Predicting Lower Body Soft Tissue Injuries in American Football with GPS Data

by

Nicholas Tice

Bachelor of Science
University of South Carolina, 2020

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in

Statistics

College of Arts and Sciences

University of South Carolina

2022

Accepted by:

Joshua Tebbs, Director of Thesis

Minusk Shin, Reader

Lianming Wang, Reader

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

Abstract

It is of utmost importance to sports organizations that they keep their players as healthy as possible and contributing to the success of the team. Advancements in technology and investments by sports clubs have allowed researchers to better understand the role of load management in high-level athletes to mitigate injury risk. Through GPS tracking data provided by a collaborating Division I American college football team, we seek to predict lower body soft tissue injuries in future training sessions and reduce the number of potentially avoidable injuries within the organization. The difficulty of analyzing the injury data set is that the frequency of the injuries is low compared to the non-injury cases, resulting in highly imbalanced classes. We address this imbalance issue through under or oversampling the underrepresented class by bagging, and we compare various classification procedures such as random forests, Gradient Boosting Machines, penalized logistic regression and support vector machines. Our empirical results show that the random forest classifier with undersampling outperforms other methods with respect to precision-recall (PR) and ROC curves.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Reducing injuries that result in a loss of playing time is a key goal of sports organizations at all levels of competition. This is especially true at the collegiate and professional level where millions of dollars are on the line and injuries can cost team performance [7, 23]. As teams are beginning to invest more into nutrition, sleep, and training programs to minimize injury risk, the field of sports science is becoming increasingly popular. The study of sport and exercise has existed for many years, but the curriculum taught at universities for these programs is undergoing a change as careers become more data driven [8, 16]. This is especially true in sports science as degrees have needed to incorporate more biostatistics and traditional statistics education to prepare their students for the data that will be available to them in this growing field [4].

One of the more widely implemented technological advancements is wearable GPS monitoring systems for athletes to use during practices and games that monitor the athlete's workload over time [20]. Many teams employ private companies that perform this service like Catapult and Zebra Technologies, but they often lack the ability to make the most of the opportunities provided by this data. This is especially true in American college football. Teams are left relying on whatever general analysis the company they subscribe to for the the tracking technology provides them, rather than formulating models that apply to their own team. At a cost of nearly $100,000 per year for these tracking data services, it is of great interest to these teams to maximize their use of this data [6].

This paper aims to expand upon the training load research of rugby and soccer leagues and apply it to American football. [2, 18] Previous studies in American football have researched the association between training load and soft tissue injury but did not attempt to predict or model the injuries [12]. To our knowledge, this is the first study with the goal of predicting injuries from GPS tracking data in American football. The most significant challenge facing anyone trying to model sports injuries is the class imbalance issue commonly seen in this type of data. Several techniques and methods were tested to combat this issue by drawing on existing research in areas with similar problems such as fraud identification scenarios. Models will be compared using area under the receiver operator characteristic (AUC), precision, and recall. By using different cutpoints for injury classification, risk for players will be divided into three categories (No risk, low risk, high risk) that the team could use to reduce the number of soft tissue injuries among the players.

Chapter 2: Material and Methods

2.1 Data Collection

The data in this observational study was received from a collaborating Division 1 American college football team. Measurements were taken from March 2016 to December 2020. In total, 23,058 training sessions from 150 unique athletes were recorded. Sessions included both practices and games. 347 non-contact lower body injuries occurred over this time. In order to be eligible for inclusion in the study, a player was required to have participated in at least 50 sessions. The team partnered with Catapult, one of the premier companies in the wearable technology space, to collect the data, and it was managed by a full-time staff member of the football team. Each player being tracked wore a vest containing a 10Hz GPS device, 100 Hz accelerometer, and heart rate monitor, all of which were measured through the Catapult ClearSky system [22]. Athletes were required to wear the device for the duration of their workout. Sessions before 2019 in the indoor facility were not recorded since the ClearSky system was not installed indoors until that point. Due to a limited supply of Catapult systems available (75 vests for approximately 115 athletes), the study is mostly made up of players on full athletic scholarships and excluded the positions of punter, kicker, or long snapper.

2.2 Variable Creation

The system records the following six variables every session and is easily accessible as a downloadable csv file: Total Player Load, Maximum Velocity, Hard Acceleration Efforts, Hard Deceleration Efforts, High Speed Distance, and Total Distance. Player Load is calculated according to the following formula where ax_i, ay$_i$, and az$_i$ are the acceleration values in the x, y, and z directions:

$$PlayerLoad = \sum_{i=1}^{n} \sqrt{(ax_i - ax_{i-1})^2 + (ay_i - ay_{i-1})^2 + (az_i - az_{i-1})^2}$$

The formula was developed by the Australian Institute of Sport to measure effort in rugby but has since been adopted by other sports [24]. Maximum velocity is the highest velocity reached by a player in the session in miles per hour. Hard Acceleration and Hard Deceleration are the total number of times that an athlete's acceleration or deceleration was greater than a set threshold (3 meters/sec^2) for at least 1 second. Both are a discrete quantitative variable. The two distance metrics give the total distance covered in yards as a continuous variable. High Speed Distance is distance covered over 12 miles per hour. A preprocessing obstacle we had to handle was missing data in the tracking set. About 5,000 observations from 2016 in the tracking data were missing all values except Player Load and were removed. This was consistent for every player in the date that was affected, and it was unclear after talking to the team why this was the case.

