management and outcomes of any patient, regardless of age or activity level.

Key Words: Machine learning, ACS-NSQIP, UCL reconstruction

INTRODUCTION

The medial ulnar collateral ligament (UCL) is a crucial soft tissue stabilizer responsible for protecting the elbow against valgus stress.[1] UCL insufficiency may present acutely or chronically, and result in an unstable ulnohumeral joint when subjected to valgus stress. Medial elbow instability commonly results from repetitive overuse, as is seen in throwing athletes, and less frequently following acute valgus load during high-energy trauma.[1-4] From an academic perspective, UCL incompetency is most comprehensively characterized in the overhead throwing athlete patient, however, elective UCL reconstruction is also commonly performed in a relatively older and larger patient population for chronic elbow laxity and instability. In this sub-group, valgus instability can manifest with insidious onset of medial-sided elbow pain and decreased range of motion in ex-athletes in their late 30's and 40's, which may require surgical intervention.[5] While the pathogenesis of UCL incompetence in this population is equivocal, the etiologic agent is hypothesized to be secondary to serial valgus loads predominately afflicting laborers and ex-athletes that either went untreated or developed symptomatic clinical instability later in life. Regardless of the mechanistic mediator, medial elbow ligamentous reconstruction nonetheless represents an effective procedure in attenuating pain and augmenting stability for a heterogeneous patient population. While the revision rate for UCL reconstruction is currently low (1-7%), it is projected to become more prevalent as the incidence of UCL injuries increases among adolescent athletes, and as older patients begin to experience chronic elbow instability.[1,6-8] Taken together, the epidemiologic projection of increased UCL reconstruction revision incidence coupled with the current rates of surgical intervention for chronic instability highlight the importance of characterizing operative outcomes in all patients undergoing reconstruction, not just overhead athletes. Due to the emphasis on performance-related outcomes in young athletes, there are no studies (to our knowledge) that characterize adverse outcomes following UCL reconstruction, such as non-home discharge or rates of transfusion, in a larger, more generalizable population. UCL reconstruction is most frequently an ambulatory surgery, and as such is considered to carry relatively low perioperative risk. Ambulatory surgery is more cost-effective than inpatient based surgery, with savings ranging from 17-43%, thus an increasing number of procedures are moving to ambulatory surgery centers.[9,10] However, current literature has focused on ways to ensure that outpatient surgeries are still providing the same high quality care for patients while taking advantage of opportunities to lower cost for the healthcare system. Especially for more complex procedures, there has been ample investigation into risk factors for postoperative admission, non-home discharge, and adverse outcomes. Even though these are relatively infrequent occurrences after UCL reconstruction, having the ability to identify key variables that predispose patients to these poor outcomes can enable providers to prophylactically address them when possible. Ultimately, accurately and efficiently predicting procedure-specific, adverse events may help attenuate deleterious post-operative outcomes for patients while simultaneously improving efficiency of healthcare delivery and financial stewardship.

This study aims to explore the potential of machine learning (ML) algorithms to identify risk factors in UCL reconstruction and predict post-intervention outcomes. ML has previously demonstrated predictive abilities within the field of orthopedics by digesting large data sets and weighing patient characteristics and surgical variables that represents independent risk factors for an outcome of interest.[11,12] In this study, we utilized ML learning algorithms to analyze an American college of surgeons national surgical quality improvement program (ACS-NSQIP) dataset and generate a procedure-specific index capable of predicting adverse outcomes of interest following any patient that undergoes elective, ambulatory UCL reconstruction.

