Original Research

Machine Learning Does Not Improve Humeral Torsion Prediction Compared to Regression in Baseball Pitchers

Garrett S Bullock[1] [a], Charles A Thigpen[2], Gary S Collins[3], Nigel K Arden[1], Thomas K Noonan[4], Michael J Kissenberth[5], Ellen Shanley[2]

[1] Department of Orthopaedic Surgery, Wake Forest School of Medicine; Centre for Sport, Exercise and Osteoarthritis Research Versus Arthritis, University of Oxford, [2] University of South Carolina Center for Rehabilitation and Reconstruction Sciences; ATI Physical Therapy, [3] Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford; Oxford University Hospitals NHS Foundation Trust, [4] Department of Orthopaedic Surgery, University of Colorado School of Medicine; University of Colorado Health, Steadman Hawkins Clinic, [5] Steadman Hawkins Clinic of the Carolinas

International Journal of Sports Physical Therapy

Background

Humeral torsion is an important osseous adaptation in throwing athletes that can contribute to arm injuries. Currently there are no cheap and easy to use clinical tools to measure humeral torsion, inhibiting clinical assessment. Models with low error and "good" calibration slope may be helpful for prediction.

Hypothesis/Purpose

To develop prediction models using a range of machine learning methods to predict humeral torsion in professional baseball pitchers and compare these models to a previously developed regression-based prediction model.

Study Design

Prospective cohort

Methods

An eleven-year professional baseball cohort was recruited from 2009-2019. Age, arm dominance, injury history, and continent of origin were collected as well as preseason shoulder external and internal rotation, horizontal adduction passive range of motion, and humeral torsion were collected each season. Regression and machine learning models were developed to predict humeral torsion followed by internal validation with 10-fold cross validation. Root mean square error (RMSE), which is reported in degrees (°) and calibration slope (agreement of predicted and actual outcome; best = 1.00) were assessed.

Results

Four hundred and seven pitchers (Age: 23.2 +/-2.4 years, body mass index: 25.1 +/-2.3 km/m^2, Left-Handed: 17%) participated. Regression model RMSE was 12° and calibration was 1.00 (95% CI: 0.94, 1.06). Random Forest RMSE was 9° and calibration was 1.33 (95% CI: 1.29, 1.37). Gradient boosting machine RMSE was 9° and calibration was 1.09 (95% CI: 1.04, 1.14). Support vector machine RMSE was 10° and calibration was 1.13 (95% CI: 1.08, 1.18). Artificial neural network RMSE was 15° and calibration was 1.03 (95% CI: 0.97, 1.09).

a **Corresponding Author:**
Garrett Bullock, PT, DPT, DPhil
Department of Orthopaedic Surgery and Rehabilitation
Wake Forest School of Medicine
email: gbllock@wakehealth.edu

Conclusion

This is the first study to show that machine learning models do not improve baseball humeral torsion prediction compared to a traditional regression model. While machine learning models demonstrated improved RMSE compared to the regression, the machine learning models displayed poorer calibration compared to regression. Based on these results it is recommended to use a simple equation from a statistical model which can be quickly and efficiently integrated within a clinical setting.

Levels of Evidence

2

INTRODUCTION

Baseball throwing generates high velocity and forces through the arm,[1–3] which contribute to unique osseous and soft tissue glenohumeral joint adaptations.[4,5] These shoulder adaptations contribute to changes in shoulder range of motion, specifically to increased external rotation and decreased internal rotation in comparison to the non-dominant arm.[6] These unique throwing specific shoulder adaptations have been associated with arm injuries in baseball players.[4,7]

While soft tissue adaptations affect throwing specific shoulder range of motion,[8] the underlying osseous structural transformations also contribute to throwing shoulder range of motion.[9] These osseous shoulder structural adaptations are termed humeral torsion (HT). HT is measured through the line that bisects the humeral head articular surface and the transepicondylar axis.[10] During youth and adolescence, the high humeral forces generated during pitching effect osseous growth and development, contributing to the diminution of humeral anatomical neck and head antetorsion that occurs with aging.[11] These structural adaptations are important for throwing development[12]; however, they are also linked to arm injury risk.[9]