In addition to the six daily variables, more features were included in the final modeling data. Three fixed variables were introduced: height, weight, and position (DB, DL, LB, OL, QB, RB, TE, WR). These were taken from the yearly rosters and updated every new year for each player. Indicators were then created for prior injury, including hamstring strains, ankle sprains, back injuries, and total number of lower body injuries along with a variable for whether or not the next session was a game. We also included an indicator for time since last injury by creating breaks between no injury, 0, 21, 60, 120, 300, and 1000. Our last time related variable measured how many days until the next training session.

Another area of interest was incorporating changes in the athlete's daily workload into the model. This was accomplished by using acute:chronic workload ratios (ACWR), which has been shown to be associated with injury rate in athletes [1]. ACWR is traditionally found by averaging the last week of data and and dividing by the average of the last 4 weeks of data. However, a recent study found that calculating the ACWR with an Exponentially Weighted Moving Average (EWMA) can be more sensitive to injury detection [15]. This was implemented in the model and calculated with the following formula where λ=2/(N+1)

$$EWMA_n = \lambda(Value_n) + (1 - \lambda)EWMA_{n-1}$$

N is 7 for the one week period or 28 for the four week period. Instead of a ratio of the straight average of the last week divided by the average of the last four weeks, the moving average gives more weight to the most recent observations. Another ratio commonly used in training load studies with athletes is Foster's Training Monotony, calculated by finding the mean of each variable for that week divided by the standard deviation [3]. This gives a measure of the variability in the weekly training load and was included for each of the six features as well. The first week of data was removed to allow ratios to calculate in the training and test sets.

The last variable created is based on the estimated workload in the next session. Since each position group does roughly the same workouts at each training session, we took the median of each variable in the next session by position group. While these would be unknown, we could model them over the course of a season and come up with reasonable estimates in conjunction with the training staff. They are intended to indicate if the next session will be a step up or step down in intensity. This future variable will then be calculated by taking the value for today's session and dividing by the median of the value in position group in the next session. A value greater than one would indicate that next session is expected to be less intense than the last session and vice versa.

Football is a contact heavy sport where athletes can sustain a wide range of injuries [19]. Since it would be difficult to account for any injuries sustained from collisions with other players, it is of greater interest to focus on non-contact injuries of the lower body. After merging the tracking data with the injury data, the response variable indicating injury in the next session was created by shifting the indicators of injury to the prior session. The last session for each player was removed since it is unclear whether injury would have occurred in the following unknown session. The response variable is binary (1 = Injury or 0 = No Injury) indicating the status of the player. Since the data only provided injury descriptions, we had to manually go through and classify each one as we saw fit. The rule used was that any injury to the legs that was described as a strain or sprain would be classified as a positive case. Another area of interest was the indicators described above for hamstring, ankle sprains, and back injuries. These also needed to be manually indicated in the dataset. Other studies have looked at creating models for specific injuries like hamstrings or ankle sprains, but we did not see any improvement performance when attempting that with our data [2].

The injury dataset required significantly more cleaning than the tracking data before it was usable. The given excel file listed the player ID, a short injury description, the date of the injury, and time missed from injury. All injuries were inputted into an excel file by the athletic training staff. This lead to multiple typos in the names, injury descriptions, and other potentially undetected data entry errors in the dates. Some injuries had no date at all and the time missed variable was left blank for the majority of cases which made that variable unusable. Injuries were matched with the tracking data by joining the two tables by player ID and date. Since the names did not always match in the injury and tracking files (i.e. one lists first name as Nick and the other as Nicholas), some names had to be manually changed before joining the two dataframes. Additionally, not all injuries had dates that matched with the tracking. In those cases, the injury would be assigned to the the prior session if it was within a day of the injury.

2.3 Statistical Analysis

Severe class imbalance is generally defined by a majority-to-minority class ratio somewhere between 100:1 and 10,000:1. Traditional classification models will run into issues when training the data and commonly used metrics like accuracy are not reflective of the true model performance because a bias can be introduced towards the majority class. Machine learning techniques generally outperform classical statistical methods at classification but can still fail to effectively distinguish between classes in these severely imbalanced settings. The two approaches to mitigating this problem are using data-level methods or algorithm-level methods. Some examples of data-level methods include data-sampling and feature selection whereas algorithm-level methods involve cost-sensitive methods and hybrid/ensemble algorithms [11].

We first looked at sampling approaches. Literature largely supports random undersampling (RUS), random oversampling (ROS), and Synthetic Minority Over-sampling Technique (SMOTE) as the most popular and effective methods. RUS randomly samples the majority case down to a predetermined ratio with the minority class, leading to a much lower computational burden by reducing the training set size. ROS and SMOTE oversample until the minority class matches the majority class with SMOTE generating synthetic instances from a random sample of the minority class from its k nearest neighbors. An advantage of this method is that it does not suffer from any loss of data, but it is much slower computationally. Other potential sampling methods include random oversampling examples (ROSE) and variants of SMOTE. ROSE also is a synthetic oversampling method but uses a smoothed bootstrap approach. In our study, we found that undersampling was much more efficient (training set of about 600 vs. 37,000 from oversampling in our example), and lead to similar, if not better, results than SMOTE or any other oversampling method.

