# LAX-Score: Quantifying Team Performance in Lacrosse and Exploring IMU Features towards Performance Enhancement

WOOSUB JUNG, Computer Science, William & Mary
AMANDA WATSON, PRECISE Center, University of Pennsylvania
SCOTT KUEHN, Strength & Conditioning, University of Arizona
ERIK KOREM, CEO, AIM7 Inc.
KEN KOLTERMANN, Computer Science, William & Mary
MINGLONG SUN, Computer Science, William & Mary
SHUANGQUAN WANG, Computer Science, Salisbury University
ZHENMING LIU, Computer Science, William & Mary
GANG ZHOU, Computer Science, William & Mary

(a) Lacrosse Athlete

(b) Lacrosse Game

Fig. 1. Women's Lacrosse — Athletes use a netted stick to carry or throw a ball, and shoot for a goal.

For the past several decades, machine learning has played an important role in sports science with regard to player performance and result prediction. However, it is still challenging to quantify team-level game performance because there is no strong ground truth. Thus, a team cannot receive feedback in a standardized way. The aim of this study was twofold. First, we

Authors' addresses: Woosub Jung, Computer Science, William & Mary, Williamsburg, Virginia, wjung01@email.wm.edu; Amanda Watson, PRECISE Center, University of Pennsylvania, Philadelphia, Pennsylvania, aawatson@seas.upenn.edu; Scott Kuehn, Strength & Conditioning, University of Arizona, Tucson, Arizona, scottkuehn88@gmail.com; Erik Korem, CEO, AIM7 Inc. Houston, Texas, erik@aim7.com; Ken Koltermann, Computer Science, William & Mary, khkoltermann@email.wm.edu; Minglong Sun, Computer Science, William & Mary, msun05@email.wm.edu; Shuangquan Wang, Computer Science, Salisbury University, Salisbury, Maryland, spwang@salisbury.edu; Zhenming Liu, Computer Science, William & Mary, lzhenming@gmail.com; Gang Zhou, Computer Science, William & Mary, gzhou@cs.wm.edu.

designed a metric called *LAX-Score* to quantify a collegiate lacrosse team's athletic performance. Next, we explored the relationship between our proposed metric and practice sensing features for performance enhancement. To derive the metric, we utilized feature selection and weighted regression. Then, the proposed metric was statistically validated on over 700 games from the last three seasons of NCAA Division I women's lacrosse. We also explored our biometric sensing dataset obtained from a collegiate team's athletes over the course of a season. We then identified the practice features that are most correlated with high-performance games. Our results indicate that *LAX-Score* provides insight into athletic performance beyond wins and losses. Moreover, though COVID-19 has stalled implementation, the collegiate team studied applied our feature outcomes to their practices, and the initial results look promising with regard to better performance.

CCS Concepts: • **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**.

Additional Key Words and Phrases: Wearable Sensing, Machine Learning, Athletic Performance Enhancement

**ACM Reference Format:**
Woosub Jung, Amanda Watson, Scott Kuehn, Erik Korem, Ken Koltermann, Minglong Sun, Shuangquan Wang, Zhenming Liu, and Gang Zhou. 2021. LAX-Score: Quantifying Team Performance in Lacrosse and Exploring IMU Features towards Performance Enhancement. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 109 (September 2021), 28 pages. https://doi.org/10.1145/3478076

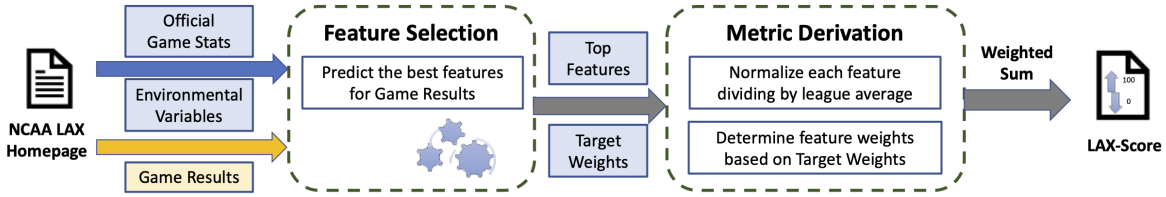
## 1 INTRODUCTION

The field of sports science has grown enormously over the past decades. As part of the growth in this field, researchers have leveraged machine learning for various purposes to improve sports performance. Insights from machine learning lead to intelligent training protocols, which can help an athlete run faster, jump higher, and lift heavier weights than before [11, 32, 41]. Researchers have also utilized machine learning techniques to prevent injuries in athletes [6, 8, 12] or assess the efficacy of training activities [18, 31, 44, 47, 51, 78]. Several elite professional sports such as basketball and baseball have already adopted machine learning in their injury management strategies [9]. However, machine learning is still underutilized in enhancing team-level athletic performance. Machine learning can draw important connections between an athlete's performance in practices and their performance in matches.

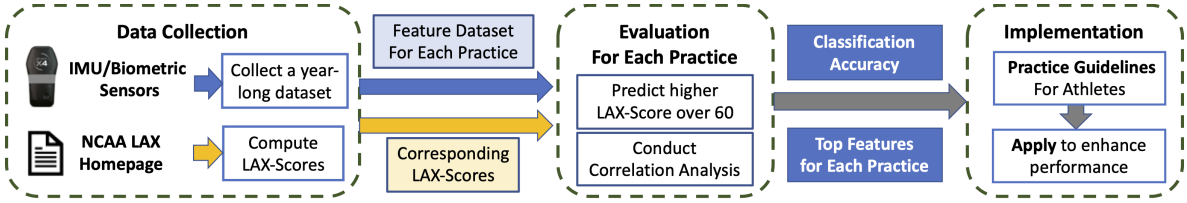
To achieve better performance over time, a team first needs to understand whether their current game performance is satisfactory. However, the field of performance enhancement in sports suffers from two related issues: disagreement on the definition of high performance and thus a lack of feedback from an official match. Some may argue that winning constitutes high performance [16], but others believe that other factors, such as skill level of the opposing team or overall game score, should be taken into account [50]. Since the definition of high performance is still controversial, it is also challenging to offer and receive feedback from the match in a standardized way.

To provide standardized feedback, we designed an intuitive metric that quantifies team-level game performance. Our metric can therefore be used to design better practice strategies for improved game performance in the future. Note that we explore the connection between practice and performance instead of the connection between practice and winning. Our reasoning is that winning and losing are not always indicators of performance and are contextual. Game outcome depends on not just a team's performance but other factors as well. Though winning games is the ultimate goal of every sports team, we believe that enhancing performance will help a team achieve an upward trajectory of winning over time. Therefore, we explore the connection between practice factors and game performance.

In this study, we aim to enhance the performance of a collegiate lacrosse team by leveraging machine learning techniques. We selected lacrosse because even though it is one of the fastest-growing collegiate sports in North America, to the best of our knowledge, no study to date has applied machine learning techniques for performance enhancement. Moreover, lacrosse is bio-mechanically and logistically similar to other team sports such as field hockey and ice hockey. For example, lacrosse athletes use a stick to carry or throw a ball, as shown in Figure 1a.




(a) Deriving LAX-Score — Sections 3 and 4




(b) Connecting LAX-Score with Practice — Sections 5 and 6

Fig. 2. Overview of LAX-Score and Connecting it with Practice Sessions

Athletes also shoot for a goal, and the team that scores the most goals wins the game, as illustrated in Figure 1b. Therefore, though this study focuses specifically on lacrosse, our approaches and findings have potential applicability to other team sports.

Our research questions in this study are as follows:

- RQ1: How can we quantify team-level athletic performance beyond wins and losses?
- RQ2: How can we evaluate this metric in light of the fact that there is no clear ground truth for performance?
- RQ3: How can our metric be utilized to enhance team practices to achieve higher game performance?

We first propose a new metric named *LAX-Score* to quantify game performance. In our study, we investigated two categories of features that are most related to game results: environmental factors, such as team ranking and game schedule information, and in-game statistics. By using feature selection, we extracted the best features for predicting game results and determined the feature weights. Figure 2a briefly shows how we derived *LAX-Score*. In short, *LAX-Score* takes the outcomes of previous games as input to evaluate the team-level performance. Our statistical analysis and comparison with the official game-note give insights into how well *LAX-Score* reflects the athletic performance beyond wins and losses.

Once we had validated the *LAX-Score* metric, we investigated the correlation between practice features and LAX-Scores. Our premise is that LAX-Scores are highly correlated to practice sessions before each game; teams that practice more efficiently are likely to perform better. To analyze this relationship, we collected Inertial Measurement Unit (IMU) sensing data and heart rate data from a women's collegiate lacrosse team during practices and games over the course of the 2019 season. Next, we explored which of these practice features are correlated with games with higher *LAX-scores*. Our results suggest that a subset of the practice features are more correlated with high athletic performance than the rest. These feature sets indicate how coaches can improve their practice sessions to yield better team performance. The data analysis was conducted in the break between the 2019 and 2020 seasons; we provided our feature recommendations to the team in advance of the 2020 season. Figure 2b provides an overview of the procedures used.

In summary, our contributions in this paper are threefold.

- We propose a game-performance metric for lacrosse, *LAX-Score*, that quantifies a variety of performance factors both on and off the field. This metric is derived from machine learning feature selection that takes both game statistics and environmental variables into account.
- We calculated over 700 *LAX-Scores* based on three seasons of official NCAA game statistics for 50 collegiate teams. Then, we confirmed that this metric is a linear function of teams' winning-percentages and follows the Gaussian distribution.
- Over the course of a season, we collected a vast IMU/biometric dataset from a collegiate team. Using this data, we identified which attributes of a practice session, such as acceleration or heart rate level, contributed to high athletic performance in a game setting.

The remainder of this paper is organized as follows: Section 2 summarizes other studies related to this paper. In Section 3, we propose a new metric called *LAX-Score*. Then, Section 4 illustrates the performance of this metric. Section 5 describes the process of collecting data from the collegiate team. In Section 6, we select the practice features that can best help athletes improve game performance. Next, we discuss implementation and future work in Section 7 and conclude our paper in Section 8.

## 2 RELATED WORK

Section 2.1 introduces performance evaluation metrics in athletics. In Section 2.2 shows how other researchers applied machine learning in other sports. Section 2.3, we describe a few studies relevant to lacrosse, which are mostly simple statistics.

### 2.1 Performance-related Metrics

While work on performance-related metrics has been done for other sports, there is a dearth of this research for lacrosse. Table 1 summarizes athletic performance metrics in other sports. James devised baseball's *Game-Score* in the 1980s [40]. Even though this metric was designed for understanding pitchers' performance, it influences several characteristics of our own metric of *LAX-Score*. This metric does not have an upper limit. While the highest score earned to date is 105, the common range is 0-100 with a mean of roughly 50. *Game-Score* also has a base point of 50 and takes the Weighted Sum Model (WSM) in its formula [21]. Unfortunately for the purposes of this metric, it is designated to evaluate an individual player's performance and only takes into account hand-picked game statistics. Basketball also has its own *Game-Score* for player evaluation [35]. Like in baseball, it is a simple WSM version of several game statistic variables. Unlike *LAX-Score*, the weights of those two *Game-Scores* are manually determined and do not change over time. Thus, the fixed weights cannot reflect any information across different seasons. Furthermore, neither of the *Game-Scores* integrates league normalization or environmental features into the metrics. They totally rely on game statistics, which can not offer broader perspectives. In soccer, while there is no widely used metric for performance evaluation, the authors in [64] introduced a weighted metric on a per-minute basis for individual soccer players' performance. This metric considers the opponent's

Table 1. Comparison with Other Performance Metrics present in Team Sports

| Metric | Sport | Target | Feature Selection | Features | Mean | Highest | Normalization | Weights | Statistical Benefit |
|---|---|---|---|---|---|---|---|---|---|
| **Game-Score [40]** | Baseball | Individual | Manual | 7 game stat | **50** | **105** | No | Fixed | Not reported |
| **Game-Score [35]** | Basketball | Individual | Manual | 12 game stat | 10 | 65 | No | Fixed | Not reported |
| **PER [36]** | Basketball | Individual | Manual | 10 game stat | 15 | 30 | **League, Team** | **Dynamic (Season)** | Fixed Mean Value (15) |
| **Schultze in [64]** | Soccer | Individual | Manual | 1 game stat, **1 env. var** | 0 | 25 | No | **Dynamic (Game)** | Not reported |
| **LAX-Score*** | Lacrosse | **Team** | **Machine-learning** | **6 game stat, 2 env vars** | **50** | **105** | **League** | **Dynamic (Season)** | **Gaussian Distribution, Linear to Winning Pct.** |

strength and weighs important features. However, the scale is not intuitive and only takes two features into account, which may suffer from overfitting.

Most of all, they all focused on individual players' performance in team sports, including PER in basketball [36]. Since the literature still lacks team-level performance indicators, it is hard for any team to assess official games and enhance their practice sessions. Another challenge in this field is that no strong ground-truth for athletic performance exists. It is also difficult to evaluate whether the existing metrics are statistically meaningful. In the absence of something more effective, however, those metrics have been empirically adopted and widely used with no clear theoretical benefits. To tackle these issues, we compare *LAX-Scores* from several games with official game-notes as well as provide statistical analysis in Section 4.

