



Machine Learning to Predict Lower Extremity Musculoskeletal Injury Risk in Student Athletes

Maria Henriquez¹, Jacob Sumner², Mallory Faherty³, Timothy Sell³ and Brinnae Bent^{4*}

¹ Department of Statistics, Duke University, Durham, NC, United States, ² Department of Biology, Duke University, Durham, NC, United States, ³ Michael W. Krzyzewski Human Performance Laboratory (K-Lab), Duke University, Durham, NC, United States, ⁴ Department of Biomedical Engineering, Duke University, Durham, NC, United States

Injury rates in student athletes are high and often unpredictable. Injury risk factors are not agreed upon and often not validated. Here, we present a random-forest machine learning methodology for identifying the most significant injury risk factors and develop a model of lower extremity musculoskeletal injury risk in student athletes with physical performance metrics spanning joint strength measured with force transducers, postural stability measured using a force plate, and flexibility, measured with a goniometer, combined with previous injury metrics and athlete demographics. We tested our model in a population of 122 student athletes with performance metrics for the lower extremity musculoskeletal system and achieved an injury risk accuracy of 79% and identified significant injury risk factors, that could be used to increase accuracy of injury risk assessments, implement timely interventions, and decrease the number of career-ending or chronic injuries among student athletes.

Keywords: injury risk, machine learning, sports science, student athlete, random forest

OPEN ACCESS

Edited by:

Valentina Camomilla,
Università degli Studi di Roma Foro
Italico, Italy

Reviewed by:

Omid Jahanian,
Mayo Clinic, United States
Federico Mari,
Foro Italico University of Rome, Italy

*Correspondence:

Brinnae Bent
brinnae.bent@duke.edu

Specialty section:

This article was submitted to
Sports Science, Technology and
Engineering,
a section of the journal
Frontiers in Sports and Active Living

Received: 26 June 2020

Accepted: 27 October 2020

Published: 19 November 2020

Citation:

Henriquez M, Sumner J, Faherty M,
Sell T and Bent B (2020) Machine
Learning to Predict Lower Extremity
Musculoskeletal Injury Risk in Student
Athletes.
Front. Sports Act. Living 2:576655.
doi: 10.3389/fspor.2020.576655

INTRODUCTION

Musculoskeletal injuries are a significant problem in student athlete populations with some reports indicating that 90% of student athletes report a sports-related injury during their athletic career as a student (Research Report: Changing the Culture of Youth Sports, 2014). Student athletes are 2.5 times more likely to report a major injury and chronic injuries than non-athletes: 67% of former Division I athletes sustain a major injury and 50% reported chronic injuries in a survey of 232 former Division I athletes (Simon and Docherty, 2014; Cowee and Simon, 2019). These injuries will have chronic effects that may influence lifelong physical activity behavior (Hrysomallis, 2007) and have profound financial implications for student athletes including costly medical bills and potential loss of scholarship (Dixon, 2017). The lower extremity musculoskeletal system consists of 42% of all injuries obtained while engaging in sports and recreation activities (Conn et al., 2003). Training interventions can help mitigate lower extremity musculoskeletal injuries (Patel et al., 2017); however, the identification of risk factors has been challenging. There is a critical need to identify risk factors for lower extremity musculoskeletal injuries.

The identification of risk factors of lower extremity musculoskeletal injuries has been challenging because of the lack of consensus on risk factors. Proposed risk factors for lower extremity musculoskeletal injury risk include over-training, inadequate nutrition, previous injury, gender, limb dominance, ankle and knee joint laxity, muscle strength, imbalance, and postural

stability (Neely, 1998; Murphy et al., 2003; Borresen and Lambertt, 2009; Saragiotto et al., 2014; Toohey et al., 2017). Machine learning models have previously been developed to explore risk factors in specific populations (Meeuwisse, 1994; Bahr and Holme, 2003; Novatchkov and Baca, 2013; Tixier et al., 2016; Bunker and Thabtah, 2019; Claudino et al., 2019). Relevant studies using machine learning to assess injury risk in sports include injury risk in elite youth football using gradient boosting algorithms with an accuracy of 85% (Rommers et al., 2020), injury risk in professional soccer players using multiple machine learning methods with an ROC score = 0.747 (López-Valenciano et al., 2018) and injury risk in the NBA using game day statistics and a sliding window with random forests, which achieved an AUC of 0.65 – 0.95 depending on the length of the sliding window (Talukder et al., 2016). However, previous machine learning models have been limited in their inclusion of measurements of muscle strength, imbalance, flexibility, and postural stability. With the inclusion of these physical performance measurements, machine learning models could be utilized to determine significant physical performance risk factors for injury risk, which would be advantageous to modeling and understanding injury risk factors.