In addition to sampling techniques, class imbalance can be addressed from an algorithmic perspective with ensemble learner methods. Common methods used include AdaBoost, Random Forest (RF), Gradient Boosting Machines (GBM), and Support Vector Machines (SVM) [9]. These rely on boosting or bagging to reduce the affect of imbalance. Bagging methods produce several randomly drawn with replacement training sets from the entire dataset. Each set has a classifier trained on it with the results combined in the final classification. Boosting also uses multiple training sets, but samples from the elements that were previously misclassified. It then iteratively assigns weights to each classifier based on their performance and weighs the final classification model accordingly. Bagging seeks to reduce variance, whereas boosting reduces bias. This causes bagging to be better at preventing overfitting.

Sampling and ensemble methods can be combined and often produce the best results. Examples of this include SMOTEBagging, ROSBagging, RUSBagging, RUSBoost, and ROSBoost [11]. These implement the desired sampling method and allow one to choose the classification method from algorithms like random forest, support vector machines, and naive bayes. This combines the benefits of both the algorithmic and data sampling approaches. It should be noted that any of these ensemble sampling boosting or bagging methods create multiple training sets and classifiers, usually 10 or more. The computational complexity for training an oversampled boosting or bagging method can be extreme if the initial training set has a high amount of observations or the number of learners is set to a large number. This is less of an issue with any of the RUS methods.

Since the model output is a continuous probability, another area of interest is picking the best cutpoint for the predictions. While this value is generally chosen to be 0.5 by default, other values can maximize or minimize measurements of interest like specificity, sensitivity, and precision. The *OptimalCutpoints* R package was originally created for clinical applications but can be applied to this problem to categorize the predicted probabilities for the test set into no risk, low risk, and high risk by selecting different optimal cutpoints based on a chosen criteria [13]. The two cutoffs selected were "MinPvalue" and "MaxKappa." The lower cutpoint is chosen with the MinPValue by minimizing the p-value associated with the χ^2 test that measures the association between the predictors and the binary result. This leads to a high negative predicted value rate where a predicted "no injury" is not injured over 99% of the time. The MaxKappa cutoff maximizes the Kappa index, a commonly used measure of inter-rater reliability. The Kappa index provides a stricter cutoff than MinPValue where risk of injury is anywhere from 10 to 30 times more likely than in the "no risk" classification.

In the results section, we will compare different sampling and algorithmic methods to determine the best model. The entire data will be split in an 80/20 split to create a training and test data set. Models will be trained on the same set of data and evaluated based on AUC, precision, recall, and F1-Score calculated from a test set in a simulation study. We will consider models combining sampling and algorithmic approaches like RUSBoost and RUSBagging as well as more traditional methods like logistic regression with RUS.

Chapter 3: Results

3.1 Model Comparison

The first area of interest was determining the best classification choice among the ensemble methods. The R package Ensemble-Based Methods for Class Imbalance Problem “ebmc”) implements four ensemble methods: SMOTEBoost, RUSBoost, UnderBagging, and SMOTEBagging [10]. Each of these methods allows the user to choose the method for classification from“cart” (Classification and Regression Tree), “c50” (C5.0 Decision Tree), “rf” (Random Forest), “nb” (Naive Bayes), and “svm” (Support Vector Machine). Training and test data was generated in an 80/20 split with each method compared on the same datasets and simulated 15 times total. For computation reasons, the five methods were trained using UnderBagging. All default parameters were used for comparison including a 1:1 minority to majority class ratio, 500 trees for the random forest, and a radial kernel for the support vector machine. 15 weak learners were trained for each model. The boxplot comparison of the AUC in Figure 3.1 and metric comparison in Table 3.1 show that the random forest algorithm is the clear best option.

The next area of interest was in determining if there was a difference between the UnderBagging and RUSBoost methods. Both are undersampling methods but differ in the bagging or boosting approach. As mentioned earlier, bagging can reduce overfitting. After doing a similar 15 run simulation comparing the two methods with the random forest classifier, we see that UnderBagging consistently outperforms RUSBoost (Figure 3.2). This leads us to believe there is some level of overfitting in the boosting method, although the difference is marginal and either algorithm would give similar results.

The last idea we want to investigate is if the ensemble method is better than traditional methods for imbalanced classification. We again simulate 15 training sets and use the same training sets for each of the models. The models considered were the UnderBagging with a random forest classifier, a standard random forest model, XGBoost, penalized logistic regression, and a decision tree. Every model except the UnderBagging was trained on a 5:1 ratio undersampled training set to reduce the data loss in the modeling process. The random forest model is tuned for mtry, the number of variables trained at each split, and the number of trees used. XGBoost is also tuned through a grid search for eta $\in (0.01, 0.05, 0.1)$, max_depth $\in (1, 3, 5, 7)$, min_child_weight $\in (3, 5, 7)$, subsample $\in (0.65, 0.8)$, and colsample_bytree $\in (0.8, 0.9, 1)$. Each XGBoost model was tuned with 5-fold cross validation on 30 rounds with early stopping after 10 rounds. Lastly, the logistic regression model was a penalized Lasso logistic regression in glmnet using the L1 norm. One of the advantages to the Lasso penalty is the automatic feature selection that would tell us which variables are contributing to injuries. The results are summarized in Figure 3.3 and Table 3.2. UnderBagging has the best AUCROC, PRROC, and Recall among the three methods. The precision and F1 score are also comparable to other methods. Based on these measures, we will use UnderBagging as our final model choice. One possible reason UnderBagging may have outperformed the other models is the lack of data loss. 10 random forest models are grown on 10 uniquely randomly undersampled datasets, so more of the majority class is being accounted for than in the other methods. It also combines both an algorithmic (RF) with a sampling approach (RUS with bagging) that could have lead to its best performance among the tested models.