Within clinical practice, HT can be calculated indirectly through ultrasonic methods.[5] However, this equipment is expensive, preventing many clinics and clinicians from assessing HT, hindering clinical examination. One way to arrive at clinical measures is through prediction modelling.[13] Statistical prediction modelling uses traditional regression based methods to obtain a risk or probability.[14] More recently, machine learning algorithms (such as random forests, gradient boosting machines, support vector machines, and neural networks) have been purported to offer increased flexibility to capture nonlinearities and higher order interactions.[15,16] Machine learning uses general purpose algorithms that identify data patterns, using minimal data assumptions,[17] and are being increasingly used in the medical setting.[18,19] As a result, there is widespread interest in exploring the usefulness of modern machine learning methods for increasing prediction accuracy compared to more regression based statistical approaches.[20]

Humeral torsion is an important osseous adaptation in throwing athletes that can contribute to arm injuries.[9] Machine learning algorithms offer an alternative strategy to predict outcomes in data with high complexity. Comparing and contrasting regression based statistical and machine learning approaches can help identify the most promising prediction model to be implemented in the clinical setting.

Therefore, the purpose of this study was to develop prediction models using a range of machine learning methods to predict professional baseball pitcher HT and compare these models to a traditional regression-based prediction model.

METHODS

STUDY DESIGN

A prospective cohort study was conducted from 2009 to 2019 on Minor League pitchers in one Major League Baseball organization. Only preseason data were utilized in this study and pitchers were only included once within the dataset. Participants were excluded from the study if 1) the athlete played a primary position other than pitcher, 2) they were being treated for a shoulder or elbow injury at the beginning of the season, or 3) they were unable to participate on the first day of practice because of upper extremity injury. Prior to data collection, all participants were informed of the risks and benefits of study participation and participants gave verbal and written consent to study participation. The PRISMA health system Institutional Review Board approved this study.

DATA COLLECTION

Before the beginning of the season, all baseball players were questioned for arm dominance, prior baseball experience, injury history, and position(s) played. Participants were then examined for current height (cm) and mass (kg). Participants were then examined for passive shoulder PROM and HT. Shoulder PROM testing was randomized for each participant, and examiners were blinded to hand dominance throughout the study.[21] Two examiners performed all measurements for the entire cohort.

PREDICTORS

Predictors included player demographics (age, hand dominance, previous baseball participation, injury history, position played, and continent of origin), shoulder PROM, and injury history. Shoulder ROM and injury history are further described below.

SHOULDER RANGE OF MOTION

All shoulder PROM (external rotation [ER], internal rotation [IR] and horizontal adduction [HA]) was measured supine on a standardized plinth table by two examiners using a digital inclinometer per previously described methods.[22–25]

Two trials were performed per shoulder measurement, and the average of these two trials was used for data analysis. Shoulder PROM was calculated on 10 participants prior to data collection for the two examiners. Shoulder PROM intra- and inter-rater reliability was excellent for ER and IR ($ICC_{(2,1)}$ and $ICC_{(2,k)}$ = 0.92-0.99) and HA ($ICC_{(2,1)}$ and $ICC_{(2,k)}$ = 0.92-0.99), and the standard error of measurement was 2°-4° for shoulder ER, IR, and HA.

INJURY HISTORY

A shoulder or elbow injury was defined as any traumatic or overuse injury that occurred during any baseball team sponsored activity (from the beginning of preseason through the last post season game) to any muscle, joint, tendon, ligament, bone, or nerve that required medical attention.[26] Injuries were further designated by dominant and nondominant arm. An independent examiner, blinded to physical measurements, reviewed medical documents to determine the diagnosis, duration of treatment, and the time to clearance for return to full sport participation.