METHODS

For this observational study, we received exempt status from the Institutional Review Board to proceed with this project due to the deidentified data collection. Rstudio (RStudio, PBC, Boston, MA) was utilized to create a filtering code that identified adult patients undergoing elective UCL reconstruction between January 2008 and December 2018, from the 722 hospitals across the United States participating in the ACS-NSQIP database. We used a retrospective cohort study design conducted at Loma Linda University (starting May 1, 2023 and ending July 3, 2023) to evaluate patients in NSQIP who underwent elective ulnar collateral ligament reconstruction using Current Procedural Terminology (CPT) code 24346. Patients with underlying malignancy, or those with missing data were excluded from analyses within our study. Patients matching the above criteria were analyzed by six supervised ML classification algorithms, namely Random Forest Classifier (RF), Gradient Boosting Classifier (GB), Support Vector Machine Classifier. (SVM), Gaussian Naive Bayes Classifier (GNB), Decision Tree (DT) and Multi-layer Perceptron Classifier (MLP). The SciKit-Learn library in the python programming language was used to construct all algorithms and tasked with predicting extended length of stay (LOS), non-home discharge (NHD), and any adverse event (AAE) based on a given set of patient variables.[13-15] As elective UCL reconstruction is an outpatient procedure extended LOS was defined as longer than 24 hours, indicating an overnight stay.[9,10] For ambulatory surgery centers, Medicare does not permit planned overnight admissions, so a patient staying longer than 24 hours following UCL reconstruction indicates a substantial adverse event.16 Home discharge included discharge locations encoded as "home," "against medical advice," or "facility which was home" in ACS-NSQIP. NHD was defined as discharge to "skilled care," "unskilled facility not home," "separate acute care," "multi-level senior community," or "rehabilitation facility." The AAE category was defined as having any one or multiple of the following: surgical site infection, renal complications, sepsis, intubation, transfusion, pneumonia, deep vein thrombosis (DVT), urinary tract infection (UTI), cerebrovascular accidents, cardiac arrest, myocardial infarction (MI), return to operating room, or death. Variables used to predict these outcomes included demographic information, operative time, and comorbidities as seen in (Table 1).

Patient variables were standardized using SciKit-Learn's StandardScaler during the preprocessing stage. This involved removing the mean and scaling the variables to unit variance, ensuring that all features were brought to the same magnitude.[17] To train and later assess the model's performance, Scikit-Learn's train-test-split method was employed to split the population data into training and testing sets. The data was divided such that 70% was used for training the model, while the remaining 30% was reserved for evaluating the model's performance at a later stage.[17,18] To identify the optimal hyperparameters, we utilized Scikit-Learn's GridSearchCV along with a fivefold cross-validation technique.[17,19] Following the selection of appropriate hyperparameters, the final models were assessed for their performance by evaluating them using the 30% testing data. Metrics used to determine algorithmic performance were classification accuracy, sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and area under the receiver operator curve (AUC).[20] The sensitivity and specificity were used to calculate the negative and positive likelihood ratios for each algorithm in predicting the outcomes of interest. The negative likelihood ratio; NLR=(1-Sensitivity)/ Specificity) is the probability that an algorithm correctly predicted a negative result in a patient without a specific adverse outcome. Conversely, positive likelihood ratio (PLR=Sensitivity/(1-Specificity) is the implied probability that a flagged positive result would correctly be assigned to a patient that has an outcome of interest.[21] Subsequently, each model was then categorized as acceptable, excellent, or outstanding based on AUC ranges of 0.70-0.79, 0.80-0.89, or 0.90 or greater, respectively.[22] The Matplotlib library was utilized to generate visualizations of the AUCs produced by each model.[23] Further analysis using Permutation feature importance (PFI) was used to identify important patient characteristics that lead to accurate predictions. This involves randomly shuffling or removing a single feature and assessing the resulting impact on the model's performance. This process disrupts the relationship between the variable and the predicted outcome, providing insights into the variable's importance. A decrease in model performance indicates the extent to which the model relies on that specific variable for its predictions.[24-26] Statistical analysis was conducted using SPSS version 28 (IBM Corporation, 2021, Armonk, NY, USA) with a significance level set at $p<0.05$. Descriptive statistics, including percentages, mean, and standard deviations (SD), were computed. Categorical differences between groups were assessed using Pearson's Chi-Square test or Fischer's exact test when the conditions for Chi-Square test were not met. Numerical differences between groups were assessed using independent sample t-tests with Levene's test for equality of variance, as well as one-way analysis of variance (ANOVA) with Bonferroni and Tukey corrections to compare the groups.

RESULTS

There were 174 patients included in the study (129 male and 45 female) with an average age of 39.3 years. Upon admission, the majority of patients (96.6%) displayed functional independence, with the remainder displaying only partial dependence. Among the observed comorbidities, diabetes was the most prevalent, occurring in 6.3% of the patients. Dyspnea was also present, with a low incidence of 4%.

Table 1: Preoperative characteristics of study population undergoing ulnar collateral ligament reconstruction.