3.2 Final Model Evaluation

After tuning the number of weak learners (10), undersampling ratio(1:1), and number of trees to grow in each random forest(1000), we simulate 50 training sets to create our evaluation metrics. We first look to Figure 3.4 i. to see that 50% of the simulated ROC falls between 0.713 and 0.751 with a median of 0.738 and reaching as high as 0.81. The model certainly shows some level of predictive power. Graphs ii., iii., and iv. show the precision, recall, and F1-Score by chosen cutpoint. We see that the cutpoint can have a fairly large impact on the different metrics and should be carefully chosen depending on what we would want to minimize or maximize. In all models, we were unable to obtain a high enough level of precision that would be desired in a practical setting. Precision, also known as the Positive Predictive Value, in our context is the probability of injury given a predicted injury. Most models had a very high number of false positives that lowered the precision and would make it difficult to advise a team to sit out every player that is predicted to have an injury in the next session.

Introducing the two cutpoints does allow for much higher precision in the MaxKappa case, and probabilities higher than that cutpoint indicating injury should warrant some action on the team's part. Recall, or sensitivity, is the probability of predicted injury given the person was actually injured. This tells us how many actual injuries are being classified as noninjuries. We can see that the MinPValue method does quite well in correctly identifying most injuries (recall Q1=0.53, Q3=0.75). However, it does mistakenly classify a lot of noninjuries as injuries which causes the precision to be so low. This precision/recall tradeoff will always occur and it is up to the practitioner to determine which measure is more valuable to maximize. Lastly, we will calculate the F1-Score by taking the harmonic mean of recall and precision:

$$F1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

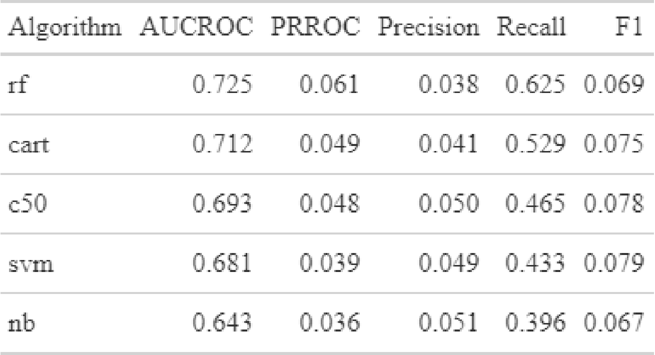
F1-Score is often used to choose between classifiers when one has a higher precision and the other has higher recall or vice versa. Based on the this metric, MaxKappa is the better cutpoint to use.

Since both cutpoints have drawbacks in the model (MinPValue low precision, MaxKappa low Recall), one way to possibly address this is combining them to create multiple prediction classes. By using both cutpoints at once, we can simultaneously address the negatives in each method. To set up the problem, our UnderBagging random forest method generates probabilities generally between 0.1 and 0.9 for the test set. The MinPValue method (average cutoff of 0.52) has high recall, meaning that it captures most of the injuries, so that will establish the first cutpoint. Any predicted probability less than that is very unlikely to be an injury and will be classified at "No Risk". The values between the MinPValue and MaxKappa (average cutoff of 0.72) are at higher risk of injury than the prior group, but still suffer from low precision and produce an undesirable amount of false positives. These values will be considered Low Risk. Any value greater than the MaxKappa will be considered High Risk and capitalize on the higher precision.

We see the effects of this formulation in Table 3.3. In the test set of 4,600 observations, there are approximately 3,200 No Risk (70%), 1,270 Low Risk (27%), and 130 High Risk (3%) predicted values on average. The table shows the probability of injury or no injury given the specified classification. An athlete predicted to have No Risk in the next session does appear to be at very little risk of injury at less than 1%. The Low Risk category also has a fairly low injury rate at 2.6% but is still 3.7 times more likely to be injured than the No Risk class. Finally, the High Risk class has a much higher chance of being hurt than the other two at 11.7% (4.5 times more likely than Low Risk and 16.7 times more likely than no risk). We have created effective groupings of the classifiers to take advantage of the varying precision and recall at different cutpoints.