OUTCOME

HUMERAL TORSION (HT)

Dominant HT was measured supine on a standardized plinth table with the shoulder in 90° of abduction. One examiner, using a 5 mHz ultrasonographic transducer (Sonosite Inc, Bothell, WA, USA) measured HT. The ultrasonographic transducer was placed level, confirmed with a bubble level, on the anterior shoulder, perpendicular to the long axis of the humerus. The humerus was then rotated until the apexes of the greater and lesser tubercles could be visualized parallel to the horizontal plane. The second examiner placed a digital inclinometer on the ulnar side of the forearm, measuring the forearm inclination angle with respect to the horizontal, which indirectly measures HT.[5] Two trials were performed per HT measurement, and the average of these two trials was used for data analysis. HT reliability was calculated on 10 participants prior to data collection for the two examiners. Humeral torsion intra- and inter-rater reliability was excellent ($ICC_{(2,1)}$ and $ICC_{(2,k)}$ 0.93-0.97) and the standard error of measure was 2-4°.

STATISTICAL ANALYSES

All data were investigated for missingness prior to analyses, using the R package *naniar*. Missing data were low (Shoulder ROM: 3%, age: <1%, HT: 2%), thus complete case analyses were performed. Descriptive statistics were reported by mean (standard deviation), median (interquartile range), and frequencies and percentages for categorical variables.

SAMPLE SIZE CONSIDERATIONS

For the statistical modelling, an *a priori* sample size calculation was performed with the R package *pmsampsize*.[27] Referencing a previous meta-analysis and meta-regression,[9] mean HT was 28°, standard deviation was 4°, and R^2 was 0.38. The *a priori* statistical regression prediction model was determined to incorporate ten degrees of freedom (i.e., pa-

rameters). As a result, a total of 246 baseball players were required to reduce the risk of overfitting.

MODEL DEVELOPMENT

The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (*TRIPOD*) were followed for all model development.[28]

STATISTICAL MODEL

A linear regression model to predict HT was developed, using predictor variables including: Predictor variables included: (1) age,[29] (2) arm dominance (Left or Right handed),[30,31] (3) shoulder IR,[9] (4) shoulder ER,[9] (5) shoulder HA,[32] (6) continent of origin (North America or Latin America),[33] (7) previous shoulder or elbow injury.[34] Linearity was not assumed; as a result, continuous predictors were assessed for non-linearity with restricted cubic splines. Restricted cubic splines were calculated with three, four, and five knots with the R package *rms*. All continuous predictors demonstrated a linear relationship to HT. Interactions were also analyzed, with no predictors observed to have an interaction relationship with HT. Internal validation was performed with a 10-fold cross validation. Internal validation is performed to reduce optimism bias, as models are overly optimistic on the developed dataset.[35,36] The R package *caret* was used to performed cross validation.

Four machine learning models (Random Forest, Gradient Boosting Machine [GBM], Support Vector Machine Regression [SVM], and Artificial Neural Networks [ANN]) were developed to predict HT using an iterative hyperparameter tuning process. Hyperparamter tuning consisted of using a grid search process. All machine learning models incorporated all the same predictors used to develop the linear regression model. The R packages *randomForest*, *gbm, kernlab* and *e1071,* and *neuralnet* were used for random forest, GBM, SVM, and ANN models. For full description of the machine learning models, tuning variables, final hyperparameters, and complete code, please refer to the Appendix. Following model development, all machine learning models, besides ANN, were internally validated with a 10-fold cross validation. The ANN model was replicated 100 times.

Model performance was assessed with root mean square error (RMSE), calibration and R^2. Root mean square error is the error of the model reported in outcome units (i.e., degrees), with lower error demonstrating improved prediction performance. Calibration is the agreement of predicted and actual outcome (i.e., HT), with a calibration of 1 equalling best calibration.[35,36]

RESULTS

A total of 407 pitchers with a mean age of 23.2 years (sd = 2.4), BMI of 25.1 km/m^2 (sd = 2.3) were eligible and included (Table 1).

GENERALIZED LINEAR REGRESSION MODEL

Final model RMSE was 12°, calibration was 1.00 (95% CI: 0.94, 1.06); Table 2; Figure 1A), and R^2 was 0.41. The mean

Table 1. Pitcher demographics, presented as mean (SD) or percentage.