Characteristics	N (%)
Demographics	
Total patients	174
Mean age (years)	39.29±22.21
Mean body mass index	28.19±5.74
Gender	
Male	129 (74)
Female	45 (26)
Race	
Asian	3 (2)
Black or African American	13 (8)
Native Hawaiian or Pacific Islander	5 (3)
White	135 (78)
Unknown/Not Reported	18 (10)
Ethnicity Hispanic	21 (12)
Functional Status	
Independent	168 (97)
Partially Dependent	6 (4)
Totally Dependent	0 (0)
Comorbidities	
Smoking	21 (12)
Diabetes	
Non-insulin Dependent	9 (5)
Insulin Dependent	2 (1)
Congestive Heart Failure	1 (1)
COPD	5 (3)
Hypertension Requiring Medication	44 (25)
Dyspnea	
Moderate Exertion	6 (4)
At Rest	1 (1)
History of Oral Steroid Use	8 (5)
Weight Loss	1 (1)
ASA Classification	
I	69 (40)
II	60 (35)
III	42 (24)
IV	3 (2)
Fragility Index	
0	127 (73)
1	31 (18)
2	11 (6)
3	3 2)
4	2 (1)
Mean Operative Time (minutes)	164.9±93.2

Total 73% of patients had a frailty index score of 0, while 17.8% of patients had a score of 1. The distribution of ASA class was mostly distributed across the first three categories, with 39.7% classified as ASA class 1, 34.5% as ASA class 2, and 24.1% as ASA class 3. (Table 1). We evaluated the performance of seven different algorithms, six ML and one logistic regression (LR) model, in predicting LOS, NHD, and AAE in total. In predicting extended LOS, SVM achieved an outstanding AUC of

0.987, sensitivity and specificity of 0.9, and overall accuracy of 92.5%, indicating extremely high predictive capability.

GB also performed well, with an outstanding AUC of 0.976, sensitivity and specificity of 1.0 and 0.9 respectively, and an overall accuracy of 94.34%. RF was predictive with an outstanding AUC of 0.979, specificity of 0.9, sensitivity of 1.0, and an overall accuracy of 92.5%. MLP achieved an outstanding AUC of 0.974, a specificity of 1.0, a sensitivity 0.5, and a predictive accuracy of 90.6%. GNB resulted in an outstanding AUC of 0.949, a specificity of 1.0, sensitivity of 0.5, and a predictive accuracy of 83.0%. DT achieved the lowest AUC of the ML algorithms with an excellent AUC of 0.856 with an accuracy of 84.9%. In comparison, LR, which yielded an outstanding AUC of 0.956, only achieved an accuracy of 79.3% for predicting LOS with a specificity and sensitivity of 0.9 and 0.5, respectively (Table 2).

Table 2: Performance metrics for machine learning algorithms. ROC-AUC, area under receiver-operator curve; PLR, positive likelihood ratio; NLR, negative likelihood ratio. PLR recorded as N/A when unable to be calculated.

Outcomes & Algorithms	ROC-AUC	Specificity	Sensitivity	Accuracy (%)	PLR	NLR
Non-home Discharge						
Random Forest Classifier	0.980	1.0	0.3	94.3	1.47	0.680
Gradient Boosting Classifier	0.940	1.0	0.3	96.2	1.50	0.667
Decision Tree Classifier	0.907	0.4	1.0	45.3	N/A	0.000
SVM Classifier	0.427	0.8	0.3	81.1	1.26	0.794
Gaussian Naive Bayes Classifier	0.867	1.0	0.7	96.2	2.94	0.340
Multi-Layer Perceptron	0.787	1.0	0.7	90.6	N/A	N/A
Logistic Regression	0.507	1.0	0.0	92.5	0.980	1.02
Extended Length of Stay						
Random Forest Classifier	0.979	0.9	1.0	92.5	41.9	0.000
Gradient Boosting Classifier	0.976	0.9167	1.0	94.3	N/A	N/A
Decision Tree Classifier	0.856	0.8	0.9	84.9	7.09	0.141
SVM Classifier	0.987	0.9	0.9	92.5	15.6	0.064
Gaussian Naive Bayes Classifier	0.949	1.0	0.5	83.0	2.07	0.484
Multi-Layer Perceptron	0.974	1.0	0.5	90.6	2.07	0.484
Logistic Regression	0.956	0.9	0.5	79.3	1.78	0.561
Any Adverse Event						
Random Forest Classifier	0.875	0.9	0.6	84.9	2.19	0.457
Gradient Boosting Classifier	0.854	1.0	0.0	90.6	1.00	1.00
Decision Tree Classifier	0.523	0.9	0.2	83.0	1.12	0.893
SVM Classifier	0.788	1.0	0.6	94.3	2.45	0.409
Gaussian Naive Bayes Classifier	0.813	1.0	0.4	94.3	1.67	0.600
Multi-Layer Perceptron	0.963	1.0	0.4	88.7	1.67	0.600
Logistic Regression	0.712	0.8	0.6	77.4	1.98	0.505