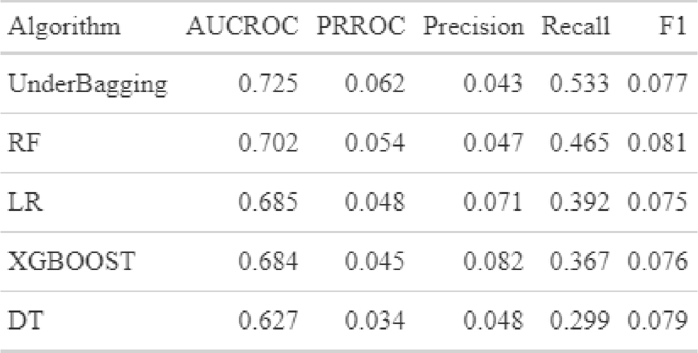
While random forests lack the interpretability of generalized linear models, we can obtain some information about which variables are contributing to the model by looking at variable importance. The two most common methods for this calculation are Gini Importance and Permutation Importance, although both can be biased [21]. The Gini Importance measures the mean decrease impurity by averaging the added weighted impurity decreases for all nodes over all trees. This is less computationally intensive than Permutation Importance but has been shown to be more biased. Permutation Importance is measured by how much the model accuracy decreases when removing the variable. Since it runs on permutations, its slower than the Gini method but generally a better indicator since it is less biased. We show the results for the permutation test in Table 3.4. These were calculated by averaging the Permutation Importance for all random forests grown in the 50 simulations. We see that the most important variable for accuracy is the amount of days until the next practice session. Measurements involving Total Distance and Total Load seem to be the most important among the tracking variables.

Table 3.1 Model comparison with UnderBagging algorithms



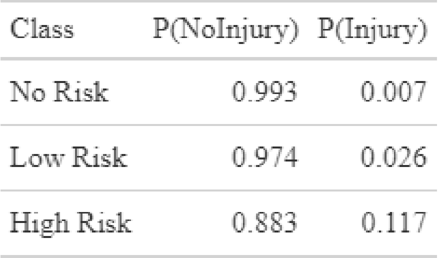
Algorithm	AUCROC	PRROC	Precision	Recall	F1
rf	0.725	0.061	0.038	0.625	0.069
cart	0.712	0.049	0.041	0.529	0.075
c50	0.693	0.048	0.050	0.465	0.078
svm	0.681	0.039	0.049	0.433	0.079
nb	0.643	0.036	0.051	0.396	0.067

Table 3.2 Model comparison for all algorithms



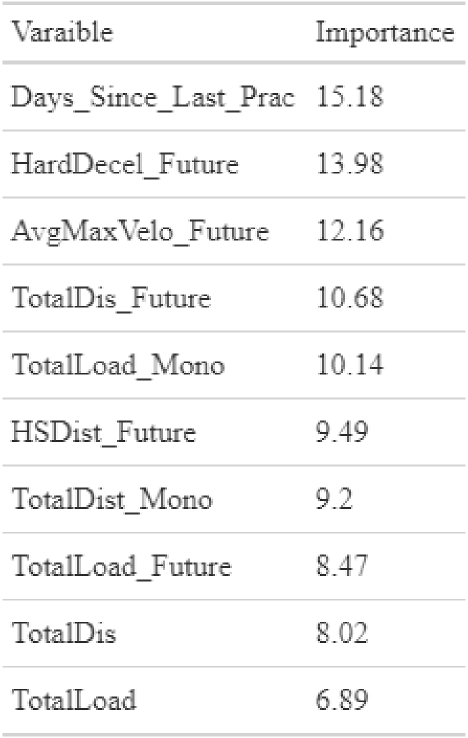
Algorithm	AUCROC	PRROC	Precision	Recall	F1
UnderBagging	0.725	0.062	0.043	0.533	0.077
RF	0.702	0.054	0.047	0.465	0.081
LR	0.685	0.048	0.071	0.392	0.075
XGBOOST	0.684	0.045	0.082	0.367	0.076
DT	0.627	0.034	0.048	0.299	0.079

Table 3.3 Injury risk table



Class	P(NoInjury)	P(Injury)
No Risk	0.993	0.007
Low Risk	0.974	0.026
High Risk	0.883	0.117

Table 3.4 Random forest variable importance



Varaible	Importance
Days_Since_Last_Prac	15.18
HardDecel_Future	13.98
AvgMaxVelo_Future	12.16
TotalDis_Future	10.68
TotalLoad_Mono	10.14
HSDist_Future	9.49
TotalDist_Mono	9.2
TotalLoad_Future	8.47
TotalDis	8.02
TotalLoad	6.89