### 2.2 Machine Learning in Athletics

The literature has been working on improving athletic performance and game prediction, which were already covered by several surveys [10, 15, 59]. Several studies have also utilized machine learning techniques [7, 39, 48]. For example, the authors in [14] conducted activity recognition using a random forest classification model. Using video files as ground truth, they classified tackle events and validated them in eight international games. The accuracy of these experiments is about 80%. In [76], the authors also classified team sport-related activities with 76 male players. The activities include stationary, walking, jogging, running, changing direction, counter-movement jumping, jumping for distance, and tackling. After extracting hand-crafted features, they used Random Forest, SVM, and Logistic Model Tree (LMT). LMT model achieved the best results with an accuracy between 79 and 92 %. Ryan et al. [62] collected in-game running data from 34 professional football players in 15 games. After collecting total and high-speed running, they constructed a linear model to examine the impact of player characteristics. The authors in [71] used GPS to assess the physical demands of Division I College Football. Gentles et al. [26] studied a women's college soccer season in its entirety, which includes 17 matches and 24 practices. Fox et al. [22] quantified and compared training and competition demands in basketball. Garrett et al. [25] explored the relationship between athletic ability assessment and maximum running velocity.

### 2.3 Lacrosse-related Research

Even though many sports have already taken advantage of machine learning skills, lacrosse is still inactive to adopt machine learning techniques despite recent growing popularity. Most of them focused on experiments for practice strategies as described below. Polley et al. [57] provided velocity band settings and measured workload. Based on the measurement, they suggested that higher intensity activities may require more recovery times. In [49], the authors mainly focused on the impact of stick carry on field test performance by position. They collected data from sprint tests or jumping tests to measure the impact. Randolph et al. [61] validated the effectiveness of different training programs. In this research, three phases of training sessions were conducted during pre-season to show the performance increase as the athletes were trained. The authors in [65] described fitness differences based on positions and starting status. During pre-season, Vescovi et al. [68] also explored positional differences from various training programs. Akiyama et al. [3] described characteristics of different positions. They observed that attackers and defenders have some better abilities than midfielders. In the thesis [45], Klemz focused on how to avoid injury through different training sessions. The authors in [34] dug into positional differences in some training programs. Steinhagen et al. [67] mentioned positional differences in skill levels as well. Hauer et al. in [33] used game data but is limited to small-sided training games during pre-season. Based on the experiments, they claimed that those small-sided games could improve athletes' performance. The authors in [20] measured and described training abilities from female athletes, while Vincent et al. [69] reviewed lacrosse-related injuries in high school and collegiate players. Plisk in [56] used logistic regression for winning determinants. Among many features, goal per game is the most related feature as a matter of course.

## 3 *LAX-SCORE*: QUANTIFYING GAME PERFORMANCE

In Section 3.1, we introduce our motivation and provide a conceptual overview of the *LAX-Score* metric. Then, we explain how we derived this metric from publicly-available data in the following subsections.

### 3.1 *LAX-Score* Motivation

We designed *LAX-Score* to quantify game performance in a standardized way. To the best of our knowledge, two existing methods can be used for performance evaluation: game results and game notes. However, they are not clear performance indicators. For example, game results whether a team won or lost can provide a rough approximation of performance. However, this binary outcome leaves out a great deal of detail that would provide a more realistic picture of a team's performance. Additionally, winning a game is not equivalent to high performance, and vice versa [50]. Thus, an athletic team needs a more thorough consideration of its performance in game settings. To obtain a more thorough picture, we also look to game notes, which provide a summary of events in each NCAA game. These game highlights, or notes, are short narratives describing general actions in the game. Even though game notes are typically written by coaches, and therefore can be subjective, they are the golden standard that describes performance beyond wins and losses for the following reasons: 1) They not only provide statistics but also convey the quality of the game in chronological order, such as who takes the lead first or how intense the game is, and 2) they use descriptive language, which provides subjective descriptions such as "outperform", "impressive", "poor", etc. Nonetheless, the official notes may seem subjective and can be more qualitative than quantitative. Therefore, we propose a single value metric called *LAX-Score* that describes team performance beyond the binary of winning and losing and the subjective quality of game notes.

*LAX-Score* captures game performance using a 0-100 scale. This scale is not strictly bound like other metrics. Thus, either a *LAX-Score* above one hundred or below zero is always possible. However, these outliers would be generated by circumstances that are extremely unlikely to occur. Of the hundreds of games tested in this study, only one game's *LAX-Score* was outside of the 0—100 range; the maximum *LAX-Score* value is 105.4, while the minimum is 2.8. This scale makes *LAX-Score* familiar as most countries use a percentage-based grading system for academic grades [74].

*LAX-Score* also provides an intuitive understanding of team performance. Table 2 shows our collegiate team's *LAX-Scores* for each of the 15 games of the 2019 season along with each game's results. For example, higher *LAX-Scores* correspond to winning games, like games 1, 2, and 6. On the other hand, losing games tend to have lower *LAX-Scores*. Meanwhile, we also observed that the team achieved high *LAX-Score* in some losing games, like games 5 and 10, while the team received poor *LAX-Scores* results when they won, like games 7 and 12. *LAX-Score* is effective, however, because it captures these nuances while also remaining a linear function of winning-percentage and following the Gaussian distribution. Thus, if a team's *LAX-Score* is in the 60s, its chance of winning a game is about 65% and its performance is roughly in the 30th percentile, which will be demonstrated in Section 4.

More importantly, *LAX-Score* is a more accurate indicator of team's performance than the game result. For example, in game 10, although the team lost, the official note describes its performance in positive terms. The game-note reads as follows: "The team outshot the opponent, 31-27, and committed two fewer turnovers [10-12]… Although the team owned an 11-10 lead following a goal with 13:39 remaining, the opponent closed the game

Table 2. The team's Game Results and *LAX-Scores* during the 2019 season

| **Game** | **#1** | **#2** | **#3** | **#4** | **#5** | **#6** | **#7** | **#8** | **#9** | **#10** | **#11** | **#12** | **#13** | **#14** | **#15** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Score** | 10-6 | 17-10 | 8-16 | 3-19 | 9-12 | 13-5 | 9-8 | 17-13 | 11-12 | 11-14 | 13-15 | 12-11 | 12-22 | 9-16 | 9-12 |
| **Result** | Win | Win | Loss | Loss | Loss | Win | Win | Win | Loss | Loss | Loss | Win | Loss | Loss | Loss |
| **LAX-Score** | 64.7 | 60.7 | 40.5 | 16.7 | 56.9 | 68.9 | 30.8 | 48.2 | 53.0 | 61.5 | 39.3 | 35.8 | 28.4 | 45.9 | 56.9 |

Table 3. Selected Features for the *LAX-Score* metric

| Feature | Type | Description | Information Gain | | Pearson's Correlation | | Target Weight |
|---|---|---|---|---|---|---|---|
| | | | *Avg.Value ± SD* | *Avg.Ranking ± SD* | *Avg.Value ± SD* | *Avg.Ranking ± SD* | |
| **Goals** | Official Statistic | An instance of scoring the team made | 0.452 ± 0.017 | 1.2 ± 0.4 | 0.682 ± 0.008 | 1.3 ± 0.046 | 56.7 |
| **Goals Allowed** | Official Statistic | Goals that the team allowed | 0.397 ± 0.014 | 1.9 ± 0.54 | 0.661 ± 0.008 | 1.7 ± 0.046 | 52.9 |
| **Shots** | Official Statistic | Throwing the ball in an attempt to score | 0.222 ± 0.022 | 4.4 ± 0.8 | 0.541 ± 0.009 | 4.1 ± 0.3 | 38.2 |
| **Shots Allowed** | Official Statistic | Shots that the team allowed | 0.159 ± 0.021 | 4.8 ± 0.4 | 0.458 ± 0.017 | 5.1 ± 0.54 | 30.9 |
| **Turnovers** | Official Statistic | When a team loses possesion of the ball | 0.009 ± 0.002 | 6.6 ± 0.4 | 0.151 ± 0.024 | 5.8 ± 0.4 | 8.0 |
| **Caused Turnovers** | Official Statistic | Turnovers that opponents made | 0.001 ± 0 | 7.6 ± 0.2 | 0.05 ± 0.024 | 7.1 ± 0.3 | 2.6 |
| **Time-Off** | Environmental Statistic | The number of days between games | 0.002 ± 0 | 6.8 ± 0.2 | 0.073 ± 0.014 | 7.4 ± 0.49 | 3.8 |
| **Ranking Difference** | Environmental Statistic | Ranking difference from opponents | 0.38 ± 0.3 | 2.6 ± 0.3 | 0.642 ± 0.008 | 3.0 ± 0 | 51.1 |

with four unanswered goals to secure the conference victory." By some metrics, the team played better than their opponent and almost won the game against the high-ranked team. Our proposed *LAX-Score* calculates this game's score as 61.5, which is an above-average performance. On the other hand, in game 7, the game-note for the winning game sounds like the team struggled against a lower-ranked team. The official note states: "Freshman scored the game-winner with just 0:18 remaining… The closely contested matchup featured six ties, and neither team led by more than two goals at any point… The opponent outshot the team, 39-18… The opponent committed just seven turnovers." In this case, even though the team won, they did not outperform the opponent, and by several metrics, the opponent played much better. These descriptions are compatible with the proposed *LAX-Score* for this game of 30.8, which is below average. These two examples illustrate that *LAX-Score* indicates team performance more accurately. We will further validate *LAX-Score* in Section 4.

## 3.2 Feature Selection

For use in *LAX-Score* derivation, we leveraged machine learning to choose better features that can demonstrate game performance. Even though there is no quantitative ground truth for team-level performance at the time of writing, the indicator of whether they won or lost could be the simplest way of defining athletic performance. Thus, we alternatively explored the best features relevant to the game results.

Accordingly, we prepared several candidate features. First, we considered the official game statistics such as shots, goals, and turnovers, which are accessible on the official website [53]. Furthermore, more factors could impact game results besides in-game statistics. Hence, we also took into account environmental variables, such as opponent rankings or rest times between games. The environmental features were manually calculated from the official website as well. Consequently, we collected our candidate features for 271 NCAA games in the 2019 season as follows: *In-game statistics = {Goals, Goals allowed, Shots, Shots Allowed, Turnovers, Turnovers Allowed, Draw Controls, Draw Controls Allowed, Ground Balls, Ground Balls Allowed}* and *Environmental statistics = {Ranking Difference, Number of days between games, Home and Away, Day and Night}*.

After collecting the candidate features, we utilized Weka [24, 75] for feature selection, which has been widely used for feature selection as well as classification problems [66]. More specifically, we fed the candidate features for the 271 games to Weka for game result prediction, as shown in Figure 2a. For the feature selection tool, we used two well-known techniques — Information Gain and Pearson's Correlation evaluator [58, 63] for comparison. The Information Gain evaluator tells entropy reduction of each attribute that is related to the target variable for

decision trees and sorts them in a high-rank order. Pearson's Correlation measures the correlation between each feature and the game result and sorts them as well.

To avoid overfitting of feature selection, we divided the 271 games into ten subsets so that each subset consists of 27 or 28 games. For each subset, we calculated each feature's feature selection value and its ranking. Then, we averaged those ten values and rankings to select top-ranked features for our metric. Table 3 summarizes the top eight features that are most related to the game results from the two evaluators. The *Value* columns show either the average value of Information Gain or Pearson's Correlation, and the *Ranking* columns display the average ranking of the corresponding feature. For example, according to the Information Gain result, *Goals*, *GoalsAllowed* and *RankingDifference* values are 0.452, 0.397, and 0.38, whereas Pearson's Correlation has 0.682, 0661, and 0.642, respectively. They are the top three features in both feature selection. These similar rankings and the small standard deviations also confirm that the feature selection results are stable across different games. Moreover, other performance metrics [21, 40] also include the same attributes like *Goals*, *Shots*, or *Turnovers* as their hand-picked features. Thus, our selected features can be considered acceptable in Sports Science.

Based on the feature selection results, we also determined a target weight for each feature. Since we wanted the *LAX-Score* to reflect the same importance of each feature, we simply computed the target weight by averaging the two values from Information Gain and Pearson's Correlation and multiplying by 100 to scale to 0-100. For example, the mean value of the two feature selections for *Goals* is 0.567, and we scaled it up to 56.7 for its target weight. Likewise, we also calculated other features' target weights, as shown in Table 3. After all, we take these eight features and their target weights to derive the *LAX-Score* in the following subsection.