When examining AAE, all ML algorithms, except for DT, demonstrated superior performance in comparison to LR. Notably, MLP exhibited the highest performance, with an outstanding AUC of 0.963, a specificity of 1.0, a sensitivity of 0.4, and a predictive accuracy of 88.7%. Following closely behind, RF achieved an excellent AUC of 0.875 with an accuracy of 84.9%, while GB achieved an excellent AUC of 0.854 with an accuracy of 90.6%. Notably, RF achieved a sensitivity and specificity of 0.6 and 0.9, respectively. GNB also performed well with an excellent AUC of 0.813, a specificity of 1.0, a sensitivity of 0.4, and an impressive accuracy of 94.3%. SVM was able to achieve an acceptable AUC of 0.788, which notably is above the accepted threshold of 0.70. DT was the only algorithm to not achieve an AUC greater than the 0.774 that LR achieved, with a disappointing 0.523. Notably, DT achieved a predictive accuracy of 83.0% whereas LR achieved a predictive capability of 77.4% (Table 2).

Regarding NHD prediction, RF demonstrated an outstanding AUC of 0.980, a specificity of 1.0, a sensitivity of 0.3, and an impressive overall accuracy of 94.3%. GB also performed well with an outstanding AUC of 0.940 and 96.2% overall predictive accuracy. Following closely behind, GNB achieved an excellent AUC 0f of 0.867 and an accuracy of 96.2% with sensitivity and specificity values of 0.7 and 1.0, respectively. MLP achieved an acceptable AUC of 0.787, which is above the threshold for significance, and an accuracy of 90.6% and achieved sensitivity and specificity values of 0.7 and 1.0, respectively. DT achieved an outstanding AUC of 0.907 with a sensitivity of 1.0 and a specificity of 0.4, however, only achieved an accuracy of 45.3%. SVM (AUC of 0.427, specificity of 0.8, sensitivity of 0.3, and accuracy of 81.1%) was the only ML to not achieve an AUC value greater than that of LR, which achieved an AUC of 0.507 with an accuracy of 92.5%. Notably, LR achieved a sensitivity of 0.0 and a specificity of 1.0 (Table 2).

Table 3: Comparative demographics between patients who did vs did not experience adverse outcomes of interest, including permutation feature importance of variables predictive of outcomes, as determined by the highest performing algorithm, p<0.05 considered significant.

Outcomes and variables	Experienced adverse event (Count or Average±SD)	Did not experience event (Count or Average±SD)	Permutation feature importance	P value
Discharge location (RF)	**Non-home discharge**	**Home discharge**		
Age (years)	71.09±12.63	37.15±21.07	0.0613	<0.001
Operative time (minutes)	289.73±169.11	156.48±79.92	0.0093	<0.001
Dyspnea: moderate exertion N (%)	3 (2)	160 (92)	0.0073	<0.001
Length of stay (SVM)	**Extended length of stay**	**Expected length of stay**		
Age (years)	63.39±17.81	27.86±13.06	0.1113	<0.001
Operative time (minutes)	236.34±116.36	131.00±53.71	0.0783	<0.001
ASA class 3 N (%)	35 (83.3)	111 (84.1)	0.0360	<0.001
Any adverse event (MLP)	**Experienced adverse event**	**No adverse event**		
Operative time (minutes)	291.07±153.93	153.00±75.92	0.0900	<0.001
Dyspnea: moderate exertion N (%)	4 (2)	168 (97)	0.0246	<0.001
ASA class 1 N (%)	13 (8)	130 (75)	0.0217	0.001

RF, Random Forest; SVM, SVM Classifier; MLP, Multi-layer Perceptron; SD, Standard Deviation.