	Professional Pitchers (n = 407)
Age (years)	23.2 (2.4)
Hand Dominance Left Right	 17% 83%
BMI (kg/m2)	25.1 (2.3)
Arm Injury History Prevalence	43%
Dominant Humeral Torsion (°)	8.2 (12.7)
Dominant Internal Rotation (°)	35.2 (11.4)
Dominant External Rotation (°)	126.9 (10.9)
Dominant Horizontal Adduction (°)	-1.4 (13)
Nondominant Humeral Torsion (°)	25.7 (13.0)
Nondominant Internal Rotation (°)	48.1 (10.6)
Nondominant External Rotation (°)	118.3 (11.6)
Nondominant Horizontal Adduction (°)	16.5 (14.6)

Table 2. Statistical and Machine Prediction Model Performance

Prediction Model	Root Mean Square Error	Calibration Slope
Generalized Linear Regression	12°	1.00 (95% CI: 0.94, 1.06)
Random Forest	9°	1.33 (95%CI: 1.29, 1.37)
Gradient Boosting Machine	9°	1.09 (95% CI: 1.04, 1.14)
Support Vector Machine Regression	10°	1.13 (95% CI: 1.08, 1.18)
Artificial Neural Network	15°	1.03 (95% CI: 0.97, 1.09)

distribution of the final model linear predictors was 17°, the standard deviation was 10°, the minimum was -19°, and the maximum was 48°. For full model report, please refer to the Appendix.

MACHINE LEARNING MODELS

The random forest and GBM demonstrated the best RMSE (Table 2). The random forest demonstrated the worst calibration (Table 2) and the ANN demonstrated the best calibration (Table 2; Figure 1B). The mean distribution of the final model linear predictors was 16° to 17°, the standard deviation was ranged from 9° to 10°, the minimum ranged from -2°1 to -11°, and the maximum was ranged from 44° to 52°. For each calibration plot and a full pictorial description of the final ANN architecture, please refer to the Appendix.

DISCUSSION

The machine learning models, besides ANN, demonstrated improved RMSE, compared to the statistical prediction model. Interestingly, the random forest and GBM RMSE difference compared to the linear regression model was similar to the HT standard error of measure (2-4°). However, all machine learning models demonstrated poor calibration compared to the linear regression prediction model. All prediction models demonstrated similar mean and variance calculations for predicted values. These findings suggest that prediction model performance should be evaluated through multiple performance metrics.

The machine learning models demonstrated decreased RMSE compared to the linear regression model. RMSE reports the average prediction model error in the units of the outcome of interest, which in this case is degrees of HT.[37] This allows for a clinically pertinent and interpretable comparison of model performance. The random forest and GBM demonstrated decreased RMSE similar to the reliability HT standard error of measure, which may demonstrate a clinically significant difference. Both the random forest and GBM methods employ ensemble methods to generate prediction models.[38,39] Ensemble methods have been shown to increase overall prediction precision due to the meta-aggregation of multiple models, allowing for increased generalizability in highly complex data.[40] Further, the SVM model demonstrated a RMSE difference just below to the standard error of measure in comparison to the statistical model. SVMs utilize spatial kernel-based methods to inform predictions. Due to the individuality affecting HT development,[12] the visual hyperplane demarcation methods used by SVM may generate improved HT prediction.

Although ML methods demonstrated decreased RMSE, calibration was poor. All machine learning methods demonstrated worse calibration compared to the statistical model, with the ANN a three-point slope difference. Calibration as-

sesses the prediction outcome versus the actual outcome, and is important in understanding the accuracy of predictions.[41] Over calibration has been reported as potentially harmful in the clinical setting, with miscalibration above 5% potentially affecting clinical decisions.[42] These worse calibration performing machine learning methods, besides the ANN model, demonstrated a calibration slope in excess of 1.09, with the random forest model having a calibration slope of 1.33. Upon visual inspection of the calibration plots, all three models had significant demarcation at both tails of the calibration slope. These calibration discrepancies may be due to the biological volatility of individual outliers. Baseball players may have different genetic, environmental, and overall baseball loading factors, which all contribute to HT. Due to the algorithmic nature of machine learning, these outliers may have indiscriminately affected overall calibration. However, the ANN model had similar calibration compared to the statistical mode. ANN's are high performers in predictions involving complex and multiple interaction data.[43] As stated above, the complex issue individual variability, may allow for ANN's to demonstrate high calibration performance.