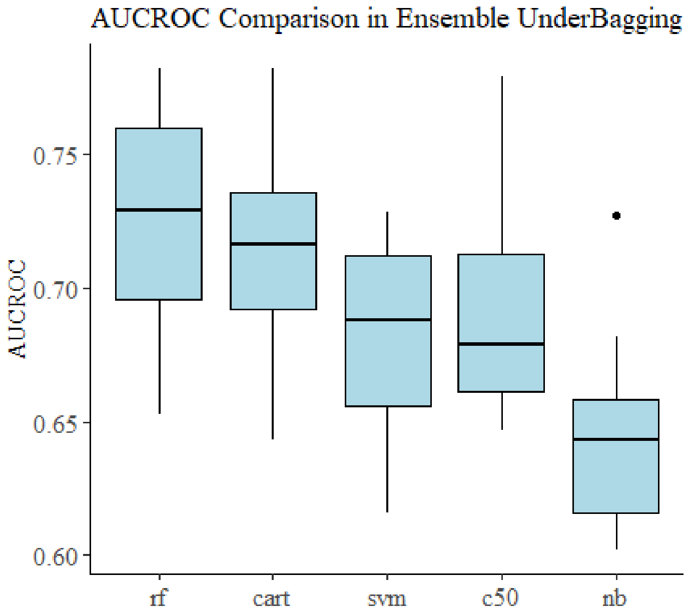


Figure 3.1 Boxplot comparison of AUCROC in UnderBagging algorithms

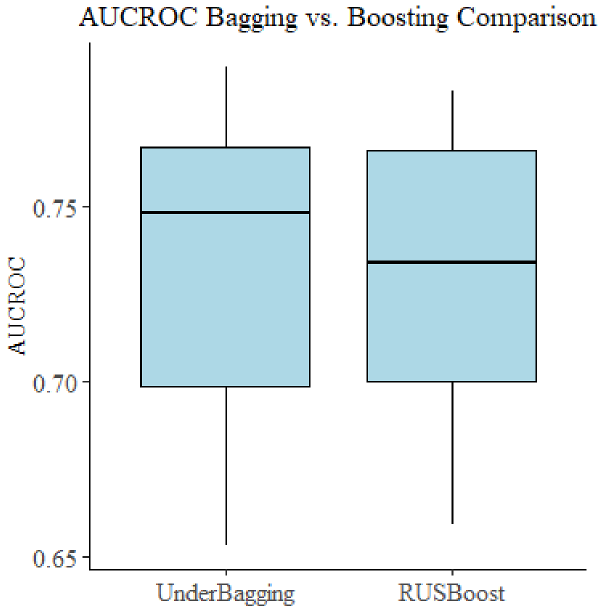


Figure 3.2 AUCROC comparison of UnderBagging vs. Boosting

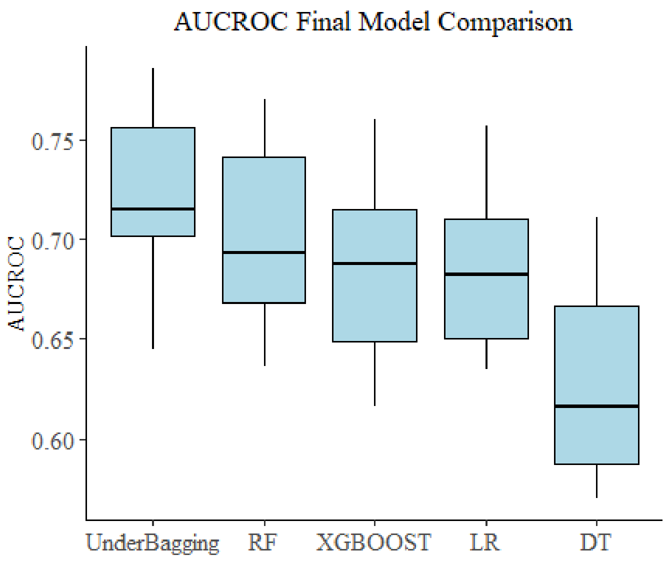


Figure 3.3 AUCROC comparison across all algorithms

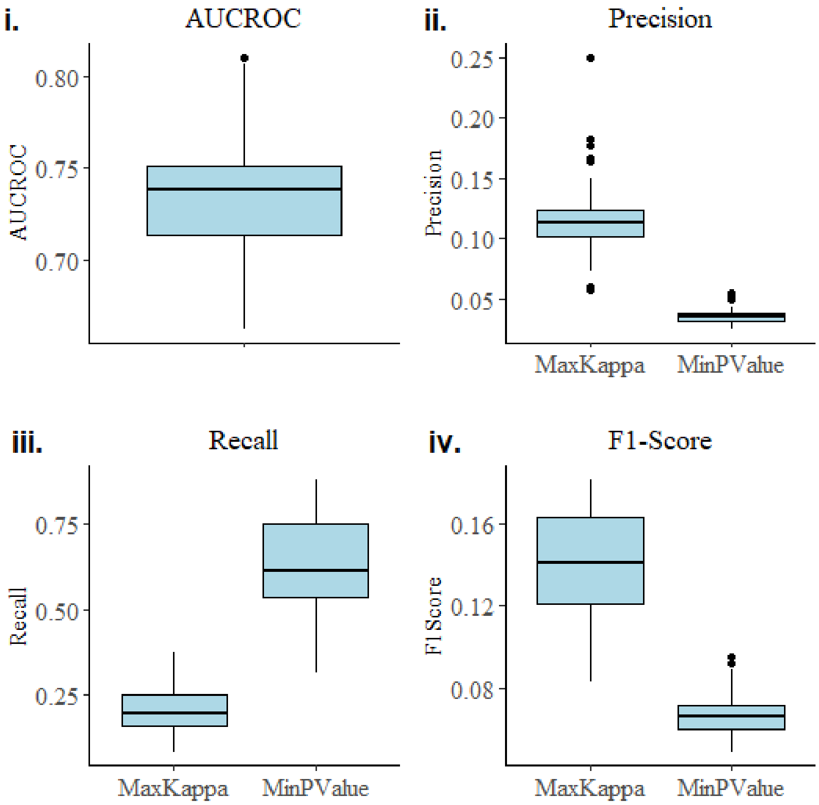


Figure 3.4 Boxplot comparison of cutpoint metrics in final model

Chapter 4: Discussion

4.1 Team Implementation

We have completed extensive testing of the best sampling and algorithmic approaches to address the class imbalance problem. Simulations show that the UnderBagging random forest approach performs better than any other combination of sampling and model choice. The two cutpoint approach to create levels of injury risk helps address some of the issues brought upon by low precision in every model tested. This leads to a more applicable product for the team implementing this method. From an athletic training standpoint, this could better direct which athletes will get the focus of the staff in terms of more treatment before and after training, as well as an awareness by the player, coaches and trainers that the athlete is at a higher risk of injury than their teammates. The players assigned a High Risk tag should almost definitely be limited in training or undergo additional testing before they start training to ensure that they remain healthy. The No Risk group causes a majority of the athletes (approx. 70%) to continue their normal training schedules and allows the training staff to target a smaller group of players. While injuries will still occur in the No Risk group, it is in the best interest of the team to work with the Low Risk and High Risk players to return to No Risk status.