To identify key variables influencing outcome prediction, we utilized the PFI technique of the top-performing algorithms.

Table 4: Adverse events.

Characteristic	N (%)
Non-home discharge	11 (6)
Extended length of stay (> 1 day)	56 (32)
All adverse events	15 (9)
Surgical site infection	2 (1)
Renal complications	1 (1)
Sepsis	3 (2)
Intubation	1 (1)
Transfusion	8 (5)
Pneumonia	0 (0)
Deep vein thrombosis	0 (0)
Urinary tract infection	5 (3)
Cardiac arrest	1 (1)
Myocardial infarction	0 (0)
Return to operating room	2 (1)
Death	1 (1)

Given the substantial variability in algorithm performance, our analysis focused solely on the PFI of the most predictive algorithm in each category as shown in (Table 3). For LOS, SVM identified age and operative time to be most predictive with PFI values of 0.111 (p<0.001) and 0.078 (p<0.001), respectively.

Regarding NHD, RF also identified age and operative time as the most predictive variables with PFI values of 0.061 (p<0.001) and 0.009 (p<0.001), respectively. Lastly, MLP identified operative time (PFI of 0.090, p<0.001) and dyspnea (PFI of 0.025, p<0.001) as the most important predictors in AAE (Table 3). The average age of those discharged home was 37.15±21.07 with an average operative time of 156.48±79.92, whereas the average age of those who were not was 71.09±12.63 with an average operative time of 289.73±169.11.

Regarding LOS, the average age for those who had an average LOS was 27.86±13.06 with an average operative time of 131.00±53.71, whereas the average age for those who experienced an extended LOS was 63.39±17.81 with an average operative time of 236.34±116.36. Lastly, regarding AAE, the average age of those who did not experience AAE was 36.75±21.09 with an average operative time of 153.00±75.92, whereas the average age of those who did experience AAE was 66.27±15.05 with an average operative time of 291.07±153.93 (Table 3).