Risk factors of lower extremity musculoskeletal injury risk need to be identified in order to implement timely interventions to reduce the prevalence of injury in student athletes. In this study, we will explore the use of machine learning algorithms to identify the most significant risk factors and develop a model of lower extremity musculoskeletal injury risk with physical performance measurements for the lower extremity musculoskeletal system in student athletes.

METHODS

Participants

The study population was composed of 122 college Division I NCAA athletes (51 females, 71 males) across four sports: men's and women's basketball (18.9%), men's football (30.3%), men's and women's soccer (37.7%), and women's volleyball (13.1%) (Table 1, Figure 1). The average age of the study population was 19.56 ± 1.33 years (Table 1). The percentage of student athletes in our data set with any occurrence of a lower extremity injury in the year following their initial testing date was approximately 43.44%. All participants gave written informed consent by a university institutional review board.

Overview of Data Collection

The data contains 50 physical metrics spanning strength, postural stability, and flexibility (dominant and non-dominant leg) combined with previous injury binary classification and demographic data (Supplementary Table 1).

The postural stability assessments are performance metrics that are tested using a force plate (Kistler 9286A, Amherst, NY) at a sampling frequency of 1,200 Hz and quantify balance under both dynamic and static conditions. Postural stability data was collected and processed according to Sell (2012). The strength assessments are performance metrics that are tested using force transducers (seated and handheld) (Lafayette

TABLE 1 | Demographics.

	Number of Participants	Percentage (%)
SPORT		
Women's Basketball	12	9.38%
Men's Basketball	11	9.02%
Women's Soccer	23	18.85%
Men's Soccer	23	18.85%
Women's Volleyball	16	13.11%
Men's Football	37	30.33%
GENDER		
Male	71	58.20%
Female	51	41.80%
AGE		
18–20	66	54.10%
20–22	42	34.40%
22+	14	11.50%

Student athlete primary collegiate sport, self-identified gender, and age range of participants in the dataset.

Instrument Company, Lafayette, IN) and assess joint strength. The flexibility assessments included performance metrics that tested joint range of motion and the muscle-tendon assessed with a standard goniometer or digital inclinometer (Saunders Group, Chaska, MN). Strength and flexibility data were collected and processed according to Faherty et al. (2020). Additionally, binary classification of previous injury in the past 12 months, age, and gender were collected.

Model Development and Evaluation

Prior to model development, features of dominant and non-dominant leg were averaged together and normalized using z-scores, as shown in Equation 1 (Altman, 1968). While normalization is not necessary for random forest modeling due to it being tree-based, we used the normalized features to get units in a comparable form in order to aid in understanding of feature differences for a web-based application for assessing injury risk in student athletes. The final feature set contained 2 categorical variables and 19 normalized numeric variables (Supplementary Table 1). The outcome variable was a binary classification of injury in the year following initial data collection. Statistical descriptions of the dataset can be found in our public repository: <https://github.com/marhenriq/Student-Athlete-Injury-Risk>.

$$\text{normVariable} = \frac{\text{VALUE} - \text{mean}(\text{VALUE})}{\text{std}(\text{VALUE})} \quad (1)$$

In order to identify the most significant risk factors and develop a model of lower extremity musculoskeletal injury risk in student athletes, we required a machine learning model that could handle the breadth of highly-correlated physiological features and distributions in our dataset while also being interpretable to enable our differentiation of the most important risk factors. Random forest models are ideal for our use case because they