All machine learning and linear regression models demonstrated similar mean and variance of predicted outcomes. These predictions are greater than those reported in a previous meta-analysis.[9] While all models demonstrated similar prediction HT outcomes, there were distinct differences in model performance parameters. These findings highlight that model performance should be evaluated on multiple parameters, and not just on one specific performance finding. Clinicians need to integrate multiple prediction model performance outcomes, including discrimination, calibration, and model error, where appropriate, when evaluating the efficacy of a prediction model.[41]

These findings warrant future research. External validation is required to evaluate the generalizability of these models. HT development may be influenced by the volume and speed of throwing.[12] Further research is needed to decipher if incorporating lifetime baseball exposure and throwing velocity could aid in prediction model precision. Other genetic factors such as collagen phenotype and familial history may also affect HT. Incorporating these predictors would be beneficial in evaluating the prediction ability of these models. Finally, the clinical utility of these prediction models needs to be evaluated. Understanding how these models may affect clinical practice and decisions in comparison to standard evidence-based practice is needed.

CLINICAL IMPLICATIONS

Model RMSE ranged from 9° to 15° for all models, with the statistical regression model RMSE was 12. The HT standard error of measure in professional pitchers is 2 degrees.[44] Professional pitchers with 5 degrees HT difference between their throwing and non-throwing arms has been previously determined to pose greater risk for arm injury.[45] As RMSE was reported in degrees, the RMSE may be beyond the clinically important error, and affect pitching arm risk assessment.[5] However, arm injury examination encompasses multiple factors,[4,46,47] and this HT prediction model could be used in conjunction of multiple other clinical tests and
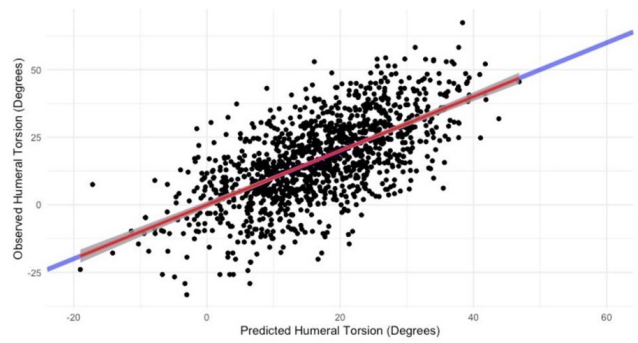
A. Regression Calibration Plot
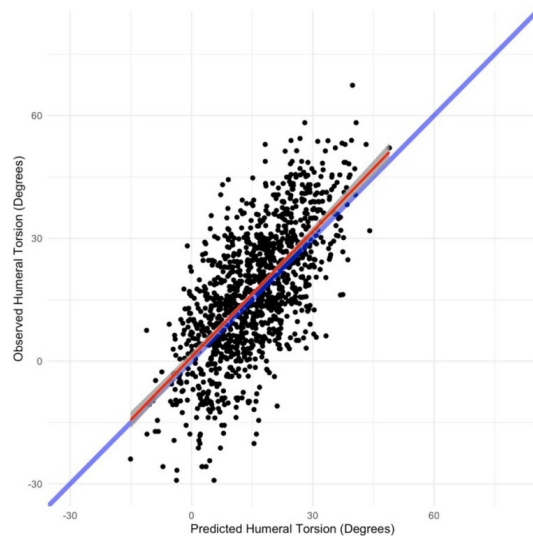
B. Artificial Neural Network Calibration Plot

Figure 1. Calibration Plot for Regression (A) and Artificial Neural Network (B)

The blue line depicts perfect calibration, while the red line reports actual calibration.

measures in order to prescribe a personalized injury mitigation program.