### 3.3 Metric Derivation

We derive our new metric, *LAX-Score*, consisting of the weighted sum of each feature as follows:

$$\begin{aligned} \textit{LAX-Score} = \textit{Base Point} \\ &+ W_1(\textit{Goals}/\textit{lgGoals}) + W_2(\textit{GoalsAllowed}/\textit{lgGoalsAllowed}) \\ &+ W_3(\textit{Shots}/\textit{lgShots}) + W_4(\textit{ShotsAllowed}/\textit{lgShotsAllowed}) \\ &+ W_5(\textit{Turnovers}/\textit{lgTurnovers}) + W_6(\textit{CausedTurnovers}/\textit{lgCausedTurnovers}) \\ &+ W_7(\textit{Time-Off}/\textit{lgTime-Off}) \\ &+ W_8(\textit{RankingDifference}/\textit{lgRankingDifference}) \end{aligned} \tag{1}$$

Since there are multiple features to determine athletic performance, we applied a form of the Weighted Sum Model (WSM), which is a popular multi-criteria decision analysis [21], to the *LAX-Score*. In Equation 1, we first divide each feature on a given day by its league average to normalize the features. The league average values begin with the prefix lg-. This normalization helps to compare each feature with other games. For example, if a

Table 4. Feature Weights in *LAX-Score* for the 2019 season

| | Target Weight | Maximum Value | Average Value | Max./Avg. | Factor Effect | Feature Weight |
|---|---|---|---|---|---|---|
| **Goals** | 56.7 | 25 | 11.96 | 2.09 | Positive | **27.13** ($W_1$) |
| **Goals Allowed** | 52.9 | 25 | 12.30 | 2.03 | Negative | **-26.06** ($W_2$) |
| **Shots** | 38.2 | 56 | 28.44 | 1.97 | Positive | **19.37** ($W_3$) |
| **Shots Allowed** | 30.9 | 49 | 28.44 | 1.72 | Negative | **-17.94** ($W_4$) |
| **Turnovers** | 8.0 | 30 | 16.26 | 1.85 | Negative | **-4.32** ($W_5$) |
| **Caused Turnovers** | 2.6 | 24 | 8.75 | 2.74 | Positive | **0.93** ($W_6$) |
| **Time-Off** | 3.8 | 14 | 4.78 | 2.93 | Negative | **-1.28** ($W_7$) |
| **Ranking Difference** | 51.1 | 0.233 | 0.077 | 3.03 | Positive | **16.87** ($W_8$) |

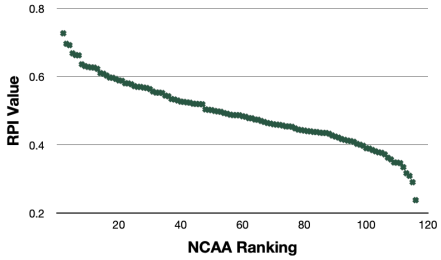



Fig. 3. NCAA Rankings vs. RPI Values in the 2019 season — RPI values follow a polynomial function

team scores more than the league average, the team should receive more credit than when scoring less than the league average.

After that, we multiply a derived weight to adjust each feature where $W_i$ denotes the feature weight. Table 4 briefly shows how to derive each weight for a given season. For example, in the 2019 season, the maximum number of goals in a single game is 25, and the mean of goals is 11.96. Thus, when we divide *Goals* by *lgGoals*, it would be 2.09 at most. Now, we divide the target weight by the *Maximum*/*Average* value to get the season feature weight. In this case, we divide 56.7 by 2.09, so the derived weight for *Goals* is 27.13. Since the derived weights depend on the maximum and league average values, the feature weights should be recalculated every season.

Besides the derived season weights, some features are considered positive factors for athletic performance, while others are negative. For instance, *Goals* scored or *Shots* made is considered to be positive features. On the other hand, *GoalsAllowed* and *ShotsAllowed* are negative. Longer days between games (*Time-Off*) infers that a team has enough time to get recovered and prepared. Hence, longer *Time-Off* is considered to be a negative factor in the *LAX-Score*. *RankingDifference* is another huge factor since its target weight is above 50. We note that we adopt RPI (Rating Percentage Index) values [72] instead of using the official rankings. This is because the official ranking is the order of RPI values in NCAA. RPI is calculated based upon a team's wins and losses and its strength of schedule in range 0 and 1. For example, the top RPI value in the 2019 season is 0.7441, while the lowest is 0.2375. Moreover, the RPI values are not linearly distributed but follow a polynomial function, as shown in Figure 3. Thus, it makes more sense to use the actual RPI values for reflecting the level of the opponent teams. In consequence, *RankingDifference* is calculated as in Eq. 2. This makes *RankingDifference* either positive or negative in which a positive value means that the opponent is a higher-ranked team.

$$RankingDifference = Opponent's\ RPI\ Value - Team's\ RPI\ Value. \tag{2}$$

The base score of the metric is 50, which is also determined for the readability of *LAX-Score*. With the base score, the mean values of *LAX-Scores* in the past three seasons are all around 50, which we will show in the following section. All the above factors considered, we determine the weighted sum of the selected features in the *LAX-Score*.

## 4 *LAX-SCORE*: METRIC EVALUATION

In Section 4.1, we compare *LAX-Scores* with game results to demonstrate that *LAX-Score* provides extra information beyond wins and losses. In Section 4.2, our results show that *LAX-Score* is a more accurate reflection of real-world performance than winning percentage alone, as validated by athletic coaches. Section 4.3 illustrates that *LAX-Score* is a linear function of winning percentage and follows the Gaussian distribution. In Section 4.4, we discuss weight optimization of our metric throughout three seasons.

Table 5. *LAX-Score* Exemplary Types with Game Statistics

| Game Type | Team (Ranking) | Opponent (Ranking) | Goals | Goals Allowed | Shots | Shots Allowed | Turnovers | Caused Turnovers | Time -Off | LAX -Score | Game-Note Tone | Game Result |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **#1** | A (22) | B (62) | 16 | 15 | 32 | 36 | 25 | 17 | 5 | 25.2 | *Negative* | *Win* |
| **#2** | C (88) | D (22) | 13 | 14 | 30 | 35 | 15 | 6 | 5 | 76.0 | *Positive* | *Loss* |
| **#3** | E (71) | F (87) | 19 | 2 | 42 | 9 | 15 | 14 | 4 | 105.4 | *Positive* | *Win* |
| **#4** | G (47) | H (22) | 5 | 24 | 17 | 39 | 18 | 11 | 5 | 6.1 | *Negative* | *Loss* |

For metric evaluation, we first computed *LAX-Scores* for 271 NCAA women's lacrosse games across two different conferences during the 2019 season, which is approximately 20% of all of the games played in that season. Before analyzing the *LAX-Score* dataset, however, we need to evaluate whether this sample is representative of the population. During the 2019 season, the sample mean ($\overline{x}$) of *LAX-Scores* is 50.18 and the standard deviation of the sample ($s$) is 17.18. In addition, the standard error of the sample mean ($SE_{\overline{x}}$) is calculated in Equation 3 [5]. Given the standard deviation of the sample ($s$) and the sample size ($n$), the standard error is 1.04. Thus, the estimation of a population mean ($\mu$) at 95% confidence is $\mu_{95\%} = 50.18 \pm 2 * 1.04$, which indicates that the collected dataset is trustworthy.

$$SE_{\overline{x}} = \frac{\sigma}{\sqrt{N}} \simeq \frac{s}{\sqrt{n}} \tag{3}$$

## 4.1 Comparing *LAX-Score* with Game Result

In this section, we will compare *LAX-Scores* and game results with the tones of official notes. As discussed in Section 3.1, official notes can provide ground truth because even though they are subjective, they still provide more information than a game result. However, it is challenging to manually compare *LAX-Scores* and game results on a game-by-game basis. Accordingly, we categorized games into four representative types of *LAX-Scores* to best exemplify the effectiveness of *LAX-Score*. The types are defined as the relationship between performance and game result as follows: *Positive-Win*, *Negative-Loss*, *Negative-Win*, and *Positive-Loss*. Note that labeling tones in these notes require specific domain knowledge. In order to remedy this, we asked two lacrosse coaches to manually label the tones of the four game notes. They were asked to label the overall sentiment of the game notes irrespective of which team won the game. The coaches agreed upon the four games as follows.

As shown in Table 5, Game 1 is an example of a *Negative-Win* game. In terms of game statistics, the higher-ranked team (A) committed more turnovers (25-17) and took less shots (32-36). The game-note of game 1 contained the following descriptions: "The opponent [B] closed the first half on a five goal run giving the team [A] a 9-7 lead at the intermission.The pressure kept up in the second half with the opponent [B] scoring four more times coming out of the break giving the opponent [B] a 11-9 lead at the 25:08 mark. The score went back-and-forth over the next nine minutes." Considering the huge ranking difference, Team A did not perform as well as their ranking would lead one to expect, but they were still able to win the game. In spite of their victory, Team A's *LAX-Score* for this game was 25.2, and the coaches also labeled the game note as negative. A *LAX-Score* below 32.4 indicates a performance below the 15th percentile in this season. A team with this score has a 75% chance of losing a game. Moreover, the 25% games that result in victory were conducted by stronger teams. This 25% chance of winning is accounted for by strong teams outperforming weak opponents by just enough to win the game.

Game 2 is a *Positive-Loss* game. According to the game statistics, the lower-ranked team (C) played well against a much higher-ranked team but ultimately lost 13-14. The official game-note for game 2 reads as follows: "It was a defeat in a spirited, gritty performance... The team [C] went toe-to-toe with the opponent [D] that sits atop their conference." Both *LAX-Score* and the game-note indicate that the lower-ranked team (C) played quite well even though they lost. Thus, Team C achieves a higher *LAX-Score* despite the defeat. The two coaches also

analyzed this game-note as positive. Team C's *LAX-Score* for this game was 76. In this season, a *LAX-Score* above 68 indicates a performance in the top 15th percentile and an 83.3% chance of winning games. The remaining 16.7% games were conducted by lower-ranked teams. This 16.7% of high *LAX-Score* games that result in defeat represent low-ranking teams performing well, but not well enough to defeat higher-ranked teams.

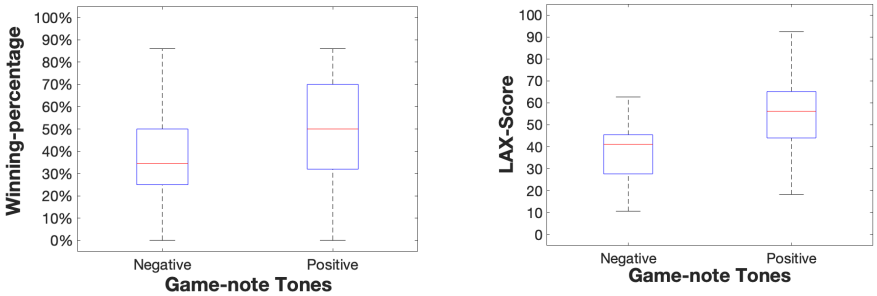
Game 3 is classified as *Positive-Win*. In this type, the tone of the game note is positive, and the result is a win. Although the two teams' NCAA rankings were similar, the winning team (E) outperformed the losing team (F) — they took more shots (42-9) and scored more goals (19-2). According to its official note, this game was one of their best performances in the century. The note says: "The impressive offensive performance marked [sic]. The 17-goal margin [19-2] was the program's largest victory since 2017. The defensive effort marked the fewest goals allowed [2] since 2001." The coaches also classified this game-note as positive. This positive tone matches a high *LAX-Score* of 105.4, which is the highest score in the last three seasons.

Finally, Game 4 exemplifies the *Negative-Loss* type. Team G was outshot (17-39) and outscored (5-24) while they allowed more turnovers (18-11) than their similarly ranked opponent (H). The official game-note for this game is extremely short, inferring that Team G's performance on that day was too poor. The official note of game 4 also states as follows: "The team [G] scored their fewest goals [5] of the season. The opponent [H] was not kind to the team. The opponent [H] led for nearly the entirety of the game. The team [G] committed 18 turnovers and recorded just 17 shots." This naturally led to a poor *LAX-Score* result, 6.1. The coaches also labeled this tone of the game-note as negative.

In summary, the four examples of *LAX-Scores* are compatible with the tones of the game notes. More importantly, the *LAX-Score* newly identified the types of games 1 and 2, which binary outcomes were inconsistent with their team performance. This property of the metric is beneficial because a team can receive feedback in a standardized way beyond the game results.

## 4.2 Comparing *LAX-Score* with *Winning-percentage*

*LAX-Score* is also a better metric than *Winning-percentage* for understanding game performance. We demonstrate how to calculate *Winning-percentage* in Equation 4 [73]. Note that a team's *Winning-percentage* value is 50% at the beginning of the season and fluctuates as the season progresses. This metric can be used to predict the likelihood of a team winning a particular game. To evaluate which metric — *LAX-Score* or *Winning-percentage* — better encapsulates game performance, we compared these two metrics with game notes across 100 games. The two coaches labeled 100 post-game write-ups randomly sampled from the original 271 games. They were given the




(a) Winning-percentage and Game-note (b) LAX-Score and Game-note

Fig. 4. Comparison of *Winning-percentage* and *LAX-Score* with Game-Notes Tones — Ranges and Variances show that *LAX-Score* is more reliable than *Winning-percentage* to understand team performance beyond game results.