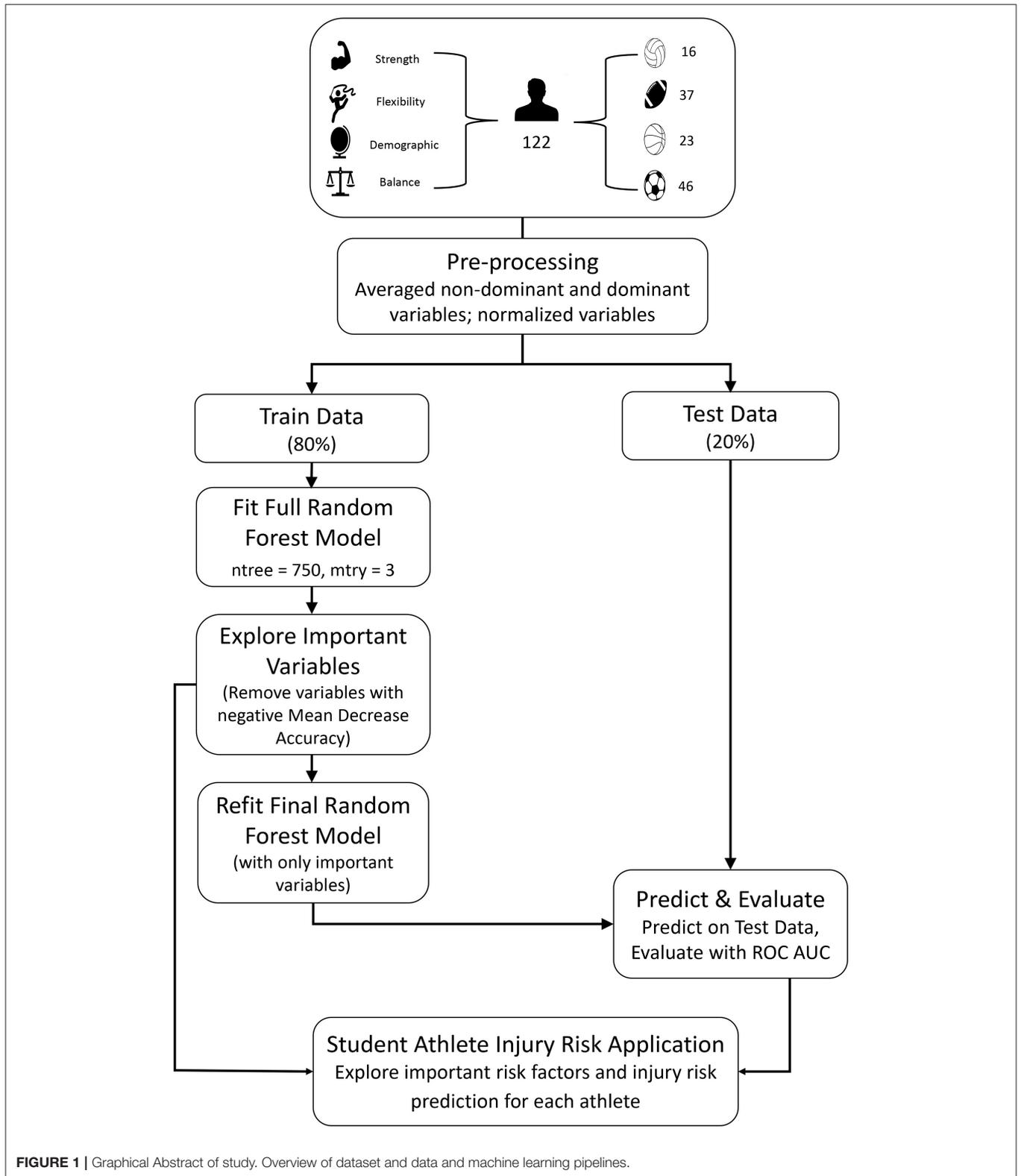
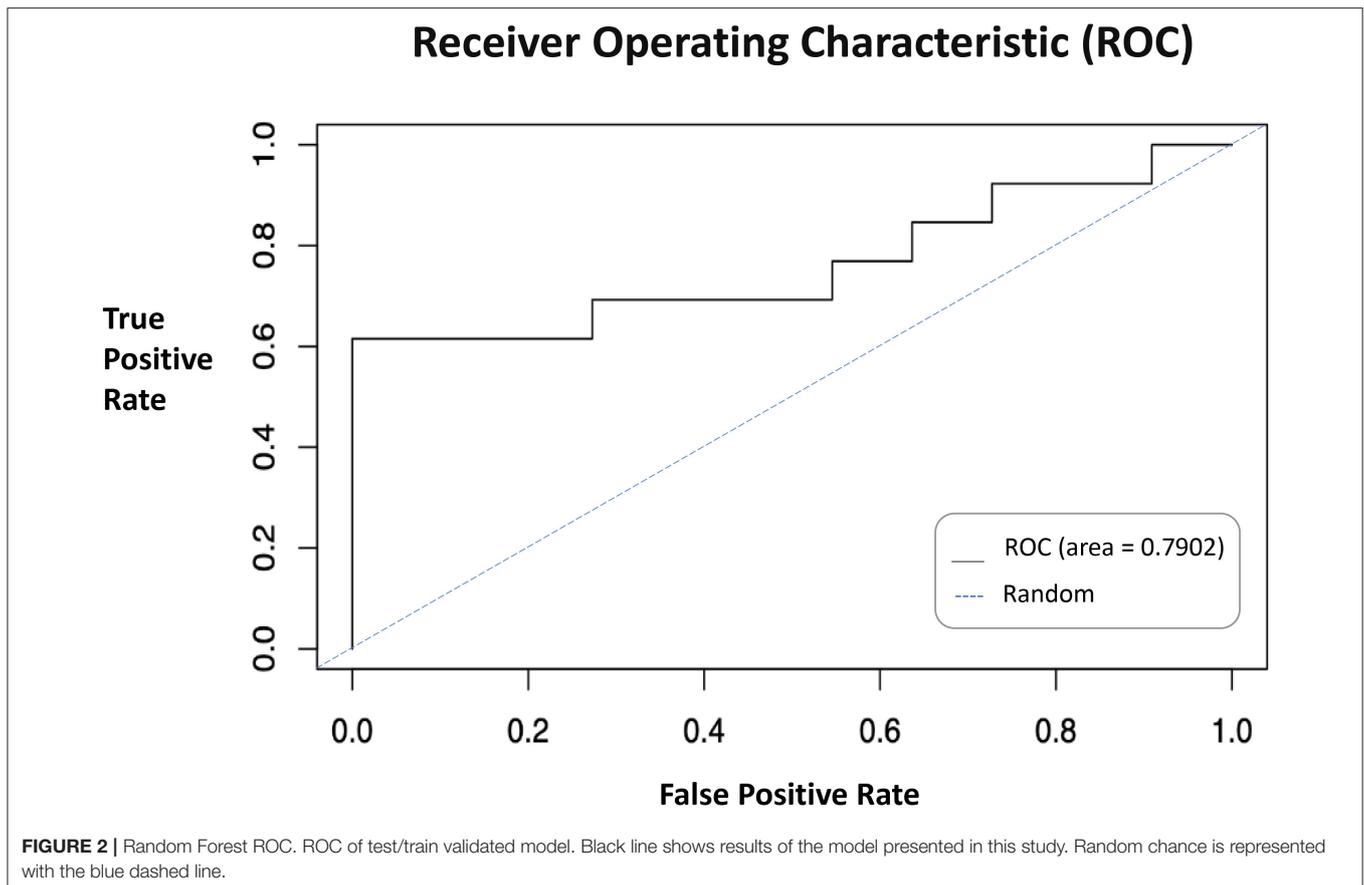