DISCUSSION

To our knowledge, there is a notable gap in the existing literature pertaining to the prediction of adverse events following UCL reconstruction. The high-level overhead athlete sub-population boasts extensive and comprehensive literature, likely owing to the competitive and financial implications of performance outcomes following surgical reconstruction. For instance, a study by Cain et al investigated a cohort of 1,281 high level athletes who underwent UCL reconstruction, revealing an impressive 83% rate of return to pre-injury performance within one year of the procedure.[27] This finding aligns with a substantial body of literature reporting similar positive outcomes in professional athletes.[28-30] Additionally, Osbahr et al presented data indicating that patients who underwent UCL reconstruction experienced an average career longevity of 2.9 years at their prior level, with retirement predominantly attributed to factors unrelated to their UCL, such as preexisting shoulder issues or earlier surgical interventions during their athletic careers.[31] Management of valgus elbow instability depends on several clinical factors and patient characteristics. However, the rate of successful RTS for athletes undergoing conservative treatment remains low and ranges from 42 to 54%.[11,32] As a result, UCL reconstruction has gained prominence due to the ability to effectively restore elbow stability and significantly improve the chances of successful RTS.[1,7,8,28-30] Therefore, UCL reconstruction is generally recommended for definitive treatment in high level athletes, and for non-high level athletes if they have failed non-operative management and experience residual functional limitations, persistent pain, and/or strong motivation to engage in labor intensive activities at their previous level.[2,33] UCL reconstruction can be performed by a variety of nuanced techniques, including the Modified Jobe, docking, and interference screw fixation, all with the end goal of restoring the competency of the medial elbow. This procedure utilizes tendon grafts, most commonly a palmaris longus autograft, to reconstruct the anterior band of the UCL. The graft is positioned and secured using bone tunnels or specialized fixation devices, aiming to replicate the anatomical and functional properties of the UCL.[1,7,8,28,29] The most common complication following UCL reconstruction is ulnar neuropathy, followed by infection, elbow stiffness, and graft rupture, regardless of the surgical method employed.[2,33,34] RTS is the most prevalent outcome evaluated after surgical UCL reconstruction, and current research focuses almost exclusively on the professional baseball player.[30] The results of this study present evidence supporting the precision and practicality of machine learning algorithms in accurately predicting adverse outcomes and identifying specific preoperative variables that are informative for clinical decision-making in a broad patient database. In contrast to subjective pre-operative tools like ASA and the Charlson Comorbidity index, which rely on an amalgam of non-specific variables to evaluate patient complexity, our ML algorithms leverage objective clinical data that is readily available upon admission and is specific to the procedure of interest. This utilization of objective data significantly enhances the effectiveness of ML algorithms in predicting outcomes. Moreover, the ability of machine learning algorithms to digest vast and nuanced clinical datasets to predict postoperative outcomes contributes to their potential for personalized medicine and tailored treatment strategies. As previously mentioned, the emphasis of post-UCL reconstruction outcomes has largely been characterized in the young athlete population. While these previous studies collectively contribute to our understanding of the long-term outcomes and factors influencing elite players' ability to successfully regain their pre-injury level of performance, there remains a paucity of literature that characterizes outcomes in non-high level athletes, including those with chronic, subclinical elbow instability. Our study, in which the average patient age is 39±22 years (Table 1), provides broad insight into the management and outcomes of patients, regardless of age, activity level, or mechanism of injury, who undergo elective, ambulatory UCL reconstruction for valgus elbow instability. Additionally, it also validates the predictive efficacy of our machine learning construct and contributes to the broader collection of literature regarding the utility of machine learning as a means for optimizing clinical management. Our study demonstrates that clinicians can achieve outstanding predictive accuracy for infrequent surgical procedures, such as UCL reconstruction, which, despite its increasing prevalence, represents a smaller portion of the overall yearly surgical procedures conducted. In this study, we queried ACS-NSQIP for elective UCL reconstruction procedures to minimize the effect of confounding variables (i.e., polytrauma patient with elbow dislocation) on patient population and outcomes. Despite the relative safety of ambulatory surgery, approximately 9% of patients experienced an adverse event, 32% needed to be admitted and stay longer than 1 day, and about 6% were unable to be discharged home (Table 4). For AAE, each ML algorithm outperformed LR in its predictive ability. Notably, MLP demonstrated a 14.6% increase in accurate AAE prediction when compared to LR, and also achieved an AUC that was 35% higher (Table 2). Interestingly, results of the PFI in MLP implicated operative time, followed by dyspnea and ASA class 1 as the greatest predictors of AAE (Table 3). LR performed relatively well in its predictive abilities of length of stay >1 day and non-home discharge. However, machine learning algorithms still outperformed LR in their predictive ability, with SVM classifier predicting LOS with a 16.6% increase in accuracy and AUC that is 3.2% higher than LR, and RF predicting non-home discharge with a 1.9% increase in accuracy and an AUC that is 93.3% higher (Table 2). PFI yielded increased age and operative time as two most important variables for predicting NHD and LOS, which is intuitive, and reflects a possible precautionary response by providers to admit relatively older patients for observation following longer and/or more complex surgeries. These statistically significant PFI variables may be independently considered by the provider when planning surgery and can be utilized to import readily available patient variables and create an automated risk profile for outcomes of interest prior to undergoing UCL reconstruction. Additionally, the results of the machine learning construct can only be used to interpret and infer outcomes by the physician as needed prior to surgery to help guide clinical decision making. For example, the positive likelihood ratio of RF for LOS is 41.9; so while only 32% of our cohort experienced NHD, the implied odds rise to 93.1% if RF positively predicts a NHD (Table 2). In this case, a pre-operative conversation that includes LOS, NHD, etc. can be discussed to help set expectations and pre-emptively allow opportunity to make necessary home/social arrangements. This study has several limitations primarily due to the small sample size, which can be attributed to UCL reconstruction being uncommon as a surgical procedure. In order to maintain a focused dataset, we only included 174 patients who voluntarily underwent the surgical procedure, excluding polytrauma cases from our dataset. Furthermore, the use of NSQIP to gather data restricts the study to a 30-day post-operative outcome assessment. This prevents the inclusion of long-term complications, such as retear and ulnar nerve paraesthesia that are commonly evaluated in studies extending well beyond the 30-day timeframe available for analysis.[6,27,28,35]