PRACTICAL EXAMPLE

To aid in clinical applicability an example is described. As the machine learning models did not improve HT prediction, the ease of use and interpretability of the statistical model is recommended for clinical implementation. The statistical model is calculated through a mathematical equation to predict HT. This equation can be inputted into a standardized Excel or other basic computer program. For example, consider a 22-year-old right-handed pitcher from North America, with 35 degrees of IR, 103 degrees of ER, and 2 degrees of HA. During the clinical interview, the pitcher did not report any current or prior arm injuries. Using the equation reported in the supplement: 33.01 (The Intercept) + 22*0.15 (Age) − 1.83 (Right-Handed) + 35*0.37 (IR) - 103*0.30 (ER) + 2*0.31 (HA) + 0 (North America) + 0 (No Injury History) the model predicts this pitcher's right HT is 18 degrees.[48]

STRENGTHS AND POTENTIAL LIMITATIONS

This study utilized a large sample of professional baseball pitchers that exceeded the *a priori* required sample size which increases the precision of these results. Multiple models were performed, incorporating both machine learning and statistical prediction model techniques, which increases the comparability of these findings. Internal validation was performed on all findings, allowing for a realistic optimism corrected model estimate, increasing the validity of these results. External validation was not performed on these models, decreasing the generalizability of these models. While many machine learning methods suggest splitting data into training and testing sets,[49,50] this decreases the power and precision of these models.[51,52] While this data met the *a priori* sample size calculations for linear regression, this sample size calculation may be too small for machine learning models.[53] Further, these data did not allow for a training and testing split to maintain proper power. Previous authors[51,52,54] have recommended to utilize all data during model development, and use robust internal validation methods to correct for optimism. As a result, cross-validation was used for internal validation on these prediction models.

CONCLUSIONS

Machine learning models demonstrated improved root mean square error and poorer calibration compared to the statistical model. Machine learning did not improve HT prediction in professional baseball players compared to a traditional statistical model. The root mean square error of all models was greater than the standard error of measure and clinically important difference, which may hinder the clinical usefulness of these models. It is recommended that clinicians use the statistical model in practice in conjunction with other examination factors, as this model provides an easy-to-use equation, that can quickly and efficiency be integrated within a clinical setting. Future research is needed to evaluate if environmental and genetic factors can improve HT prediction.

REFERENCES

1. Reinold MM, Macrina LC, Fleisig GS, Aune K, Andrews JR. Effect of a 6-Week weighted baseball throwing program on pitch velocity, pitching arm biomechanics, passive range of motion, and injury rates. *Sports Health*. 2018;10(4):327-333. doi:10.1177/1941738118779909

2. Fleisig GS, Barrentine SW, Zheng N, Escamilla RF, Andrews JR. Kinematic and kinetic comparison of baseball pitching among various levels of development. *J Biomech*. 1999;32(12):1371-1375.

3. Fleisig GS, Andrews JR, Dillman CJ, Escamilla RF. Kinetics of baseball pitching with implications about injury mechanisms. *Am J Sport Med*. 1995;23(2):233-239.

4. Wilk KE, Macrina LC, Fleisig GS, et al. Deficits in glenohumeral passive range of motion increase risk of shoulder injury in professional baseball pitchers: a prospective study. *Am J Sport Med*. 2015;43(10):2379-2385.

5. Noonan TJ, Thigpen CA, Bailey LB, et al. Humeral torsion as a risk factor for shoulder and elbow injury in professional baseball pitchers. *Am J Sports Med*. 2016;44(9):2214-2219. doi:10.1177/0363546516648438

6. Reinold MM, Gill TJ. Current concepts in the evaluation and treatment of the shoulder in overhead-throwing athletes, part 1: physical characteristics and clinical examination. *Sports Health*. 2010;2(1):39-50.

7. Wilk KE, Macrina LC, Fleisig GS, et al. Correlation of glenohumeral internal rotation deficit and total rotational motion to shoulder injuries in professional baseball pitchers. *Am J Sport Med*. 2011;39(2):329-335.

8. Reinold MM, Wilk KE, Macrina LC, et al. Changes in shoulder and elbow passive range of motion after pitching in professional baseball players. *Am J Sport Med*. 2008;36(3):523-527.

9. Helmkamp JK, Bullock GS, Rao A, Shanley E, Thigpen C, Garrigues GE. The relationship between humeral torsion and arm injury in baseball players: A systematic review and meta-analysis. *Sports Health*. 2020;12(2):132-138.