Table 6. Comparison between *Winning-percentage* and *LAX-Score* of Each Type (Average ± Standard Deviation)

| | Negative-Loss | Negative-Win | Positive-Loss | Positive-Win |
|---|---|---|---|---|
| **Winning-percentage (%)** | 35.2 ± 23.2 | 59.8 ± 25.2 | 31.7 ± 20.7 | 56.2 ± 24.2 |
| **LAX-Score** | 37.6 ± 14.8 | 46.5 ± 19.5 | 47.4 ± 16.6 | 58.0 ± 17.2 |

same set of games. It took each coach approximately 6 hours to label these game notes. Each game constituted a datapoint. We chose the final label for each datapoint as follows: If the coaches agreed on the sentiment of the game notes, we labeled the datapoint according to their consensus. If the coaches disagreed, we abstained from labeling that datapoint. In total, the coaches agreed on 91 of the 100 labels. Overall, the breakdown of our labels is as follows: Positive: 60, Negative: 31, and Abstain: 9.

$$\textit{Winning-percentage (\%)} = \{\text{the number of winning games}\}/\{\text{the number of all games played}\} \times 100 \tag{4}$$

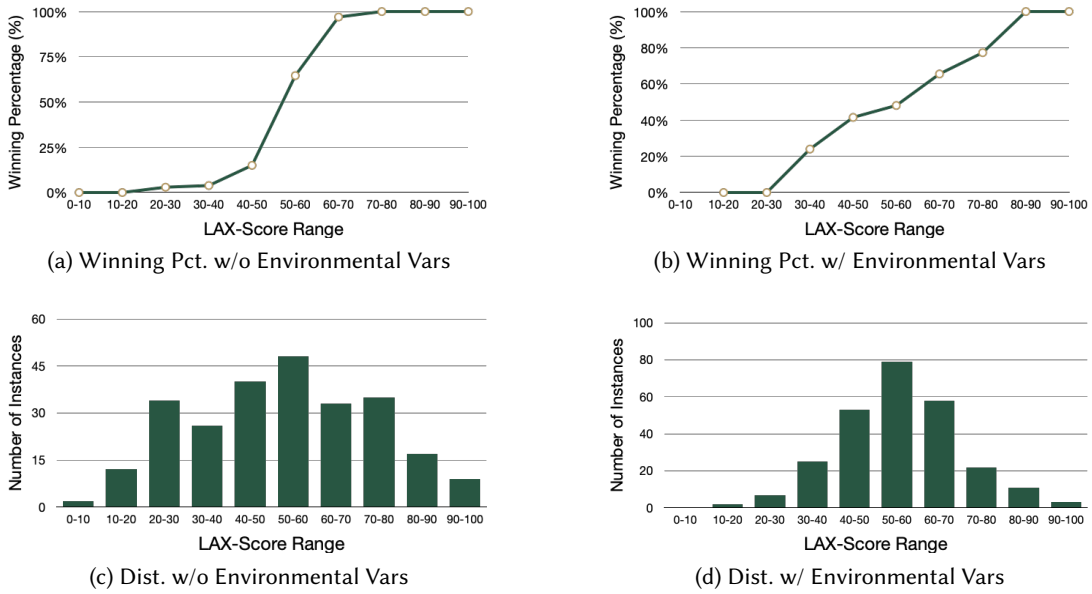
Next, we conducted a correlation analysis on this dataset to determine which metric best aligned with the labeled tones of the game notes. Figure 4 shows that *Winning-percentages* have the same ranges for positive- and negative-toned games, while *LAX-Scores* are partially overlapped. More specifically, in Figure 4a, some games with high *Winning-percentages* (up to 85%) still have negative tones, while games with low *Winning-percentages* can also yield positive performances. On the other hand, in Figure 4b, most of the games that have negative tones achieved a *LAX-Score* of 60 or below, and the games with positive tones have much higher *LAX-Scores*. Overall, the correlation between *LAX-Scores* and game-note tones is 0.40, and the correlation between *Winning-percentages* and game-note tones is 0.18.

Our findings were similar when we examined *LAX-Score* and *Winning-percentage* based on the four types articulated in Section 4.1. Table 6 shows the average values of each type. For the *Negative-Loss* and *Positive-Win* types, *LAX-Score* and *Winning-percentage* have almost the same values. Those are the games that occurred as expected based on *Winning-percentage*. However, for the *Negative-Win* type, even though *Winning-percentages* are comparatively high, the tones of the games note are negative. Meanwhile, the average of the *LAX-Scores* is below 50, but still higher than the *Negative-Loss* type of games (37.6). For *Positive-Loss* games, low *Winning-percentages* are incompatible with the positive tones of the game notes, while *LAX-Score* averaged 47.4, which is much higher than the average of *Negative-Loss* games' scores. Moreover, *LAX-Score* has lower standard deviation values than *Winning-percentage* in general. This indicates that *LAX-Score* is more reliable and trustworthy than *Winning-percentage*. Therefore, we find that *LAX-Score* provides a more accurate reflection of real-world performance than *Winning-percentage*.

## 4.3 Statistical Benefits

Unlike existing metrics in other sports, *LAX-Score* takes into account environmental factors such as team rankings and game schedules. This gives the metric several statistical advantages.

*4.3.1 Linear function of Winning-Percentage.* Figures 5a and 5b illustrate that environmental variables make *LAX-Score* a linear function of winning-percentage. In Figure 5a, *LAX-Score* without environmental features is not intuitive for quantifying athletic performance, since there is no substantial difference in winning-percentage between games with *LAX-Scores* in the 60s and game with *LAX-Scores* in the 90s. However, Figure 5b shows that applying environmental variables flattens the curve so that *LAX-Score* is a linear function. For example, a *LAX-Score* in the 60s gives a team over a 60% chance of winning, while a *LAX-Score* in the 90s gives over a 90% chance of winning.




(a) Winning Pct. w/o Environmental Vars

(b) Winning Pct. w/ Environmental Vars

(c) Dist. w/o Environmental Vars

(d) Dist. w/ Environmental Vars

Fig. 5. Impacts of Environmental Variables in *LAX-Score*

*4.3.2 Gaussian Distribution.* Probability distribution analysis is essential in athletics because it enables a model to predict future performance [28]. Figures 5c and 5d indicate that *LAX-Score* also follows a Gaussian distribution with environmental factors. We recall that the mean *LAX-Score* of our dataset is 50.18, and the standard deviation is 17.18. In Figure 5d, about 70.38% of our *LAX-Scores* fall between $\mu \pm \sigma$ (33.0 and 67.36). Moreover, about 95% of *LAX-Scores* are within $\mu \pm 2\sigma$, and 100% of the games fall between $\mu \pm 3\sigma$. Thus, the distribution indeed follows the 3-sigma rule [29], which comes from the cumulative distribution function of the Gaussian distribution. In addition, the skewness of the distribution is $-0.03$, and the kurtosis is $-0.3$. These ranges are considered acceptable to prove the symmetricity and normality of the distribution [17]. Hence, a team can target a specific *LAX-Score* level more easily for their desired athletic performance. For example, in this *LAX-Score* distribution, $P(X \geq 59.2) = 0.30$, so a *LAX-Score* over 60 can be considered to fall in the top 30% of team-level performance, whereas targeting a *LAX-Score* of 68 is equivalent to targeting to the top 15% of performance since $P(X \geq 68) = 0.15$.

## 4.4 Feature Weights Validation

We claim that the *LAX-Score* metric and its weights are not overfitted to a specific season's data. To validate whether *LAX-Score* overfits a specific season, we collected conference-wide data from the last three seasons and calculated feature-weights and *LAX-Scores* for each of those seasons as well. At the beginning of each season, we use the last season's weights since the season weights are similar to each other. Then, those weights are updated

Table 7. Feature Weights and Statistical Outcomes for the last three seasons

| Season | Goals ($W_1$) | Goals Allowed ($W_2$) | Shots ($W_3$) | Shots Allowed ($W_4$) | Turnovers ($W_5$) | Caused Turnovers ($W_6$) | Time-off ($W_7$) | Ranking Difference ($W_8$) | Linear to Win Pct. ($R^2$) | Gaussian Distribution | Mean ±SD | Standard Error |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2019** | 27.13 | 26.03 | 19.37 | 17.91 | 4.34 | 0.93 | 1.28 | 16.87 | 0.96 | Yes | 50.18 ± 17.18 | 1.04 |
| **2018** | 29.43 | 28.27 | 21.48 | 18.45 | 4.11 | 0.96 | 1.34 | 18.11 | 0.96 | Yes | 50.33 ± 16.69 | 1.06 |
| **2017** | 26.74 | 25.69 | 20.11 | 15.67 | 4.71 | 1.05 | 1.32 | 13.8 | 0.93 | Yes | 51.22 ± 15.09 | 0.92 |

(a) X4, Catapult

(b) T31, Polar

(c) How to wear the devices

Fig. 6. Data Collection Devices — Each player wore X4 for IMU data and T31 for Heart Rate data.

every week as the season progresses. NCAA also updates its RPI values every week. Thus, we can use up-to-date RPI values and update the weights for calculating *LAX-Scores* during the season. Note that the derived feature weights can fluctuate at the beginning of the season but will eventually stabilize as the season proceeds.

Table 7 illustrates that the last three seasons' feature weights and statistical outcomes are similar to each other. These weights allow *LAX-Score* to have similar results over the seasons. For the 2018 season, we collected statistics from 231 games across two different conferences. In this season, the mean of the collected data is 50.33, the standard deviation is 16.69, and the standard error is 1.06. *LAX-Score* still follows a Gaussian distribution, and the coefficient of determination ($R^2$) [27] to the winning-percentage is 0.96, which shows the goodness of a fit model. Likewise, we collected data for 267 games from the 2017 season. In this season, the mean value is 51.22, the standard deviation is 15.09, and the standard error is 0.92. The $R^2$ value is still 0.93, and the distribution also follows the Gaussian distribution. In total, the *LAX-Scores* of 700 games across three conferences from the last three seasons fall between zero and 100, which provides athletes and coaches with intuitive insight into game performance. In addition, all of these seasons have the same statistical properties as *LAX-Score*.

## 5 DATA COLLECTION ON PRACTICE SESSIONS

Our ultimate goal of this paper is to enhance practice sessions and therefore win more games in the long run. To explore this connection, we collected practice data first. In Section 5.1, we describe data collection procedure from our athletes. We then demonstrate the collected sensing dataset in Section 5.2.

### 5.1 Methodology

To obtain the best practice features, we explored the entire 2019 season. Table 8 describes the summary of the 2019 season schedule. The William & Mary women's lacrosse team had 15 games in total. Each game day consists of a warm-up session followed by a game, which has two halves. Between the games, up to five different types of

Table 8. A Collegiate Lacrosse Team's 2019 Season Schedule — The team had 15 games and 35 practices

| | | Official Game | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **1st** | **2nd** | **3rd** | **4th** | **5th** | **6th** | **7th** | **8th** | **9th** | **Break** | **10th** | **11th** | **12th** | **13th** | **14th** | **15th** |
| **Practice Session** | **Recovery** | | | Day 10 | | | Day 25 | | | Day 37 | | | | | | | Day 70 |
| | **High-load** | Day 1 | | Day 12 | | | | | | Day 38 | Day 46 | Day 50 | | | | Day 64 | |
| | **Low-load** | | | | | Day 21 | | Day 29 | | Day 39 | Day 47 | Day 51 | | Day 59 | | Day 65 | |
| | **Taper** | Day 3 | | Day 13 | | Day 22 | | Day 30 | | Day 40 | | Day 52 | | | | | Day 71 |
| | **Potentiation** | Day 4 | Day 8 | | Day 17 | Day 23 | Day 26 | Day 31 | Day 34 | | | Day 53 | | Day 60 | | Day 67 | Day 72 |
| **Game day** | | Day 5 | Day 9 | Day 15 | Day 19 | Day 24 | Day 27 | Day 32 | Day 35 | Day 42 | | Day 54 | Day 56 | Day 61 | Day 63 | Day 68 | Day 73 |

practice sessions could take place, depending on scheduling. For ease of readability, we omit the opponents and the game dates but use relative scheduling in which the first practice for the first game is Day 1.

In the days leading up to a game, the team follows a sequence of sessions. First, in order to prepare for a game day, the team usually starts with a *High-load* session, where training intensity is as high as a game day. This practice helps athletes keep the same ability that they had at the beginning of the season. After one or two days of *High-load* sessions, *Low-load* sessions are conducted. This session has much less intensity than *High-load* as to not overload athletes. These two types of practice concepts have been proposed and studied for several decades in the field of Sports Science [23]. Then, *Taper* and *Potentiation* sessions can be performed a day or two before the game. The intensity of these two training sessions are similar to each other, but *Potentiation* is supposed to be shorter and has faster execution without fatigue than *Taper*. On a game day, the team plays against another collegiate lacrosse team. The athletes put forth the highest efforts on that day. After games, *Recovery* sessions are conducted if needed. On the *Recovery* days, the team typically utilizes inside recovery methods of extended warm-ups or corrections.

During the above schedule of games and practices, we collected IMU and heart rate data from the team's athletes. This procedure was approved by the William & Mary university IRB and conducted under the supervision of athletic coaches. Due to the need for confidentiality, the sensing dataset cannot be accessible to outside researchers. Throughout the entire season, every athlete wore the X4 device manufactured by Catapult for IMU data collection, as shown in Figure 6a. X4 consists of a 10Hz GPS, 100Hz tri-axial accelerometers, and 100Hz tri-axial gyroscopes. Heart rate data were also collected by the T31 device, which is compatible with X4, as shown in Figure 6b. Figure 6c displays how the athletes wore the devices. The size of T31 is 0.4"x0.4", and X4 is about 4"x2", which is small enough to carry. Thus, T31 can be worn at chest level, whereas X4 can be put in sports bras. After each activity, athletic coaches gathered the devices, and all the data including heart rate data were transferred to a local laptop via a USB cable connection. Then, the collected data were synchronized with a web-based application provided by Catapult. Through this web application, our researchers accessed metabolic data as well as workload data by date and by athlete. Note that these devices have been widely used for sports data collection [4]. In some sports, data collection using these products is even allowed in official matches [13]. Thus, this wide-use confirms that they have plenty of references for credibility.

### 5.2 Season-long Dataset

As shown in Table 8, the William & Mary team had 50 activities in the 2019 season, including 15 games and 35 practices. There are 16 players in total; we have six attackers, six midfielders, three defenders, and one goalkeeper who consented to the IRB data collection. Every athlete generates 56 sensing features on a given day. Hence, each activity consists of several hours of sensing data from the 16 athletes. Overall, we collected over 2,000 hours of data on 56 features in the 2019 season. The demographic data of the athletes are as follows. The age ranges are 18—23 with a mean of 20.3 and a standard deviation of 1.5. The height ranges are 5'1"—5'8" with a mean of 5'5" and a standard deviation of 2.1 inches. The weight ranges are 54—75 (kg) with a mean of 64.4 and a standard deviation of 6.6.