FIGURE 1 | Graphical Abstract of study. Overview of dataset and data and machine learning pipelines.

have no formal distributional assumptions and are therefore useful in handling features with a variety of distributions (Ho, 1998; Kleinberg, 2000; Breiman, 2001). Random forest models

can handle multi-modal data and they are interpretable, allowing us to decipher important relationships between features and outcome variables (Ho, 1998; Kleinberg, 2000; Breiman, 2001).



We developed a random forest (RF) model with the randomForest library in R that is based on Breiman's Random Forests (Breiman, 2001; Liaw, 2018) using a test/train split validation strategy (80:20 split, separating on subjects) and tuned hyperparameters using a GridSearch (**Figure 1**). Our RF model was composed of 750 trees (number of variables available for splitting at each tree node, $mtry = 3$). We tuned the RF model by removing all variables that yielded a negative Mean Decrease Accuracy score. Mean decrease accuracy (also known as permutation importance) is the scaled decrease in accuracy if the variable is removed from the features when developing the model (Han et al., 2016). The variables removed from the final model because they had a negative mean decrease accuracy were eyes closed balance test composite score, DPSI composite score, gender, gastrocnemius flexibility, and knee extension strength.

We set the output of our model to be probabilities of a participant being injured in the following year (Subject Injury Risk). We chose probabilities as the output of our model rather than using a scoring system with no maximum value because probabilities are more interpretable due to being bound from values 0 to 1. We used the confusion matrix to define the class error rates of false negative and false positive errors in our model. The performance metric we utilized to evaluate the performance of our model was receiver operating characteristic (ROC) area under the curve (AUC). We completed a secondary validation of our RF model using k-fold cross validation (folds = 5, $mtry = 2$).

All data analyses and model development were performed in R (version 3.2.3).

Web Application

To facilitate the communication of results with athletes, coaches, and physicians, we developed a web-application through Dash by Plotly¹ that allows the viewer to see how respective individual statistics/testing values compare with that of the individual's team (**Figure 1**). Backend development of the application was coded in Python, while frontend development was coded in HTML and CSS. The application was deployed through Heroku (Salesforce, San Francisco, CA).

RESULTS

We developed a model for injury risk in student athletes using random forest machine learning. With our initial validation strategy of test/train split, our ROC AUC accuracy metric was 79.02% (**Figure 2**). The false negative class error rate was 15.52% and the false positive class error rate was 77.50%. So, errors in our model were more likely to be false positives than false negatives. Our secondary validation with k-fold cross validation resulted in an average ROC AUC of 68.90%. Comprehensive analysis of the

¹*Dash Documentation and User Guide*. Plotly. Available online at: <https://dash.plotly.com/> (accessed October 5, 2020).

TABLE 2 | Random Forest Variable Importance.

Variable	Category	Mean Decrease Accuracy
Hip adductor strength	Strength	5.3465738
Hip external rotation strength	Strength	4.3168850
Straight leg raise	Flexibility	4.1723271
Height	Demographic	3.6593198
Hip abductor strength	Strength	3.6520573
Hip internal rotation strength	Strength	3.4898969
Eyes open balance test composite score	Balance	3.2547060
Ankle dorsiflexion strength	Strength	2.9879077
Ankle plantarflexion strength	Strength	2.8693613
Primary sport type	Demographic	2.7218000
Knee flexion strength	Strength	2.3598642
Eyes closed balance test composite score	Balance	2.1693963
Active knee extension	Flexibility	1.7146625
Ankle eversion strength	Strength	1.6240054
Ankle inversion strength	Strength	0.8680607

Variable Importance represented by Mean Decrease Accuracy for 15 variables used in tuned model.

random forest model is available in a public repository at <https://github.com/marhenriq/Student-Athlete-Injury-Risk>.