10. Krahl VE. The torsion of the humerus: its localization, cause and duration in man. *Am J Anatomy*. 1947;80(3):275-319.

11. Sabick MB, Kim YK, Torry MR, Keirns MA, Hawkins RJ. Biomechanics of the shoulder in youth baseball pitchers: implications for the development of proximal humeral epiphysiolysis and humeral retrotorsion. *Am J Sport Med*. 2005;33(11):1716-1722.

12. Whiteley R, Adams R, Ginn K, Nicholson L. Playing level achieved, throwing history, and humeral torsion in Masters baseball players. *J Sport Sci*. 2010;28(11):1223-1232.

13. Hudda MT, Fewtrell MS, Haroun D, et al. Development and validation of a prediction model for fat mass in children and adolescents: meta-analysis using individual participant data. *BMJ*. 2019;366:l4293.

14. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:b375.

15. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. *J Clin Epidemiol*. 2016;76:175-182.

16. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med*. 2016;35(2):214-226.

17. Bzdok D, Altman N, Krzywinski M. Points of significance: statistics versus machine learning. *Nature*. Published online 2018:1-7.

18. Mccarthy JF, Marx KA, Hoffman PE, et al. Applications of machine learning and high-dimensional visualization in cancer detection, diagnosis, and management. *Ann New York Acad Sci*. 2004;1020(1):239-262.

19. Lisboa PJ, Taktak AFG. The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Networks*. 2006;19(4):408-415.

20. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *New Eng J Sport Med*. 2016;375(13):1216.

21. Shanley E, Rauh MJ, Michener LA, Ellenbecker TS, Garrison JC, Thigpen CA. Shoulder range of motion measures as risk factors for shoulder and elbow injuries in high school softball and baseball players. *Am J Sport Med*. 2011;39(9):1997-2006.

22. Ellenbecker TS, Roetert EP, Bailie DS, Davies GJ, Brown SW. Glenohumeral joint total rotation range of motion in elite tennis players and baseball pitchers. *Med Sci Sports Exerc*. 2002;34(12):2052-2056.

23. Tyler TF, Roy T, Nicholas SJ, Gleim GW. Reliability and validity of a new method of measuring posterior shoulder tightness. *J Orthop Sports Phys Ther*. 1999;29(5):262-269; discussion 270-4.

24. Shanley E, Thigpen C, Clark JC, et al. Changes in passive range of motion and development of glenohumeral internal rotation deficit (GIRD) in the professional pitching shoulder between spring training in two consecutive years. *J Shoulder Elbow Surg*. 2012;21(11):1605-1612. doi:10.1016/j.jse.2011.11.035

25. Shanley E, Kissenberth MJ, Thigpen CA, et al. Preseason shoulder range of motion screening as a predictor of injury among youth and adolescent baseball pitchers. *J Shoulder Elbow Surg*. 2015;24(7):1005-1013. doi:10.1016/j.jse.2015.03.012

26. Rauh MJ, Margherita AJ, Rice SG, Koepsell TD, Rivara FP. High school cross country running injuries: a longitudinal study. *Clin J Sport Med*. 2000;10(2):110-116.

27. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368.

28. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594. doi:10.1136/bmj.g7594

29. Kibler WB, Chandler TJ, Livingston BP, Roetert EP. Shoulder range of motion in elite tennis players: effect of age and years of tournament play. *Am J Sport Med*. 1996;24(3):279-285.

30. Takenaga T, Goto H, Sugimoto K, et al. Left-handed skeletally mature baseball players have smaller humeral retroversion in the throwing arm than right-handed players. *J Shoulder Elbow Surg*. 2017;26(12):2187-2192.

31. Takeuchi S, Yoshida M, Sugimoto K, Tsuchiya A, Takenaga T, Goto H. The differences of humeral torsion angle and the glenohumeral rotation angles between young right-handed and left-handed pitchers. *J Shoulder Elbow Surg*. 2019;28(4):678-684.

32. Myers JB, Oyama S, Goerger BM, Rucinski TJ, Blackburn JT, Creighton RA. Influence of humeral torsion on interpretation of posterior shoulder tightness measures in overhead athletes. *Clin J Sport Med*. 2009;19(5):366-371.