All the features of our data collection are described in Table 9 through 12, which are also our candidate features for our *LAX-Score* classification problems in Section 6. We briefly explain several features here, and more details are explained in the *Description* column of each table.

First of all, the duration and the distance of different acceleration runs were collected, as shown in Table 9. We call different ranges of metabolic data in *Band*. For example, acceleration data include eight bands in total where each band has a different range of acceleration, which we followed what the literature proposed [43]. This feature set suggests how much workload the athletes put into practices. For example, if the acceleration band 8 features (A15 and A16) are higher than usual, the athletes sprinted more on the given day.

Table 9. Acceleration-related Features

| Feature | | Description | Index |
|---|---|---|---|
| Acceleration Band 1 | Total Duration | Deceleration Time between -10 and -2.78 ($m/s^2$) | A1 |
| | Total Distance | Deceleration Distance between -10 and -2.78 ($m/s^2$) | A2 |
| Acceleration Band 2 | Total Duration | Deceleration Time between -2.78 and -1.47 ($m/s^2$) | A3 |
| | Total Distance | Deceleration Distance between -2.78 and -1.47 ($m/s^2$) | A4 |
| Acceleration Band 3 | Total Duration | Deceleration Time between -1.47 and -0.65 ($m/s^2$) | A5 |
| | Total Distance | Deceleration Distance between -1.47 and -0.65 ($m/s^2$) | A6 |
| Acceleration Band 4 | Total Duration | Deceleration Time between -0.65 and 0 ($m/s^2$) | A7 |
| | Total Distance | Deceleration Distance between -0.65 and 0 ($m/s^2$) | A8 |
| Acceleration Band 5 | Total Duration | Acceleration Time between 0 and 0.65 ($m/s^2$) | A9 |
| | Total Distance | Acceleration Distance between 0 and 0.65 ($m/s^2$) | A10 |
| Acceleration Band 6 | Total Duration | Acceleration Time between 0.65 and 1.47 ($m/s^2$) | A11 |
| | Total Distance | Acceleration Distance between 0.65 and 1.47 ($m/s^2$) | A12 |
| Acceleration Band 7 | Total Duration | Acceleration Time between 1.47 and 2.78 ($m/s^2$) | A13 |
| | Total Distance | Acceleration Distance between 1.47 and 2.78 ($m/s^2$) | A14 |
| Acceleration Band 8 | Total Duration | Acceleration Time between 2.78 and 10 ($m/s^2$) | A15 |
| | Total Distance | Acceleration Distance between 2.78 and 10 ($m/s^2$) | A16 |
| Accel/Decel Count | | The number of acceleration and deceleration runs | A17 |

Similarly, the duration and the distance of different velocity runs were also collected, as shown in Table 10. We also processed a high-speed running feature wherein an athlete is running at 75% of their maximal velocity or higher. To measure the maximal velocity of each athlete, we ascertained them at the beginning of the season. Like the acceleration features, the velocity features are highly related to the intensity of the practices. For instance, the high-speed features (V13 and V14) are mostly observed in *High-load* practices.

Furthermore, as illustrated in Table 11, heart rate data were collected to determine each athlete's time spent and distance run in each heart rate band. The literature suggested that heart rate data is another powerful factor to measure athletes' status and reveals fatigue levels of athlete [30, 42, 46, 55]. For example, higher values of the H19 feature suggest that they need more time to recover from the activity.

Lastly, we collected other features including *PlayerLoad*, as shown in Table 12. *PlayerLoad* is a widely used metric in sports science [22, 54], which was introduced by Catapult. This metric represents acceleration changes at any given time and thus intuitively shows the fatigue of each player, as calculated in Equation 5. In the equation, *fwd* denotes acceleration value of forward directions, while *side* represents sideway acceleration, *up* represents upwards, and $t$ denotes time.

Table 10. Velocity-related Features

| Feature | | Description | Index |
|---|---|---|---|
| Velocity Band 1 | Total Duration | Time between 0 and 20% of Maximal Velocity | V1 |
| | Total Distance | Distance between 0 and 20% of Maximal Velocity | V2 |
| Velocity Band 2 | Total Duration | Time between 20 and 50% of Maximal Velocity | V3 |
| | Total Distance | Distance between 20 and 50% of Maximal Velocity | V4 |
| Velocity Band 3 | Total Duration | Time between 50 and 75% of Maximal Velocity | V5 |
| | Total Distance | Distance between 50 and 75% of Maximal Velocity | V6 |
| Velocity Band 4 | Total Duration | Time between 75 and 90% of Maximal Velocity | V7 |
| | Total Distance | Distance between 75 and 90% of Maximal Velocity | V8 |
| Velocity Band 5 | Total Duration | Time over 90% of Maximal Velocity | V9 |
| | Total Distance | Distance over 90% of Maximal Velocity | V10 |
| Max Velocity | Value | Maximal Velocity (mph) in a game | V11 |
| | Percentage | Proportion of Maximal Velocity to the best record | V12 |
| High Speed Distance | | Distance of Velocity Bands 4 and 5 | V13 |
| High Speed Duration | | Duration of Velocity Bands 4 and 5 | V14 |

Table 11. Heart Rate-related Features

| Feature | | Description | Index |
|---|---|---|---|
| Heart Rate Band 1 | Total Duration | Time between 0 and 60% of Maximal HR | H1 |
| | Total Distance | Distance between 0 and 60% of Maximal HR | H2 |
| Heart Rate Band 2 | Total Duration | Time between 60 and 70% of Maximal HR | H3 |
| | Total Distance | Distance between 60 and 70% of Maximal HR | H4 |
| Heart Rate Band 3 | Total Duration | Time between 70 and 80% of Maximal HR | H5 |
| | Total Distance | Distance between 70 and 80% of Maximal HR | H6 |
| Heart Rate Band 4 | Total Duration | Time between 80 and 90% of Maximal HR | H7 |
| | Total Distance | Distance between 80 and 90% of Maximal HR | H8 |
| Heart Rate Band 5 | Total Duration | Time between 90 and 100% of Maximal HR | H9 |
| | Total Distance | Distance between 90 and 100% of Maximal HR | H10 |
| Heart Rate Band 6 | Total Duration | Time between 100 and 110% of Maximal HR | H11 |
| | Total Distance | Distance between 100 and 110% of Maximal HR | H12 |
| Heart Rate Band 7 | Total Duration | Time between 110 and 120% of Maximal HR | H13 |
| | Total Distance | Distance between 110 and 120% of Maximal HR | H14 |
| Heart Rate Band 8 | Total Duration | Time over 120% of Maximal HR | H15 |
| | Total Distance | Distance over 120% of Maximal HR | H16 |
| Max Heart Rate | Value | Maximal Heart Rate in a game | H17 |
| | Percentage | Proportion of Maximal Heart Rate to the best record | H18 |
| High Heart Rate Duration | | Duration of Heart Rate Band over 4 | H19 |
| Higher Heart Rate Duration | | Duration of Heart Rate Band over 5 | H20 |
| Heart Rate Exertion Per Minute | | Heart Rate Exertion per Minute | H21 |

Table 12. Miscellaneous Features

| Feature | Description | Index |
|---|---|---|
| Meterage Per Minute | Meterage per Minute | M1 |
| Exertion Index Per Minute | Exertion Index per Minute | M2 |
| Total PlayLoad | Total PlayLoad metric Load | M3 |
| Total Duration | Total Duration of Activity | M4 |

$$PlayerLoad = \sqrt{(fwd_{t+1} - fwd_t)^2 \ + \ (side_{t+1} - side_t)^2 \ + \ (up_{t+1} - up_t)^2} \tag{5}$$

All in all, each feature could influence an athlete's workload or fatigue level that eventually impacts athletic game performance. Therefore, athletes and coaches need to manage desired levels of the features more intelligently for performance enhancement.

## 6 CONNECTING PRACTICE WITH *LAX-SCORE*

Section 6.1 introduces experimental environments. In Section 6.2, we explore what practice features are correlated with high-performance games, which represent games with high LAX-Scores. In Section 6.3, our linear regression results demonstrate that there are causal relationships between a subset of practice features and *LAX-Scores*. Therefore, our feature analysis will help coaches and athletes to enhance their practice sessions more efficiently.

### 6.1 Experimental Environment

Our hypothesis is that effective practices will enable athletes to perform better on a game day and achieve a higher *LAX-Score* game. Thus, we first examine a binary classification problem, in which practice features are used to predict whether the *LAX-Score* of a game is over 60. Note that the top 30 colleges in the 2019 season achieved an average winning-percentage of 62% [53]. Moreover, a *LAX-Score* of 60 or higher indicates a performance in the top 28% under the Gaussian distribution. Therefore, 60 can be a respectable and reasonable *LAX-Score* to achieve

for most collegiate teams. Next, we conducted correlation analysis on each practice. These results demonstrate that a subset of the practice features are more correlated to games with a *LAX-Score* greater than 60.

In total, we evaluated four classification problems for the different types of practice: *High-load*, *Low-load*, *Taper*, and *Potentiation*. Note that we excluded *Recovery* sessions from this classification evaluation because there were only two sessions the team collected data from in the 2019 season. In each classification problem, a single input instance represents each athlete's 56 features per practice, as described in Section 5.2. According to the team's schedule in Table 8, the same types of practice were conducted multiple times during the same season. For example, the team had six *High-loads*, seven *Low-loads*, seven *Tapers*, and eleven *Potentiation* sessions. Hence, the number of instances for the *High-load* classification problem is $96(16 \times 6)$. Likewise, the number of instances for *Low-load*, *Taper*, and *Potentiation* are $112(16 \times 7)$, $112(16 \times 7)$, and $176(16 \times 11)$, respectively.

## 6.2 Correlation Results for Each Practice

We chose three different machine learning classifiers — Random Forest, Decision Tree, and Multilayer Perceptron (MLP). The authors in [10] conducted a thorough survey that covers machine learning studies in sport science. According to this survey, Artificial Neural Networks (ANN) is the most widely used machine learning algorithm in sports science, followed by Decision Trees. However, ANN does not explain the features of the network, so it cannot highlight actionable features to athletes in this study. To enhance athletic performance with our classification results, we also considered Random Forest and Decision Tree as our classifiers. Decision Tree provides feasible outcome features, which can be applied to athletic practices. Random Forest is an ensemble learning method for multiple decision trees. As a result, we chose MLP as an ANN classifier for comparison, Decision Tree, and Random Forest classifiers to get performance-enhancing features.

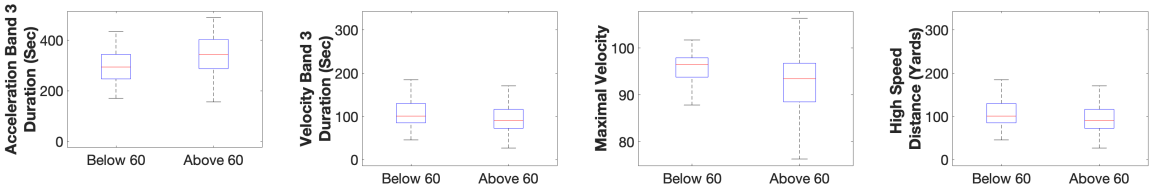
We set the maximum depth to 5 for Decision Tree, but this classifier might still fall prey to overfitting. For Random Forest, we let each tree choose 8 out of 56 features to avoid overfitting with 100 trees. For MLP, we set the learning rate to 0.1 and used 36 hidden units. We also used *Sigmoid* as an activation function. To avoid possible overfitting, we conducted a 10-fold cross-validation test for each classifier as follows: We combined practice sessions' instances and randomly split them into 10 subsets. Then, we trained nine subsets and tested on the remaining dataset. We repeated the same procedure 10 times to get a classification result.

Furthermore, we computed Pearson's correlation and its p-value for each practice feature, and *LAX-Scores* from the corresponding games. To validate the effectiveness of our analysis, we also compared the correlation between practice features and *LAX-Scores* with the correlation between practice features and game results.

*6.2.1 High-load Session.* *High-load* sessions are conducted with the highest intensity after athletes recover from a game. Those are sessions that emphasize speed of execution and utilize larger field sizes to train full-field

Table 13. Evaluation Results of *LAX-Score* Classification

| Practice Type | Classifier | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| High-load | Random Forest | 87.7 | 87.7 | 87.7 | 87.7 |
| | Decision Tree | 76.7 | 77.3 | 76.7 | 76.6 |
| | MLP | 84.3 | 83.7 | 84.3 | 83.9 |
| Low-load | Random Forest | 95.3 | 95.3 | 95.3 | 95.3 |
| | Decision Tree | 86.8 | 87.0 | 86.8 | 86.8 |
| | MLP | 96.2 | 96.5 | 96.2 | 96.2 |
| Taper | Random Forest | 96.2 | 96.6 | 96.2 | 96.2 |
| | Decision Tree | 85.0 | 85.2 | 85.0 | 85.0 |
| | MLP | 95.3 | 95.5 | 95.3 | 95.3 |
| Potentiation | Random Forest | 90.6 | 90.7 | 90.6 | 90.1 |
| | Decision Tree | 79.1 | 77.8 | 79.1 | 78.2 |
| | MLP | 88.7 | 88.9 | 88.7 | 88.3 |




(a) $A5$ ($r = 0.22$) (b) $V5$ ($r = -0.27$) (c) $V11$ ($r = -0.31$) (d) $V13$ ($r = -0.3$)

Fig. 7. Best Features for *High-load* Session where X-axes represent *LAX-Score* games.

concepts and accumulate high-speed distance and exposures to high-velocity changes of direction. According to the literature, if *High-load* is not intense enough, athletes will suffer from a decline in athletic performance over the season [23].