While UCL reconstruction is generally considered a safe surgical procedure, the increasing prevalence of this surgery necessitates a deeper investigation into the associated risks. Previous research has focused on high level athletes and their ability to return to their sport of choice, revealing promising outcomes that have contributed to the rising popularity of the surgery. However, there remains a significant gap in the literature, as individuals who are not-high level athletes also choose to undergo this procedure for improvements in pain management and joint mobility, leading to an enhanced quality of life. Our machine learning model demonstrated a high level of accuracy in predicting outcomes such as extended length of stay, non-home discharge, and any adverse event. Notably, age emerged as a significant variable strongly associated with each of the categories. These findings provide valuable insights for individuals considering elective UCL reconstruction, enabling them to evaluate the potential risks associated with the surgical procedure.

The integration of machine learning into this research represents a significant advancement in understanding the outcomes and risks of UCL reconstruction beyond the exclusive context of high-level athletes. By incorporating a broader range of patients, our study contributes to a more comprehensive understanding of the potential benefits and adverse events associated with this procedure, ultimately allowing for more informed decisions regarding UCL reconstruction.

CONCLUSION

The highly predictive capability of ML indicates the possibility to represent a procedure-specific complementary tool for the preoperative risk stratification process. This study provides a comprehensive analysis of UCL reconstruction in the management and outcomes of any patient, regardless of age or activity level.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Jensen AR, LaPrade MD, Turner TW, Dines JS, Camp CL. The History and Evolution of Elbow Medial Ulnar Collateral Ligament Reconstruction: from Tommy John to 2020. Curr Rev Musculoskelet Med. 2020; 13(3):349-360.
2. Erickson BJ, Harris JD, Chalmers PN, Bach BR Jr, Verma NN, Bush-Joseph CA, et al. Anatomy, Indications, Techniques, and Outcomes. Sports Health. 2015;7(6):511-7.
3. Nicolette GW, Gravlee JR. Ulnar collateral ligament injuries of the elbow in female division I collegiate gymnasts: a report of five cases. Open Access J Sports Med. 2018;9:183-9.
4. Liman MNP, Avva U, Ashurst JV. Elbow Trauma. Treasure Island: Stat Pearls Publishing; 2020.
5. Armañanzas R, Alonso-Nanclares L, Defelipe-Oroquieta J, Kastanauskaite A, de Sola RG, Defelipe J, et al. Machine learning approach for the outcome prediction of temporal lobe epilepsy surgery. PLoS One. 2013;8(4):e62819.
6. Carr JB, Camp CL, Dines JS. Elbow Ulnar Collateral Ligament Injuries: Indications, Management, and Outcomes. Arthroscopy. 2020;36(5):1221-2.
7. Chen FS, Rokito AS, Jobe FW. Medial elbow problems in the overhead-throwing athlete. J Am Acad Orthop Surg. 2001;9(2):99-113.
8. Loftice J, Fleisig GS, Zheng N, Andrews JR. Biomechanics of the elbow in sports. Clin Sports Med. 2004;23(4):519-30
9. Munnich EL, Parente ST. Procedures take less time at ambulatory surgery centers, keeping costs down and ability to meet demand up. Health Aff (Millwood). 2014;33(5):764-9.
10. Fabricant PD, Seeley MA, Rozell JC, Fieldston E, Flynn JM, Wells LM, et al. Cost Savings From Utilization of an Ambulatory Surgery Center for Orthopaedic Day Surgery. J Am Acad Orthop Surg. 2016;24(12):865-871.
11. Carlstrom LP, Helal A, Perry A, Lakomkin N, Graffeo CS, Clarke MJ. Too frail is to fail: Frailty portends poor outcomes in the elderly with type II odontoid fractures independent of management strategy. J Clin Neurosci. 2021;93:48-53.
12. Gowd AK, Agarwalla A, Amin NH, Romeo AA, Nicholson GP, Verma NN, et al. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. J Shoulder Elbow Surg. 2019;28(12): e410-e421.
13. Yu T, Zhu H. Hyper-parameter optimization: a review of algorithms and applications. ArXiv. 2020;10:485.