We determined the features most important in determining risk of injury in our model using mean decrease accuracy. The most important variables include Hip External Rotation Strength, Hip Adductor Strength, and Straight Leg Raise (Table 2, Supplementary Figure 1, Supplementary Table 2). We examined which feature type (strength, flexibility, balance, and demographics) were most important in determining risk of injury and found that all feature types were important in model development; the relative importance of each feature type were within 0.03 of each other (Figure 3).

The outputs of our model were used in an application for assessing injury risk in student athletes (Supplementary Figure 2). This application allows the comparison of personal injury risk to team injury risk, creating a normalized injury risk to the student athletic team. Additionally, this application enables the comparison of important features of injury risk among student athletes on the same team, which can be used to develop injury risk mitigation training programs for student athletes.

DISCUSSION

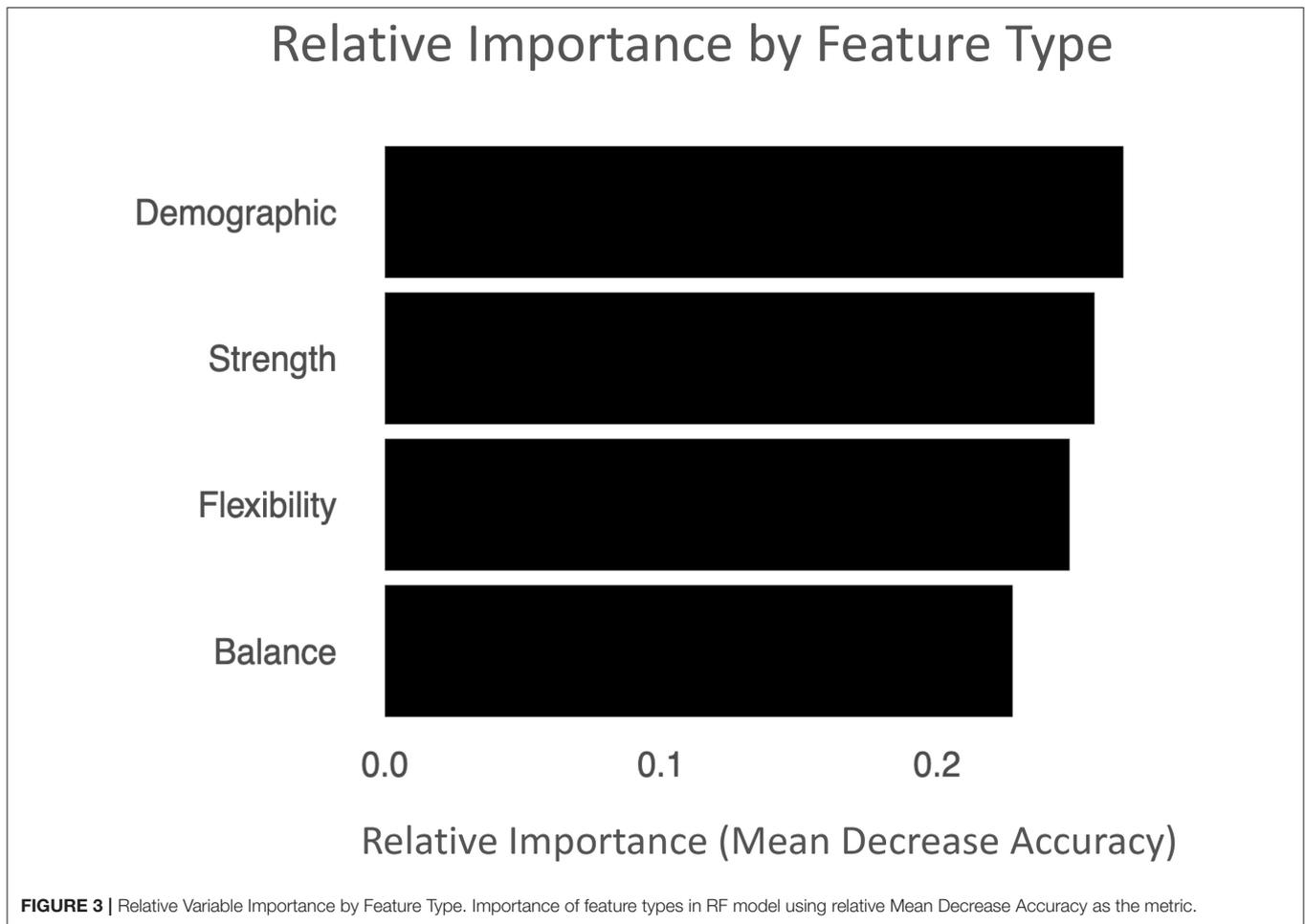
It is critical to identify risk factors of lower extremity musculoskeletal injury risk in order to implement timely interventions to reduce the prevalence of injury in student athletes. In this study, we utilized random forest machine learning algorithms to identify the most significant risk factors and develop a model of lower extremity musculoskeletal injury risk with physical performance measurements for the lower extremity musculoskeletal system in student athletes. Specifically, we developed a random forest machine learning model to explore risk factors of lower extremity musculoskeletal injury in