In Table 13, we calculate the accuracy of the classification problems. Given an athlete's instance, we predict whether it is likely to predict higher *LAX-Score* in the upcoming game. The accuracy of Random Forest is about 88%. Decision Tree achieves the worst results with an accuracy of 76%. For all the classifiers, *Precision, Recall* and *F1-Score* metrics also have similar values. This result indicates that the dataset is not overfitted despite of its small size.

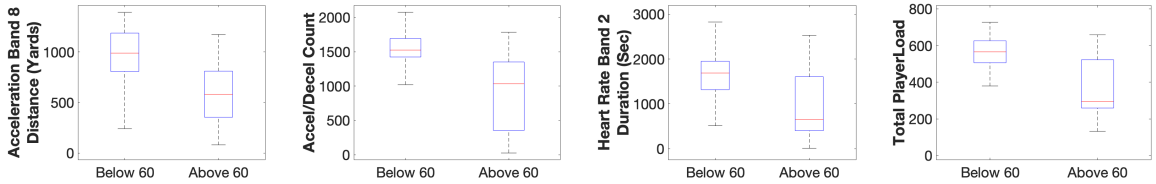
Furthermore, it appears that some practice features correlate more noticeably to high *LAX-Score* games than others. Figure 7 shows how those features constitute a *High-load* session. In Figure 7a, higher deceleration ($A5$) has a positive correlation with high performance. On the other hand, as shown in Figures 7b, 7c, and 7d, high-speed running ($V5, V11, V13$) are inversely correlated to higher *LAX-Scores*. Our overall analysis shows that there is a 'sweet spot' of intensity in high-load sessions: too little intensity decreases the effectiveness of the practice, whereas excessive intensity risks injury or exhaustion. Table 14 summarizes the top features of each practice that have higher correlation values. Out of 56 total features, there are 8 *High-load* features that are correlated with *LAX-Score* ($p < 0.05$). However, there are fewer significant correlation values in *High-load* sessions compared to other practices.

Table 14. Correlation Analysis between LAX-Scores and Sensing Features

| Practice Type | High-load | Low-load | Taper | Potentiation |
|---|---|---|---|---|
| **Top 5 Features** | -0.31 (V11) | -0.69 (A17) | -0.64 (V1) | 0.48 (A15) |
| | -0.3 (V13) | -0.53 (A16) | -0.61 (A17) | 0.38 (A16) |
| | -0.27 (V5) | -0.53 (A13) | -0.61 (M4) | -0.35 (A9) |
| | -0.26 (V8) | -0.51 (M3) | -0.58 (A9) | 0.33 (A3) |
| | 0.22 (A5) | -0.43 (H3) | -0.54 (A7) | -0.33 (V1) |
| **# Correlated Features (p-value < 0.05)** | 8 | 38 | 32 | 35 |
| **# Highly Correlated Features (p-value < 0.001)** | 0 | 11 | 12 | 9 |

Table 15. *High-load* Session Guideline — Features that may decrease *LAX-Score*

| **Feature values related to low *LAX-Scores*** | |
|---|---|
| (A9) Accel. Band 5 Duration < 4,500 | (A10) Accel. Band 7 Distance > 300 |
| (V1) Vel. Band 1 Duration < 6,800 | (V5) Vel. Band 3 Duration > 100 |
| (M3) Total Player Load > 500 | (M4) Total Duration < 7,500 |
| (H6) HR Band 3 Distance > 1,200 | (H8) HR Band 4 Distance > 1,500 |




(a) $A16$ ($r = -0.53$) (b) $A17$ ($r = -0.69$) (c) $H3$ ($r = -0.43$) (d) $M3$ ($r = -0.51$)

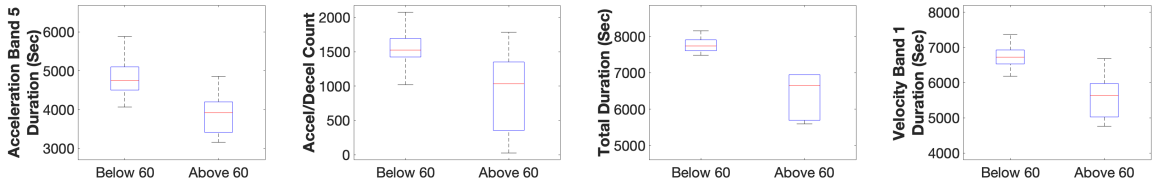
Fig. 8. Best *Low-load* Features that are correlated to *LAX-Score*

Based on this data, we provided actionable guidelines for each practice to the collegiate team studied after the 2019 season. Table 15 summarizes part of our suggested feature values for *High-load* practice sessions. Note that these are the feature values that may decrease team-level game performance. In general, the feature results suggest that an effective *High-load* session should exceed a minimum level of intensity, such as Features $V1, A9$, and $M4$. For example, the total duration of *High-load* activity ($M4$) is recommended to last longer than two hours. The coaches and athletes can now redesign their practice sessions based off of the reported outcomes.

*6.2.2 Low-load Session.* These are more metabolically intense sessions compared to *High-load* sessions. The field size for practice is small, so there are not many opportunities for high speeds or high-velocity changes of direction to occur, but this is still a solid training day with the intent to improve execution under fatigue.

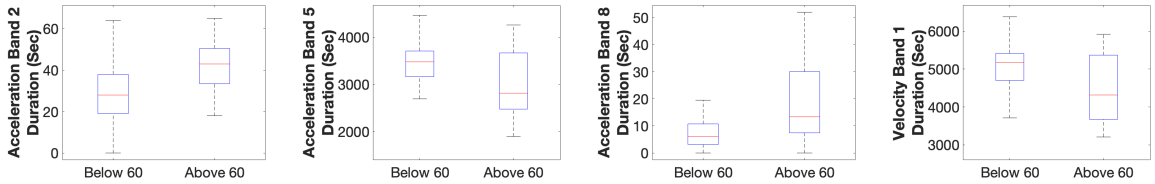
While Decision Tree had an accuracy of about 87%, Random Forest and MLP achieved pretty high results with an accuracy above 95%. These better results suggest that the *Low-load* session features are more correlated with *LAX-Score* than *High-load*. *Low-load* sessions also had similar results for other metrics.

Our results indicate that lower values of some practice features result in a higher *LAX-Score*, as illustrated in Table 14. Thus, it is desirable to set the limit of these features at certain thresholds. Otherwise, *Low-load* practices would negatively influence game performance. For example, Figure 8 displays the best *Low-load* features for games with a *LAX-Score* over 60. Figures 8a and 8b show that high-acceleration running ($A16, A17$) should be kept below a certain threshold. Figure 8c illustrates that running drills conducted at lower heart rates ($H3$) are correlated with high *LAX-Scores*. On the other hand, higher *PlayerLoad* ($M3$) may impact team-performance negatively, as shown in Figure 8d. Overall, this dataset demonstrates that restraining certain practice feature values during *Low-load* practices can be helpful. In *Low-load*, the highest correlation value between practice features and game results is about $-0.54$ ($A17$), which is still less than the highest correlation value between practice features and *LAX-Score* ($-0.69$).




(a) $A9$ ($r = -0.58$) (b) $A17$ ($r = -0.61$) (c) $M4$ ($r = -0.61$) (d) $V1$ ($r = -0.64$)

Fig. 9. Best *Taper* Features that are correlated to *LAX-Score*




(a) $A3$ ($r = 0.33$) (b) $A9$ ($r = -0.35$) (c) $A15$ ($r = 0.48$) (d) $V1$ ($r = -0.33$)

Fig. 10. Best *Potentiation* Features that are correlated to *LAX-Score*

*6.2.3 Taper Session.* Between the *Low-load* and *Potentiation* sessions, *Taper* is performed to facilitate recovery from previous *High/Low -load* training sessions, or games, depending upon the turnaround time from the previous game. *Taper* sessions are by far the lowest intensity session going into a game. This session is a day where the team uses a small field and walk or jog through tactics to clean up from work done on the *Load* sessions. The athletes do not run high-speed in this session.

As illustrated in Table 13, the overall accuracy of *Taper* sessions is quite high for all of the classifiers. Random Forest achieves an accuracy of 96.2%, whereas Decision Tree has about an 85% accuracy, and MLP has over the accuracy of 95%. Our results show that MLP and Random Forest do not have significant performance differences. Thus, Random Forest would be a preferred classifier because it can offer the importance of features towards performance enhancement.

In Table 14, we also observe that the characteristics of *Taper* are similar to *Low-load* sessions. When the coaches constrained features like *Total Duration* ($M4$), the team achieved higher *LAX-Scores*, as shown in Figure 9c. Figure 9 shows the top *Taper* features for high athletic performance. Thus, it is beneficial for coaches to place limits on *Low Acceleration* ($A9$), *Accel/Decel Count* ($A17$) or *Low Velocity* runnings ($V1$), as shown in Figures 9a, 9b, and 9d. In *Taper*, the most correlated feature between practice features and game results is $M2$ ($r = -0.37$), which is also much less than the highest correlation with *LAX-Score* ($-0.64$).

*6.2.4 Potentiation Session.* *Potentiation* is performed right before a game day. In this practice session, athletes play intensively to get ready for the game but not as much as a game day. According to the coaches, this session is supposed to be intense but short for better results. It is also important to look into the *Potentiation* dataset because the team conducted this practice prior to most of the games, as detailed in Table 8.

In terms of classification accuracy, Random Forest achieves an accuracy of 90.6%, and this classifier performs the best again in the *Potentiation* dataset. Decision Tree provides an accuracy of 79.1%, which is lower than other practice types. In MLP tests, the accuracy is about 89% with a learning rate of 0.1. Overall, MLP and Random forest still achieve the best results, whereas Random Forest even provides actionable feature items.

When we examine each feature, we still observe high correlation with *LAX-Score*. In Figure 10c, the athletes ran with high-acceleration ($A15$) longer before higher *LAX-Score* games. Figure 10a also suggests that more deceleration running ($A3$) is better for higher athletic performance during *Potentiation*. On the other hand, as shown in Figures 10b and 10d, the athletes ran shorter low-speed drills ($A9, V1$) before *LAX-Score* games over 60. These observations validate the purpose of *Potentiation*: it should be intense but short to achieve the greatest efficiency. In *Potentiation*, feature $A1$ has the highest correlation value with game results, which is only 0.23.

## 6.3 Causal Relationship between Practice Features and LAX-Score

In the previous subsection, we found that a subset of practice features is more correlated with LAX-Scores than the remaining practice features. However, correlation does not necessarily imply causation. If we find a causal

Table 16. Multivariable Linear Regression[e] Results

| Best Model | # Instances | R-squared | Adjusted R-squared | F-statistic | P-value |
|---|---|---|---|---|---|
| **High-load Model** | 96 | 0.120 | 0.100 | 5.870 | 0.004 [a] |
| **Low-load Model** | 112 | 0.630 | 0.623 | 86.127 | $< 0.001$ [b] |
| **Taper Model** | 112 | 0.316 | 0.309 | 48.020 | $< 0.001$ [c] |
| **Potentiation Model** | 176 | 0.154 | 0.139 | 10.464 | $< 0.001$ [d] |

[a] Predictors: (constant), V11, and A5 features
[b] Predictors: (constant), A17, and M3 features
[c] Predictors: (constant) and M4 features
[d] Predictors: (constant), V1, A3, and A15 features
[e] Dependent Variable: *LAX-Score*

relationship between practice features and the LAX-Score metric, our research can help coaches better instruct athletes to improve their game performance.

To address the causality between practice features and the LAX-Score, we performed multivariable linear regressions on each practice to examine which practice features affect game performance. Since the unit of measurement of each practice feature is different, we first normalized the data with min/max values from all practices. In order to select the best model for each practice, we used the IBM SPSS Statistics [37] with the top five features that have a high correlation with *LAX-Scores*, which were introduced in Table 14. In other words, the five normalized features are candidate predictors, and the *LAX-Score* is the target variable. Overall, our null and test hypotheses for the four practice sessions are as follows:

- $H_0$ : Practice features do not have an effect on game performance (*LAX-Score*).
- $H_1$ : Practice features have an effect on game performance (*LAX-Score*).

Table 16 lists the best model for each practice. In our study, the best model is a model that has the highest adjusted $R^2$ value among candidate models for each practice. Note that a higher adjusted $R^2$ value represents a stronger model that can explain a causal relationship between practice features and *LAX-Scores*. Our results

Table 17. Coefficient Results[a] of the Multivariable Linear Regressions

| Model | Feature | Unstandardized Coefficient | | Standardized Coefficient | t-statistic | P-value |
|---|---|---|---|---|---|---|
| | | B | Std. Error | $\beta$ | | |
| **High-load** | (Intercept) | 58.803 | 3.345 | | 17.582 | $< 0.001$ |
| | V11 | -10.455 | 4.017 | -0.264 | -2.603 | 0.011 |
| | A5 | 6.967 | 3.355 | 0.212 | 2.208 | 0.040 |
| **Low-load** | (Intercept) | 66.351 | 0.657 | | 28.814 | $< 0.001$ |
| | A17 | -21.347 | 3.222 | -0.517 | -6.626 | $< 0.001$ |
| | M3 | -12.499 | 3.443 | -0.283 | -3.631 | $< 0.001$ |
| **Taper** | (Intercept) | 74.697 | 0.907 | | 22.405 | $< 0.001$ |
| | M4 | -37.53 | 5.416 | -0.562 | -6.93 | $< 0.001$ |
| **Potentiation** | (Intercept) | 48.972 | 1.070 | | 45.772 | $< 0.001$ |
| | A15 | 14.554 | 6.270 | 0.189 | 2.321 | 0.021 |
| | A3 | 13.489 | 5.630 | 0.193 | 2.396 | 0.018 |
| | V1 | -14.202 | 5.872 | -0.171 | -2.419 | 0.017 |

[a] Dependent Variable: *LAX-Score*

demonstrate that all the selected models are valid to provide statistically significant data for each practice (P-value from F-statistic $< 0.05$). Thus, we rejected our null hypotheses and examined more detailed results.