14. Van RG, Drake FL. Python 3 Reference Manual. Create Space. 2009.
15. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825-30.
16. Provider enrollment and certification- overnight stays in ambulatory surgical centers. center for medicare and medicaid services. Available at: https://www.cms.gov/Medicare/Provider-Enrollment-andCertification/SurveyCertificationGenInfo/downloads/SCLetter04-22.pdf. Accessed on 20 February 2023.
17. Gramfort A. Scikit-learn: Machine Learning in Python. JMLR. 2011;12:2825-30.
18. Gholamy A, Kreinovich V, Kosheleva O. Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. Mach Learn Res. 2020.
19. Yu T, Zhu H. Hyper-parameter optimization: a review of algorithms and applications. ArXiv. 2020;10:485.
20. Zhou J, Gandomi AH, Chen F, Holzinger A. Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics. Electronics. 2021;10(5):593.
21. Shreffler J, Huecker MR. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios. Treasure Island (FL): StatPearls Publishing; 2023.
22. Hosmer DW, Lemeshow S. Applied logistic regression, 2nd ed. USA: Wiley; 2000:156-64.
23. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Engineer. 2020;9(3), 90-5.
24. Kaneko H. Cross-validated permutation feature importance considering correlation between features. Analyt Sci Adv. 2022;3(9-10):278-87.
25. Cava W, Bauer C. Interpretation of machine learning predictions for patient outcomes in electronic health records. AMIA. 2019;10:572-81.
26. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics. 2010;26(10): 1340-7.
27. Cain EL, Andrews JR, Dugas JR. Outcome of Ulnar Collateral Ligament Reconstruction of the Elbow in 1281 Athletes: Results in 743 Athletes with Minimum 2-Year Follow-up. Am J Sports Med. 2010;38(12): 2426-34.
28. Erickson BJ, Gupta AK, Harris JD, Bush-Joseph C, Bach BR, Abrams GD, et al. Rate of return to pitching and performance after Tommy John surgery in Major League Baseball pitchers. Am J Sports Med. 2014; 42(3):536-43.
29. Jack RA 2nd, Burn MB, Sochacki KR, McCulloch PC, Lintner DM, Harris JD. Performance and Return to Sport After Tommy John Surgery Among Major League Baseball Position Players. Am J Sports Med. 2018;46(7):1720-6.
30. Glogovac G, Grawe BM. Outcomes With a Focus on Return to Play for Revision Ulnar Collateral Ligament Surgery Among Elite-Level Baseball Players: A Systematic Review. Am J Sports Med. 2019;47(11): 2759-63.
31. Osbahr DC, Cain EL, Raines BT, Fortenbaugh D, Dugas JR, Andrews JR. Long-term Outcomes After Ulnar Collateral Ligament Reconstruction in Competitive Baseball Players: Minimum 10-Year Follow-up. Am J Sports Med. 2014;42(6):1333-42.
32. Chauhan A, McQueen P, Chalmers PN, Ciccotti MG, Camp CL, D'Angelo J, et al. Nonoperative Treatment of Elbow Ulnar Collateral Ligament Injuries With and Without Platelet-Rich Plasma in Professional Baseball Players: A Comparative and Matched Cohort Analysis. Am J Sports Med. 2019;47(13):3107-19.
33. Ebert AB, Andrews JR. Ulnar Collateral Ligament (UCL) Reconstruction Technique. In: Bain G, Eygendaal D, van Riet R, eds. Surgical Techniques for Trauma and Sports Related Injuries of the Elbow. Berlin: Springer; 2007;10:1-30.
34. Clain JB, Vitale MA, Ahmad CS, Ruchelsman DE. Ulnar Nerve Complications After Ulnar Collateral Ligament Reconstruction of the Elbow: A Systematic Review. Am J Sports Med. 2019;47(5):1263-9.
35. Saper M, Shung J, Pearce S, Bompadre V, Andrews JR. Outcomes and Return to Sport After Ulnar Collateral Ligament Reconstruction in Adolescent Baseball Players. Orthop J Sports Med. 2018;6(4):232-5.

Cite this article as: Case TR, Rice RC, Cabrera AJ, Kricfalusi M, Dasari P, Jesurajan J, et al. Employing machine learning to predict adverse acute post-surgical outcomes following elective ulnar collateral ligament reconstruction. Int J Res Orthop 2023;9:888-95.