student athletes. We used these important risk factors to create a model for student athlete injury risk. By combining analysis of identified risk factors and injury risk relative to the entire student athlete team, coaches, sports injury clinicians, and athletes can quantitatively summarize the strengths and weaknesses of an athlete, enabling the implementation of timely interventions to reduce student athlete injury risk.

This model and the web-based application is meant to enable athletes, coaches, and trainers to assess relative injury risk and provide guidance on what physical performance metrics need to be improved to decrease this risk.

On average, the relative importance values of each feature type (strength, flexibility, balance, and demographics) with respect to Mean Decrease Accuracy are quite similar (Figure 3, Table 2). This suggests that a combination of these areas is important to comprehensively assess student athlete injury, as supposed to one or two feature types. Removing one feature type entirely would significantly decrease the predictive power of the RF model. This model can serve as validation to utilize measurements across these feature types in injury risk assessment. The most important variables in our model of injury risk in student athletes were two hip strength-based metrics, Hip Adductor Strength and Hip External Rotation Strength, and a flexibility metric, Straight length Raise. These metrics were more important than ankle strength, knee strength, and balance metrics. This supports a body of literature that found relationships between hip strength and risk for injury (Holcomb et al., 2012; Khayambashi et al., 2016; Stastny et al., 2016; Mucha et al., 2017; Nguyen et al., 2017; Powers et al., 2017) and relationships between flexibility and risk for injury (Gleim and McHugh, 1997; Thacker et al., 2004; Leppänen et al., 2014). Surprisingly, ‘Primary Sport’ was not one of the most important variables in our model of injury risk. This could be due to our limited amount of sports and/or the small sample size in the dataset.

The large number of confounding variables present when recording injury is a significant obstacle in discovering the true accuracy of any injury risk model. A student athlete might have a very high risk for injury but can never get injured due to lack of playtime or successful training/physical therapy sessions and an athlete with a low injury risk could easily get injured through aggressive contact injury or unexpected confounding variables. Thus, it is nearly impossible to predict injury with a model. However, through the development of injury risk models, we can better understand the physiological parameters that have an impact on risk for injury. Accordingly, this study serves to normalize the significant physical performance metrics and demographic variables for the student athletes in order to quantitatively provide an overview of a student athlete’s strengths and weaknesses. The risk score given by our model, normalized to teammates, should serve as a tool, that combined with insights from clinicians and coaches, can aid in the development of an injury-prevention plan to quantitatively lower an athlete’s risk for injury. The use of any machine learning model for injury risk should be accompanied by additional visualizations and summary statistics of the student athlete as well as the student athlete’s team to provide further insight into an athlete’s risk for injury, as we have done in our student athlete injury risk application.



The limitations of this study include the small sample size (122 student athletes) and not having features addressing student athlete nutrition habits, stress factors, and game play statistics. While we have attempted to reduce training time and the potential for overfitting with careful feature selection methods, random forest modeling has inherent limitations, which include high model complexity requiring computational resources and longer training periods than other machine learning frameworks. We use Mean Decrease Accuracy for feature selection, which has been known to have limitations due to the multicollinearity problem (variable impact calculation is less accurate when there are high numbers of correlated variables) (Hur et al., 2017). Future studies with significantly larger sample sizes will be required to improve upon this framework for general student athlete injury risk. Different models for specific sports should be explored to determine specific injury risk triggers. While our model did not find Primary Sport to be a significant variable in injury risk, it should still be explored and validated. Models generated for specific types of lower extremity musculoskeletal injury may enable greater specificity and interpretability of injury risk. In future studies, gradient tree boosting, another tree-based machine learning model, could be considered to reduce the computational resources necessary for random forest modeling. Additionally, new methods of feature selection that solve the

multicollinearity problem, such as the Shapley Value method (Hur et al., 2017), could be considered for future selection in future studies.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of participant confidentiality and privacy restrictions. Requests to access the datasets should be directed to Brinnae Bent, brinnae.bent@duke.edu.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Pittsburgh Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MH, JS, and BB were involved in data analysis and interpretation, model development, and application development. MF and TS were involved in study design and data collection. MH and BB were involved in manuscript preparation. MH,

BB, TS, and MF were involved in manuscript editing. All authors contributed to the article and approved the submitted version.