4.2 Limitations and Future Research

The biggest limitation in this study was the means in which we received the data. Since it was an observational study retroactively looking at data collected by the team, we were not involved in collecting the data nor given complete access to the data set. Only six variables were shared with us even though other data captured by the Catapult system such as heart rate activity could have improved model performance. There is a sense that the true extents of the data were not being properly used by the organization. Within their data collection process, there were possibilities of missing data in workouts or system error that could have affected the cumulative load variables, especially in the earlier years when the system was first introduced.

Another area that produced questions was in the process associated with recording the injury data. There were a few injuries that had no dates associated with them and were unusable in this process. It is unclear if these were actual injuries or not and could have had an impact on our results. An aspect of injuries that we wish could have been introduced to the model was time missed from injury. However, the documentation of this was far too sparse to gain any insights from what was given to us. If we were to work with the team moving forward, we would recommend a much more stringent documentation of injuries with greater detail placed on description of the injury and time it took to recover. Another complication to note is staff turnover. The team has had multiple head strength and conditioning coaches as well as other staff turnover during the course of the study. This could influence the philosophy in practice in terms of intensity and could have an affect on confounding variables that have an impact on soft tissue injuries that we were unable to capture like nutrition and recovery.

Future research could be conducted to generalize the results to any football team. At the moment, it is only applicable to the team in the study. This would require other teams providing similar data, and it may be difficult for that many teams to agree to a study like that without some incentives. Some other variables that we would have liked to include would be some measures of flexibility, general strength, sleep, and weight or body fat. Other studies have investigated the connection between flexibility and muscle strength with soft tissue injuries and it could have been informative in regards to hamstring or calf strains, two of the most common injuries seen in the study [5, 25]. This could be measured by a standardized exam at the beginning and end of the season and offseason. Information such as max lifts in squat, deadlift, and bench press were also of interest but unable to be used due to the turnover in strength coaches and the lack of records that were transferred. This caused similar issues with the weight and body fat measurements, as there was not enough continuity over the 4 years of the study for accurate records. Weight was used in the model but had to be scraped from online rosters that may not be that reliable. Sleep has become much more prioritized in professional athletes for its impact on recovery and sleep monitoring technology is becoming more popular. Whoop is one of the systems that has been worn by athletes like LeBron James, Michael Phelps, and Patrick Mahomes [17]. While this would be an interesting component to add to our analysis, the cost to outfit the entire team with bands and to get complete buy in from the players to always wear them to bed may be too challenging.

Chapter 5: Conclusion

This paper shows that combining sampling and algorithmic approaches offers the best solution to the imbalanced classification problem in predicting lower body soft tissue injuries. The model shows some predictive power but suffers from low precision which is not ideal in an injury model when it could lead to players missing games or practices that they do not need to miss. Therefore, we reframed how we view the model in terms of tiers of injury risk by using different cutpoints that took advantage of the higher sensitivity with the MinPValue method and the higher precision with MaxKappa. The model could likely be improved if we were given some level of control over the injury recording methods and more access to the catapult system in the future.

Works Cited

[1] Bowen, Laura, et al. "Spikes in Acute:chronic Workload Ratio (ACWR) Associated with a 5–7 Times Greater Injury Rate in English Premier League Football Players: A Comprehensive 3-Year Study." British Journal of Sports Medicine, Feb. 2019, https://doi.org/10.1136/bjsports-2018-099422.

[2] Carey, D. L., et al. "Predictive Modelling of Training Loads and Injury in Australian Football." International Journal of Computer Science in Sport, vol. 17, no. 1, July 2018, pp. 49–66, https://doi.org/10.2478/ijcss-2018-0002.

[3] CARL FOSTER. "Monitoring Training in Athletes with Reference to Overtraining Syndrome." Medicine Science in Sports Exercise, vol. 30, no. 7, 1998, pp. 1164–68, https://doi.org/10.1097/00005768-199807000-00023.

[4] Casals, Martí, and Caroline F. Finch. "Sports Biostatistician: A Critical Member of All Sports Science and Medicine Teams for Injury Prevention." Injury Prevention, vol. 23, no. 6, Dec. 2016, pp. 423–27, https://doi.org/10.1136/injuryprev-2016-042211.

[5] Comfort, Paul, et al. "Training Considerations after Hamstring Injury in Athletes." Strength and Conditioning Journal, vol. 31, no. 1, Feb. 2009, pp. 68–74, https://doi.org/10.1519/ssc.0b013e318195d225.