33. Thomas SJ, Sheridan S, Reuther KE. Humeral retroversion and participation age in professional baseball pitchers by geographic region. *J Athl Train*. 2020;55(1):27-31.

34. Sweitzer BA, Thigpen CA, Shanley E, et al. A comparison of glenoid morphology and glenohumeral range of motion between professional baseball pitchers with and without a history of SLAP repair. *Arthroscopy*. 2012;28(9):1206-1213.

35. Bullock GS, Hughes T, Sergeant JC, Callaghan MJ, Riley R, Collins G. Methods matter: clinical prediction models will benefit sports medicine practice, but only if they are properly developed and validated. *Br J Sport Med*. 2021;55(23):1319-1321.

36. Bullock GS, Hughes T, Sergeant JC, Callaghan MJ, Riley RD, Collins GS. Clinical prediction models in sports medicine: A guide for clinicians and researchers. *J Orthop Sport Phys Ther*. 2021;51(10):517-525.

37. Willmott CJ, Ackleson SG, Davis RE, et al. Statistics for the evaluation of model performance. *J Geophys Res*. 1985;90(C5):8995-9005.

38. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. Published online 2001:1189-1232.

39. Liaw A, Wiener M. Classification and regression by randomForest. *R news*. 2002;2(3):18-22.

40. Martelli PL, Fariselli P, Casadio R. An ENSEMBLE machine learning approach for the prediction of all-alpha membrane proteins. *Bioinformatics*. 2003;19(suppl_1):i205-i211.

41. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiol*. 2010;21(1):128.

42. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. *Med Dec Making*. 2015;35(2):162-169.

43. Lancashire LJ, Lemetre C, Ball GR. An introduction to artificial neural networks in bioinformatics—application to complex microarray and mass spectrometry datasets in cancer studies. *Briefings Bioinformatics*. 2009;10(3):315-329.

44. Myers JB, Oyama S, Goerger BM, Rucinski TJ, Blackburn JT, Creighton RA. Influence of humeral torsion on interpretation of posterior shoulder tightness measures in overhead athletes. *Clin J Sport Med*. 2009;19(5):366-371. doi:10.1097/JSM.0b013e3181b544f6

45. Noonan TJ, Thigpen CA, Bailey LB, et al. Humeral torsion as a risk factor for shoulder and elbow injury in professional baseball pitchers. *Am J Sport Med*. 2016;44(9):2214-2219.

46. Byram IR, Bushnell BD, Dugger K, Charron K, Harrell Jr FE, Noonan TJ. Preseason shoulder strength measurements in professional baseball pitchers: identifying players at risk for injury. *Am J Sport Med*. 2010;38(7):1375-1382.

47. Scher S, Anderson K, Weber N, Bajorek J, Rand K, Bey MJ. Associations among hip and shoulder range of motion and shoulder injury in professional baseball players. *J Athl Train*. 2010;45(2):191-197.

48. Bullock GS, Shanley E, Collins GS, et al. Development and internal validation of a humeral torsion prediction model in professional baseball pitchers. *J Shoulder Elbow Surg*. 2021;(12):2832-2838.

49. Hansen K, Montavon G, Biegler F, et al. Assessment and validation of machine learning methods for predicting molecular atomization energies. *J Chem Theory Comp*. 2013;9(8):3404-3419.

50. Chen PH, Fan RE, Lin CJ. A study on SMO-type decomposition methods for support vector machines. *IEEE transactions Neural Networks*. 2006;17(4):893-908.

51. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. *J Clin Epidemiol*. 2016;69:245-247.

52. Steyerberg EW, Harrell FE Jr, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54(8):774-781.

53. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol*. 2014;14(1):137.

54. Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. *BJOG*. 2017;124(3):423-432.

SUPPLEMENTARY MATERIALS

Appendix 1

Download: https://ijspt.scholasticahq.com/article/32380-machine-learning-does-not-improve-humeral-torsion-prediction-compared-to-regression-in-baseball-pitchers/attachment/81315.docx