FUNDING

The Rhodes Information Initiative Data+ program funded the work for data analysis and model development. BB was funded through a Duke Forge fellowship. We would like to acknowledge the Duke University Compact for Open Access Publishing Equity (COPE) Fund for supporting this publication.

REFERENCES

- Altman, I. (1968). *Altman Z-Score*. FCS Commercial Finance Group.
- Bahr, R., and Holme, I. (2003). Risk factors for sports injuries - a methodological approach. *Br. J. Sports Med.* 37, 384–392. doi: 10.1136/bjism.37.5.384
- Borresen, J., and Lambert, M. I. (2009). The quantification of training load, the training response and the effect on performance. *Sport. Med.* 39, 779–795. doi: 10.2165/11317780-000000000-00000
- Breiman, L. (2001). Random forests. *Mach. Lang.* 45, 5–32. doi: 10.1023/A:1010933404324
- Bunker, R. P., and Thabtah, F. (2019). A machine learning framework for sport result prediction. *Appl. Comput. Inform.* 15, 27–33. doi: 10.1016/j.aci.2017.09.005
- Claudino, J. G., de Olivera Capanema, D., de Souza, T. V., Serrão, J. C., Machado Pereira, A. C., Nassis, G. P., et al. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sport. Med. Open* 5:28. doi: 10.1186/s40798-019-0202-3
- Conn, J. M., Annett, J. L., and Gilchrist, J. (2003). Sports and recreation related injury episodes in the US population, 1997–1999. *Inj. Prev.* 9, 117–123. doi: 10.1136/ip.9.2.117
- Cowee, K., and Simon, J. E. (2019). A history of previous severe injury and health-related quality of life among former collegiate athletes. *J. Athl. Train.* 54, 64–69. doi: 10.4085/1062-6050-377-17
- Dixon, C. (2017). *When Student-Athletes Get Injured, Who Pays?* Noodle. Available online at: <https://www.noodle.com/articles/when-student-athletes-get-injured-who-pays134> (accessed June 15, 2020).
- Faherty, M., Varnell, M., Csonka, J., Salehi, K., Gomez, S., and Sell, T. (2020). Sex differences in lower extremity musculoskeletal and neuromuscular characteristics in intercollegiate soccer athletes. *Sci. Med. Footb.* 4, 45–51. doi: 10.1080/24733938.2019.1628351
- Gleim, G. W., and McHugh, M. P. (1997). Flexibility and its effects on sports injury and performance. *Sport. Med.* 24, 289–299. doi: 10.2165/00007256-199724050-00001
- Han, H., Guo, X., and Yu, H. (2016). “Variable selection using mean decrease accuracy and mean decrease gini based on random forest,” in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS* (Beijing: IEEE Computer Society), 219–224.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 832–844. doi: 10.1109/34.709601
- Holcomb, W. R., Miller, M. G., and Rubley, M. D. (2012). Importance of comprehensive hip strengthening. *Strength Cond. J.* 34, 16–19. doi: 10.1519/SSC.0b013e31822ecac9
- Hrysomallis, C. (2007). Relationship between balance ability, training and sports injury risk. *Sport. Med.* 37, 547–556. doi: 10.2165/00007256-200737060-00007
- Hur, J. H., Ihm, S. Y., and Park, Y. H. (2017). A variable impacts measurement in random forest for mobile cloud computing. *Wirel. Commun. Mob. Comput.* 2017:6817627. doi: 10.1155/2017/6817627
- Khayambashi, K., Ghoddosi, N., Straub, R. K., and Powers, C. M. (2016). Hip muscle strength predicts noncontact anterior cruciate ligament injury in male and female athletes: a prospective study. *Am. J. Sports Med.* 44, 355–361. doi: 10.1177/0363546515616237
- Kleinberg, E. M. (2000). On the algorithmic implementation of stochastic discrimination. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 473–490. doi: 10.1109/34.857004
- Leppänen, M., Aaltonen, S., Parkkari, J., Heinonen, A., and Kujala, U. M. (2014). Interventions to prevent sports related injuries: a systematic review and meta-analysis of randomised controlled trials. *Sport. Med.* 44, 473–486. doi: 10.1007/s40279-013-0136-8
- Liaw, A. (2018). *randomForest package in R: Breiman and Cutler's Random Forests for Classification and Regression*. Available online at: <https://www.stat.berkeley.edu/~breiman/RandomForests/> (accessed September 29, 2020).
- López-Valenciano, A., Ayala, F., Puerta, J. M., de Ste Croix, M. B. A., Vera-García, F. J., Vera-García, F. J., et al. (2018). A preventive model for muscle injuries: a novel approach based on learning algorithms. *Med. Sci. Sports Exerc.* 50, 915–927. doi: 10.1249/MSS.00000000000001535
- Meeuwisse, W. H. (1994). Assessing causation in sport injury: a multifactorial model. *Clin. J. Sport Med.* 4, 166–170. doi: 10.1097/00042752-199407000-00004
- Mucha, M. D., Caldwell, W., Schlueter, E. L., Walters, C., and Hassen, A. (2017). Hip abductor strength and lower extremity running related injury in distance runners: a systematic review. *J. Sci. Med. Sport* 20, 349–355. doi: 10.1016/j.jsams.2016.09.002
- Murphy, D. F., Connolly, D. A. J., and Beynon, B. D. (2003). Risk factors for lower extremity injury: a review of the literature. *Br. J. Sports Med.* 37, 13–29. doi: 10.1136/bjism.37.1.13
- Neely, F. G. (1998). Intrinsic risk factors for exercise-related lower limb injuries. *Sport. Med.* 26, 253–263. doi: 10.2165/00007256-199826040-00004
- Nguyen, A. D., Zuk, E. F., Baellow, A. L., Pfile, K. R., DiStefano, L. J., and Boling, M. C. (2017). Longitudinal changes in hip strength and range of motion in female youth soccer players: Implications for ACL injury, a pilot study. *J. Sport Rehabil.* 26, 358–364. doi: 10.1123/jsr.2015-0197
- Novatchkov, H., and Baca, A. (2013). Artificial intelligence in sports on the example of weight training. *J. Sport. Sci. Med.* 12, 27–37.
- Patel, D. R., Yamasaki, A., and Brown, K. (2017). Epidemiology of sports-related musculoskeletal injuries in young athletes in United States. *Transl. Pediatr.* 6, 160–166. doi: 10.21037/tp.2017.04.08
- Powers, C. M., Ghoddosi, N., Straub, R. K., and Khayambashi, K. (2017). Hip strength as a predictor of ankle sprains in male soccer players: a prospective study. *J. Athl. Train.* 52, 1048–1055. doi: 10.4085/1062-6050-52.11.18
- Research Report: Changing the Culture of Youth Sports (2014). Safe Kids Worldwide. Available online at: <https://www.safekids.org/research-report/research-report-changing-culture-youth-sports-august-2014> (accessed June 15, 2020).
- Rommers, N., Rössler, R., Verhagen, E., Vandecasteele, F., Verstockt, S., Vaeyens, R., et al. (2020). A machine learning approach to assess injury risk in elite youth football players. *Med. Sci. Sports Exerc.* 52, 1745–1751. doi: 10.1249/MSS.00000000000002305
- Saragiotto, B. T., Di Pierro, C., and Lopes, A. D. (2014). Fatores de risco e prevenção de lesões em atletas de elite: estudo descritivo da opinião de