[6] Konrad, Alex. "The Australian Tech That's Improving the World's Best Athletes." Forbes, 8 May 2013, www.forbes.com/sites/alexkonrad/2013/05/08/aussie-tech-catapult-gps/?sh=29af7e624c02.

[7] Hägglund, Martin, et al. “Injuries Affect Team Performance Negatively in Professional Football: An 11-Year Follow-up of the UEFA Champions League Injury Study.” British Journal of Sports Medicine, vol. 47, no. 12, May 2013, pp. 738–42, https://doi.org/10.1136/bjsports-2013-092215.

[8] Halperin, Israel, Andrew D. Vigotsky, Carl Foster, and David B. Pyne. ” Strengthening the Practice of Exercise and Sport-Science Research”. International Journal of Sports Physiology and Performance, vol. 13, no. 2, pp. 127-134, https://doi.org/10.1123/ijspp.2017-0322.

[9] Hasanin, Tawfiq, et al. “Severely Imbalanced Big Data Challenges: Investigating Data Sampling Approaches.” Journal of Big Data, vol. 6, no. 1, Nov. 2019, https://doi.org/10.1186/s40537-019-0274-4.

[10] Hsiang, Hao, and Chen (2022). ebmc: Ensemble-Based Methods for Class Imbalance Problem. R package version 1.0.1. https://CRAN.R-project.org/package=ebmc

[11] Leevy, Joffrey L., et al. “A Survey on Addressing High-Class Imbalance in Big Data.” Journal of Big Data, vol. 5, no. 1, Nov. 2018, https://doi.org/10.1186/s40537-018-0151-6.

[12] Li, Ryan T., et al. “Does Overexertion Correlate with Increased Injury? The Relationship between Player Workload and Soft Tissue Injury in Professional American Football Players Using Wearable Technology.” Sports Health: A Multidisciplinary Approach, vol. 12, no. 1, Aug. 2019, pp. 66–73, https://doi.org/10.1177/1941738119868477.

[13] López-Ratón, Mónica, et al. “OptimalCutpoints: AnRPackage for Selecting Optimal Cutpoints in Diagnostic Tests.” Journal of Statistical Software, vol. 61, no. 8, 2014, https://doi.org/10.18637/jss.v061.i08.

[14] Malone, Shane, et al. “The Acute:chonic Workload Ratio in Relation to Injury Risk in Professional Soccer.” Journal of Science and Medicine in Sport, vol. 20, no. 6, June 2017, pp. 561–65, https://doi.org/10.1016/j.jsams.2016.10.014.

[15] Murray, Nicholas B., et al. “Calculating Acute:chronic Workload Ratios Using Exponentially Weighted Moving Averages Provides a More Sensitive Indicator of Injury Likelihood than Rolling Averages.” British Journal of Sports Medicine, vol. 51, no. 9, Dec. 2016, pp. 749–54, https://doi.org/10.1136/bjsports-2016-097152.

[16] Rein, Robert, and Daniel Memmert. “Big Data and Tactical Analysis in Elite Soccer: Future Challenges and Opportunities for Sports Science.” SpringerPlus, vol. 5, no. 1, Aug. 2016, https://doi.org/10.1186/s40064-016-3108-2.

[17] Newcomb, Tim. “A Wearable for Elites like LeBron and Phelps.” Sports Illustrated, 2 Oct. 2015, www.si.com/edge/2015/10/02/tech-talk-wearable-fitness-tracker-whoop-lebron-james-michael-phelps.

[18] Rossi, Alessio, et al. “Effective Injury Forecasting in Soccer with GPS Training Data and Machine Learning.” PLOS ONE, edited by Jaime Sampaio, vol. 13, no. 7, July 2018, p. e0201264, https://doi.org/10.1371/journal.pone.0201264.

[19] Saal, J. A. “Common American Football Injuries.” Sports Medicine (Auckland, N.Z.), vol. 12, no. 2, 1991, pp. 132–47, https://doi.org/10.2165/00007256-199112020-00005.

[20] Seshadri, Dhruv R., et al. “Wearable Sensors for Monitoring the Internal and External Workload of the Athlete.” NPJ Digital Medicine, vol. 2, Nature Publishing Group UK, 2019, p. 71, https://doi.org/10.1038/s41746-019-0149-2.

[21] Strobl, Carolin, et al. “Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution.” BMC Bioinformatics, vol. 8, no. 1, Jan. 2007, https://doi.org/10.1186/1471-2105-8-25.

[22] “Vector.” Catapult, www.catapultsports.com/solutions/vectorspecifications.

[23] Walia, Bhavneet, and Christopher John Boudreaux. “The Cost of Players’ Injuries to Professional Sports Leagues and Other Sports Organizations.” Managerial Finance, vol. ahead-of-print, no. ahead-of-print, July 2020, https://doi.org/10.1108/mf-06-2020-0301.

[24] “What Is Player Load?” Catapult Support, support.catapultsports.com/hc/en-us/articles/360000574716-What-is-Player-Load-.

[25] Worrell, Teddy W., and David H. Perrin. “Hamstring Muscle Injury: The Influence of Strength, Flexibility, Warm-Up, and Fatigue.” Journal of Orthopaedic Sports Physical Therapy, vol. 16, no. 1, July 1992, pp. 12–18, https://doi.org/10.2519/jospt.1992.16.1.12.