First, we observe that all the adjusted $R^2$ values are close to the original $R^2$ values; thus, we confirm that the number of selected features and instances is not overfitting. Among the four models, the *Low-load* model performs the best; this model can explain 62.3% of the variance of the *LAX-Scores* from the selected features (Adjusted $R^2 = 0.623$). In addition, the *Taper* session model has roughly 30% explanatory power (Adjusted $R^2 = 0.309$). However, the other adjusted $R^2$ values are relatively small (Adjusted $R^2 < 0.3$), which indicate weaker causal relationships.

Next, Table 17 illustrates the coefficient results of the best models for *LAX-Score* prediction. The intercept values are considered as base points, and the selected features can affect the *LAX-Score* models. Since we already normalized the input feature values, the unstandardized and standardized coefficient results do not have any significant difference. In the *Low-load* model, we find that features $A17$ (Accel/Decel count) and $M3$ (PlayerLoad) are statistically significant since the P-values from the t-statistic are less than 0.001. Moreover, the coefficient values are negative; thus, it infers that lower feature values affect *LAX-Scores* positively. This observation is consistent with the correlation analysis results presented in Figure 8. Likewise, in *Taper* sessions, feature $M4$ (Total Duration) is statistically significant (P-value $< 0.001$). In addition, higher values affect game performance negatively since the coefficient value of $M4$ is negative (-37.53). This result is also consistent with the correlation results presented in Figure 9. On the contrary, for *High-load* and *Potentiation* sessions, although there are several variables that are statistically significant (P-value from t-statistic $< 0.05$), the models do not explain much about *LAX-Scores* (Adjusted $R^2 < 0.30$).

In conclusion, our results demonstrate that the *Low-load* and *Taper* sessions have stronger causal relationships than the *High-load* and *Potentiation* sessions.

## 7 DISCUSSION

In this section, we share our thoughts regarding implementation, limitations, and future work.

### 7.1 Implementation

Our formulation with custom features shows promise for adaptation to real-world athletes. In particular, our physiological observations about practice types are consistent with numerous theories of high-low planning in sports science [23]. At the beginning of the 2020 season, we provided the practice guidelines based on our classification results to the same team that provided the data for this study. Before we shared our recommendations, the team's coaches had required the team to practice at a greater volume than our observation would later show is most effective and consequently seen an uptick in injuries. As demonstrated in Section 6.2, overshooting practices can negatively impact in-game athletic performance and eventually lead to serious injuries. Thus, the coaches applied our actionable outcomes to prevent practice-session injuries by adjusting the intensity of *High-load* and *Low-load*.

Unfortunately, due to the COVID-19 pandemic, NCAA suspended the women's lacrosse season in March 2020, and at the time of writing, it has not resumed. For this reason, we were unable to see more meaningful results for this season. Nevertheless, we are eager to apply our outcomes to the team next season and utilize their feedback in our future work.

### 7.2 Limitation

One possible limitation of our research is the size of the sensing dataset. Our dataset was limited to data gathered from a collegiate team of 16 athletes over the course of a single season. Thus, it is necessary to collect more datasets for more in-depth research. For example, game performance can be influenced by sequencing in types of

practice sessions. If we collect enough datasets in the future, it is also worth looking into the impacts of different practice combinations. Still, collaboration with other domain experts is challenging especially in collegiate sports. This is because data collection is limited to certain times of the year and thus time-consuming for both researchers and athletes. Despite the above limitations, we were still effective on a team scale. In future work, more techniques such as data augmentation can be used to satisfy the amount of data.

Furthermore, although we presume that highly effective practice leads to better team performance in general, it should be noted that the high correlation between certain practice features and high *LAX-Score* does not mean that those features' presence in a specific session guarantee higher game performance. In fact, the inherent nature of the sport as a complex system makes it hard to focus on causality. To determine better causality, the practice stimulus would have to be removed entirely and the team would have to only play games during the competitive season, which is hard during collegiate seasons. Thus, more thorough studies can be conducted with confounding factors that might impact the performance, such as an athlete's sleep quality or academic schedule. Nevertheless, the literature implicitly accepts the cause-and-effect relationship between training and performance [52]. From this perspective, our observations demonstrate a trend of important practice features that may yield high athletic performance.

### 7.3 Future Work

We believe that the machine learning and computer science community can contribute more to sports science. Specifically, Natural Language Processing (NLP) could help validate athletic performance from official game-notes. When we analyzed the overall tone of the game-notes, we also considered using a commonly-used sentiment analysis platform, MonkeyLearn [38]. This pre-trained tool analyzes text tones to determine whether the given text sounds positive or negative [60] [70] [1]. However, the accuracy of this tool against the coaches' labeling was only 57.6%. This suggests that there is still a need for a domain-specific classifier. This kind of classifier requires labeling the game-notes from athletic experts for model training. In the future, we are planning to build an athletics-specific classifier to interpret game-notes. In this way, we can further validate *LAX-Score* with the game-notes more easily.

Our approaches are also applicable to other team-sports even though we focused on lacrosse in this paper. As shown in Figure 2, we have exploited public statistics to evaluate team-performance and enhance practices via sensing features and ML techniques. However, the most important lesson learned is that it is crucial to secure a meaningful dataset. Thus, we propose to collect a conference-wide or region-wide dataset in order to enhance athletic performance. Those datasets could tremendously improve athletic performance in many aspects, such as in other studies [2, 19].

Furthermore, there are several possible research areas in this field besides performance enhancement. For example, a player substitution recommendation system could be a potential research project [77]. Based on real-time passive sensing features, ML could recommend which player needs to be substituted at a given time. It requires an online prediction system and can be done in real-time during games. Moreover, research on injury prevention can be another major topic. Identifying injury-prone activities or athletes will benefit athletic performance in the long run. Even though many athletes have used IMU sensors, the literature mostly applies legacy machine learning techniques [15]. Overall, we believe that increased collaboration between computer scientists and sports scientists can contribute immensely.

## 8 CONCLUSION

In this paper, we aimed to enhance athletic performance on an NCAA Division I women's lacrosse team. Since there is no strong ground truth for evaluating athletic performance, we first leveraged machine learning feature selection to devise a new scoring metric called *LAX-Score* in order to capture game performance. The proposed

metric takes into account both game statistics and environmental factors in measuring team performance. We statistically validated *LAX-Score* on over 700 official NCAA game statistics for 50 collegiate teams. Our results demonstrated that *LAX-Score* is a linear function of the winning percentage and follows the Gaussian distribution.

To draw connections between high game performance and practice features, we collected an IMU/biometric dataset from a collegiate team throughout the 2019 season. Then, we investigated the correlation between sensing data features and games with a *LAX-Score* greater than 60. The correlation analysis suggested that a subset of the features are more related to high athletic performance than the rest. Moreover, machine learning classifiers predict high *LAX-Score* games with a maximum accuracy of 96%. Finally, we provided actionable recommendations to the collegiate team based on our observations to yield better performance. We were unable to directly observe effective performance improvement due to the COVID-19 pandemic's effects on collegiate sports; however, our results do show promise in preventing injuries in practice sessions.

## ACKNOWLEDGMENTS

The authors would like to thank all the athletes who participated in this study. The authors would also like to thank the anonymous reviewers for their valuable comments and helpful suggestions.

## REFERENCES

[1] Nur Sakinah Diyanah Abdullah and Izzal Asnira Zolkepli. 2017. Sentiment Analysis of Online Crowd Input towards Brand Provocation in Facebook, Twitter, and Instagram. *Proceedings of the International Conference on Big Data and Internet of Thing (BDIOT)* (2017), 67–74. https://doi.org/10.1145/3175684.3175689

[2] Mahboubeh Ahmadalinezhad, Masoud Makrehchi, and Neil Seward. 2019. Basketball lineup performance prediction using network analysis. In *ASONAM '19: International Conference on Advances in Social Networks Analysis and Mining*. ACM, New York, NY, USA, 519–524. https://doi.org/10.1145/3341161.3342932

[3] Kei Akiyama, Takaya Sasaki, and Masahiro Mashiko. 2019. Elite Male Lacrosse Players' Match Activity Profile. *Journal of Sports Science and Medicine* 18, 2 (2019), 290–294. https://www.jssm.org/2jssm-18-290.xml

[4] Alf Alderson. 2019. Making Rugby Safer | E&T Magazine.

[5] Douglas G Altman and J Martin Bland. 2005. Standard deviations and standard errors. *BMJ (Clinical research ed.)* 331, 7521 (2005), 903. https://doi.org/10.1136/bmj.331.7521.903

[6] Jonathan D Bartlett, Fergus O'Connor, Nathan Pitchford, Lorena Torres-Ronda, and Samuel J Robertson. 2017. Relationships Between Internal and External Training Load in Team-Sport Athletes: Evidence for an Individualized Approach. *International Journal of Sports Physiology and Performance* 12, 2 (Feb. 2017), 230–234. https://doi.org/10.1123/ijspp.2015-0791

[7] Andres Mauricio Cifuentes Bernal, Robin Alfonzo Blanco Cañon, and Mauricio Plaza Torres. 2017. An Rn Surface to Describe Sport Performance in Women's Basketball Using Kohonen Nets on ECG Signals. In *Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics (ICCBB 2017)*. Association for Computing Machinery, New York, NY, USA, 24–28. https://doi.org/10.1145/3155077.3155082

[8] Tim Op De Beéck, Arne Jaspers, Michel S. Brink, Wouter G.P. Frencken, Filip Staes, Jesse J. Davis, and Werner F. Helsen. 2019. Predicting Future Perceived Wellness in Professional Soccer: The Role of Preceding Load and Wellness. *International Journal of Sports Physiology and Performance* 14, 8 (Jan. 2019), 1–25. https://doi.org/10.1123/ijspp.2017-0864

[9] Bill Briggs. 2019. How machine learning is unlocking the secrets of human movement – and reshaping pro sports. https://news.microsoft.com/transform/machine-learning-unlocking-secrets-human-movement-reshaping-pro-sports/

[10] Rory Bunker and Teo Susnjak. 2019. The Application of Machine Learning Techniques for Predicting Results in Team Sport: A Review. *arXiv.org* (Dec. 2019). arXiv:cs.LG/1912.11762v1 https://arxiv.org/abs/1912.11762

[11] David L Carey, Kok-Leong Ong, Meg E Morris, Justin Crow, and Kay M Crossley. 2016. Predicting ratings of perceived exertion in Australian football players - methods for live estimation. *International Journal of Computer Science in Sport* 15, 2 (2016), 64–77. https://doi.org/10.1515/ijcss-2016-0005

[12] David L Carey, Kok-Leong Ong, R Whiteley, Kay M Crossley, Justin Crow, and Meg E Morris. 2017. Predictive Modelling of Training Loads and Injury in Australian Football. *International Journal of Computer Science in Sport* 17, 1 (June 2017), 49–66. https://doi.org/10.2478/ijcss-2018-0002

[13] Catapult.com. 2018. Catapult Technologies become FIFA-approved Wearable Tracking Devices. https://www.catapultsports.com/blog/fifa-approved-wearable-tracking

[14] Ryan M Chambers, Tim J Gabbett, Ritu Gupta, Casey Josman, Rhodri Bown, Paul Stridgeon, and Michael H Cole. 2019. Automatic detection of one-on-one tackles and ruck events using microtechnology in rugby union. *Journal of Science and Medicine in Sport* (Jan. 2019). https://doi.org/10.1016/j.jsams.2019.01.001

[15] João Gustavo Claudino, Daniel de Oliveira Capanema, Thiago Vieira de Souza, Julio Cerca Serrão, Adriano C. Machado Pereira, and George P. Nassis. 2019. Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports Med - Open* 5, 28 (2019). https://doi.org/10.1186/s40798-019-0202-3

[16] Gabor Csataljay, Peter O'Donoghue, Mike Hughes, and Henriette Dancs. 2017. Performance indicators that distinguish winning and losing teams in basketball. *International Journal of Performance Analysis in Sport* 9, 1 (April 2017), 60–66. https://doi.org/10.1080/24748668.2009.11868464

[17] Paul Mallery Darren George. 2010. *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 Update* (10th. ed.). Allyn & Bacon.