ACKNOWLEDGMENTS

The authors would like to acknowledge Paul Bendich and the entire Data+ team for bringing this team together and their guidance in presenting this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fspor.2020.576655/full#supplementary-material>

- fisioterapeutas, médicos e treinadores. *Brazilian J. Phys. Ther.* 18, 137–143. doi: 10.1590/S1413-35552012005000147
- Sell, T. C. (2012). An examination, correlation, and comparison of static and dynamic measures of postural stability in healthy, physically active adults. *Phys. Ther. Sport.* 13, 80–86. doi: 10.1016/j.ptsp.2011.06.006
- Simon, J. E., and Docherty, C. L. (2014). Current health-related quality of life is lower in former division I collegiate athletes than in non-collegiate athletes. *Am. J. Sports Med.* 42, 423–429. doi: 10.1177/0363546513510393
- Stastny, P., Tufano, J. J., Golas, A., and Petr, M. (2016). Strengthening the gluteus medius using various bodyweight and resistance exercises. *Strength Cond. J.* 38, 91–101. doi: 10.1519/SSC.0000000000000221
- Talukder, H., Vincent, T., Foster, G., Hu, C., Huerta, J., Kumar, A., et al. (2016). “Preventing in-Game Injuries for NBA Player,” in *MIT Sloan Sports Analytics Conference* (Boston, MA).
- Thacker, S. B., Gilchrist, J., Stroup, D. F., and Kimsey, C. D. (2004). The impact of stretching on sports injury risk: a systematic review of the literature. *Med. Sci. Sports Exerc.* 36, 371–378. doi: 10.1249/01.MSS.0000117134.83018.F7
- Tixier, J.-P. A., Hallowell, M., Rajagopalan, B., and Bowman, D. (2016). Application of machine learning to construction injury prediction. *Autom. Constr.* 69, 102–114. doi: 10.1016/j.autcon.2016.05.016
- Toohey, L. A., Drew, M. K., Cook, J. L., Finch, C. F., and Gaida, J. E. (2017). Is subsequent lower limb injury associated with previous injury? A systematic review and meta-analysis. *Br. J. Sports Med.* 51, 1670–1678. doi: 10.1136/bjsports-2017-097500

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Henriquez, Sumner, Faherty, Sell and Bent. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.