[18] Han Ding, Longfei Shangguan, Zheng Yang, Jinsong Han, Zimu Zhou, Panlong Yang, Wei Xi, and Jizhong Zhao. 2015. FEMO: A Platform for Free-weight Exercise Monitoring with RFIDs. *SenSys '15: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, 141–154. https://doi.org/10.1145/2809695.2809708

[19] Jordi Duch, Joshua S. Waitzman, and Luís A. Nunes Amaral. 2010. Quantifying the Performance of Individual Players in a Team Activity. *PLoS ONE* 5, 6 (June 2010), 1–7. https://doi.org/10.1371/journal.pone.0010937

[20] Sharon R. Rana Emily A. Enemark-Miller, Jeff G. Seegmiller. 2009. Physiological profile of women's Lacrosse players. *Journal of Strength and Conditioning Research* 23, 1 (2009), 39–43. https://doi.org/10.1519/JSC.0b013e318185f07c

[21] Peter C Fishburn. 1967. Additive utilities with incomplete product set: Applications to priorities and assignments. *Operations Research* 15, 3 (1967), 537–542. https://www.jstor.org/stable/168461

[22] Jordan L Fox, Robert Stanton, and Aaron T Scanlan. 2018. A Comparison of Training and Competition Demands in Semiprofessional Male Basketball Players. *Research Quarterly for Exercise and Sport* 89, 1 (Jan. 2018), 103–111. https://doi.org/10.1080/02701367.2017.1410693

[23] Charlie Francis. 1992. *Charlie Francis Training System.* TBLI Publications.

[24] Eibe Frank, Mark A. Hall, and Ian H. Witten. 2016. Data Mining: Practical Machine Learning Tools and Techniques. https://www.cs.waikato.ac.nz/ml/weka/

[25] Joel M Garrett, Ian McKeown, Darren J Burgess, Carl T Woods, and Roger G Eston. 2017. A preliminary investigation into the discriminant and ecological validity of the athletic ability assessment in elite Australian rules football:. *International Journal of Sports Science and Coaching* 13, 5 (Oct. 2017), 679–686. https://doi.org/10.1177/1747954117736168

[26] Jeremy Gentles, Christine Coniglio, Matthew Besemer, Joshua Morgan, and Michael Mahnken. 2018. The Demands of a Women's College Soccer Season. *Sports* 6, 1 (March 2018), 16–11. https://doi.org/10.3390/sports6010016

[27] S A Glantz, B K Slinker, and T B Neilands. 1990. Primer of Applied Regression and Analysis of Variance. McGraw-Hill.

[28] Brian Godsey. 2014. Comparing and Forecasting Performances in Different Events of Athletics Using a Probabilistic Model. *arXiv.org* 2 (Aug. 2014). arXiv:stat.AP/1408.5924v1

[29] E W Grafarend. 2006. Linear and nonlinear models: fixed effects, random effects, and mixed models.

[30] Ben Greenfield. 2013. How Heart Rate Zones Work. https://bengreenfieldfitness.com/article/fitness-articles/how-heart-rate-zones-work/

[31] Benjamin H Groh, Frank Warschun, Martin Deininger, Thomas Kautz, Christine Martindale, and Björn Eskofier. 2017. Automated Ski Velocity and Jump Length Determination in Ski Jumping Based on Unobtrusive and Wearable Sensors. *IMWUT* 1, 3 (2017), 1–17. https://doi.org/10.1145/3130918

[32] Mahmoud Hassan, Florian Daiber, Frederik Wiehr, Felix Kosmalla, and Antonio Krüger. 2017. FootStriker: An EMS-based Foot Strike Assistant for Running. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 1 (March 2017), 2–18. https://doi.org/10.1145/3053332

[33] Richard Hauer, Antonio Tessitore, Nicole Binder, and Harald Tschan. 2018. Physiological, perceptual, and technical responses to continuous and intermittent small-sided games in lacrosse players. *PLoS ONE* 13, 10 (Oct. 2018), e0203832–14. https://doi.org/10.1371/journal.pone.0203832

[34] Jay R. Hoffman, Nicholas A. Ratamess, Kate L. Neese, Ryan E. Rose, Jie Kang, Jason F. Magrelli, and Avery D. Faigenbaum. 2019. Physical Performance Characteristics in NCAA Division III Champion Female Lacrosse Athletes. 23, 5 (Sept. 2019), 1524–1529. https://doi.org/10.1519/JSC.0b013e3181b3391d

[35] John Hollinger. 2005. Game Score. https://en.wikipedia.org/wiki/John_Hollinger

[36] John Hollinger. 2005. PER. https://www.basketball-reference.com/about/per.html

[37] IBM. 2009. SPSS (Statistical Package for the Social Sciences) Statistics. https://www.ibm.com/analytics/spss-statistics-software

[38] MonkeyLearn Inc. 2013. MonkeyLearn - TextAnalysis. https://monkeylearn.com

[39] Sushma Jain and Harmandeep Kaur. 2017. Machine learning approaches to predict basketball game outcome. *2017 3rd International Conference on Advances in Computing,Communication & Automation (ICACCA) (Fall)* (2017), 1–7. https://doi.org/10.1109/ICACCAF.2017.8344688

[40] B James. 1988. The Bill James historical baseball abstract. https://www.mlb.com/glossary/advanced-stats/game-score

[41] Arne Jaspers, Tim Op De Beéck, Michel S Brink, Wouter G P Frencken, Filip Staes, Jesse J Davis, and Werner F Helsen. 2018. Relationships Between the External and Internal Training Load in Professional Soccer: What Can We Learn From Machine Learning? *International Journal of Sports Physiology and Performance* 13, 5 (May 2018), 625–630. https://doi.org/10.1123/ijspp.2017-0299

[42] Bob Kaleal. 2016. Heart Rate Training and the Science Behind its Success. https://www.ptonthenet.com/articles/heart-rate-training-and-the-science-behind-its-success-4053

[43] Stephen J Kelly, Mark L Watsford, Michael J Rennie, Rob W Spurrs, Damien Austin, and Matthew J Pine. 2019. Match-play movement and metabolic power demands of elite youth, sub-elite and elite senior Australian footballers. *PLoS ONE* 14, 2 (Feb. 2019), e0212047–10. https://doi.org/10.1371/journal.pone.0212047

[44] Aftab Khan, James Nicholson, and Thomas Plötz. 2017. Activity Recognition for Quality Assessment of Batting Shots in Cricket using a Hierarchical Representation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sept. 2017), 1–31. https://doi.org/10.1145/3130927

[45] Bridgett Lynn Klemz. 2014. Lacrosse: Biomechanics, Injuries, Prevention and Rehabilitation. (April 2014), 1–63. http://purl.flvc.org/fgcu/fd/Klemz_fgcu_1743_10051

[46] Kyle D Peterson . 2018. Resting Heart Rate Variability Can Predict Track and Field Sprint Performance. *OA Journal - Sports* 1 (Sept. 2018), 1–9. https://doi.org/10.24294/sp.v1i1.159

[47] Cassim Ladha, Nils Y Hammerla, Patrick Olivier, and Thomas Plötz. 2013. ClimbAX - skill assessment for climbing enthusiasts. *UbiComp '13: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (2013), 235–244. https://doi.org/10.1145/2493432.2493492

[48] Yongjun Li, Lizheng Wang, and Feng Li. 2021. A data-driven prediction approach for sports team performance and its application to National Basketball Association. *Omega* 98 (2021), 102123. https://doi.org/10.1016/j.omega.2019.102123

[49] Robert G Lockie, Samantha A Birmingham-Babauta, John J Stokes, Tricia M Liu, Fabrice G Risso, Adrina Lazar, Dominic V Giuliano, Ashley J Orjalo, Matthew R Moreno, Alyssa A Stage, and DeShaun L Davis. 2018. An Analysis of Collegiate Club-Sport Female Lacrosse Players: Sport-Specific Field Test Performance and the Influence of Lacrosse Stick Carrying. *International Journal of Exercise Science* 11, 4 (Jan. 2018), 269–280.

[50] Cliff Mallett and Jean Côté. 2006. Beyond Winning and Losing: Guidelines for Evaluating High Performance Coaches. *The Sport Psychologist* 20, 2 (June 2006), 213–221. https://doi.org/10.1123/tsp.20.2.213

[51] Nazanin Mehrasa, Yatao Zhong, Frederick Tung, Luke Bornn, and Greg Mori. 2018. Deep learning of player trajectory representations for team activity analysis. *11th MIT Sloan Sports Analytics Conference* (2018).

[52] Iñigo Mujika. 2017. Quantification of training and competition loads in endurance sports: methods and applications. *The International Journal of Sports Physiology and Performance (IJSPP)* (2017). https://doi.org/10.1123/ijspp.2016-0403

[53] NCAA. 2020. D1 Women's College Lacrosse. https://www.ncaa.com/sports/lacrosse-women/d1

[54] Daniel P Nicolella, Lorena Torres-Ronda, Kase J Saylor, and Xavi Schelling. 2018. Validity and reliability of an accelerometer-based player tracking device. *PLoS ONE* 13, 2 (Feb. 2018), e0191823–13. https://doi.org/10.1371/journal.pone.0191823

[55] Kyle D Peterson. 2018. Recurrent Neural Network to Forecast Sprint Performance. *Applied Artificial Intelligence* 32, 7-8 (Oct. 2018), 692–706. https://doi.org/10.1080/08839514.2018.1505214

[56] Steven S. Plisk. 1994. Regression Analyses of NCAA Division I Final Four Men's Lacrosse Competition. *Journal of Strength and Conditioning Research* 8, 1 (Aug. 1994), 28–42.

[57] Chris S. Polley, Stuart J. Cormack, Tim J. Gabbett, and Ted Polglaze. 2019. Activity Profile of High Level Australian Lacrosse. *Journal of Strength and Conditioning Research* 29, 1 (June 2019), 126–136. https://doi.org/10.1519/JSC.0000000000000599

[58] JR Quinlan. 1986. Induction of decision trees. *Machine Learning* 1 (1986), 81–106. https://doi.org/10.1007/BF00116251

[59] Alen Rajšp and Jr. Izdok Fister. 2020. A Systematic Literature Review of Intelligent Data Analysis Methods for Smart Sport Training. *Applied Sciences* 10, 9 (2020). https://doi.org/10.3390/app10093013

[60] C L Rakshitha and S Gowrishankar. 2018. Machine Learning based Analysis of Twitter Data to Determine a Person's Mental Health Intuitive Wellbeing. *International Journal of Applied Engineering Research* 13, 21 (2018). http://www.ripublication.com/ijaer18/ijaerv13n21_20.pdf

[61] Aaron M Randolph. 2012. Analysis of the Effectiveness of a Preseason Strength and Conditioning Program for Collegiate Men's and Women's Lacrosse. (Dec. 2012), 1–208.

[62] Samuel Ryan, Aaron J Coutts, Joel Hocking, and Thomas Kempton. 2017. Factors Affecting Match Running Performance in Professional Australian Football. *International Journal of Sports Physiology and Performance* 12, 9 (Oct. 2017), 1199–1204. https://doi.org/10.1123/ijspp.2016-0586

[63] Patrick Schober, Christa Boer, and Lothar A Schwarte. 2018. Correlation coefficients: appropriate use and interpretation. *Anesth Analg.* 126, 5 (2018), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

[64] Steven R Schultze and Christian-Mathias Wellbrock. 2018. A weighted plus/minus metric for individual soccer player performance. *Journal of Sports Analytics* 4, 2 (March 2018), 121–131. https://doi.org/10.3233/JSA-170225

[65] Katie M. Sell, James M. Prnedergast, Jamie J. Ghigiarelli, Adam M. Gonzalez, Lauren M. Biscardi, Adam R. Jajtner, and Alexander S. Rothstein. 2018. Comparison of Physical Fitness Parameters For Starters vs. Nonstarters in an NCAA DIVISION I MEN's Lacrosse Team . 32, 11 (Nov. 2018), 3160–3168. https://doi.org/10.1519/JSC.0000000000002830

[66] Shweta Srivastava. 2014. Weka: a tool for data preprocessing, classification, ensemble, clustering and association rule mining. *International Journal of Computer Applications* 88, 10 (2014). https://doi.org/10.5120/15389-3809

[67] Michelle R. Steinhagen, Michael C. Meyers, Howard H. Erickson, Larry Noble, and Melanie T. Richardson. 1998. Physiological Profile of College Club Sport Lacrosse Athletes. *Journal of Strength and Conditioning Research* 12, 4 (Aug. 1998), 226–231. https://doi.org/10.1519/00124278-199811000-00004

[68] Jason D Vescovi, Todd D Brown, and Teena M Murray. 2007. Descriptive characteristics of NCAA Division I women lacrosse players. *Journal of Science and Medicine in Sport* 10, 5 (Oct. 2007), 334–340. https://doi.org/10.1016/j.jsams.2006.07.010

[69] Heather K Vincent, Laura Ann Zdziarski, and Kevin R Vincent. 2015. Review of Lacrosse-Related Musculoskeletal Injuries in High School and Collegiate Players. *Sports Health: A Multidisciplinary Approach* 7, 5 (Aug. 2015), 448–451. https://doi.org/10.1177/1941738114552990

[70] Zhengkui Wang, Guangdong Bai, Soumyadeb Chowdhury, Quanqing Xu, and Zhi Lin Seow. 2017. TwiInsight: Discovering Topics and Sentiments from Social Media Datasets. *arXiv.org* (May 2017). arXiv:cs.IR/1705.08094v1 http://arxiv.org/abs/1705.08094v1

[71] Aaron Wellman, Sam Coad, Grant Goulet, and Christopher McLellan. 2016. Quantification of competitive game demands of NCAA division I college football players using global positioning systems. *Journal of Strength and Conditioning Research* 30, 1 (2016), 11–19. https://doi.org/10.1519/JSC.0000000000001206

[72] Wikipedia. 2004. Rating percentage index - Wikipedia. https://en.wikipedia.org/wiki/Rating_percentage_index

[73] Wikipedia. 2005. Winning Percentage - Wikipedia. https://en.wikipedia.org/wiki/Winning_percentage

[74] Wikipedia. 2014. Grading systems by country - Wikipedia. https://en.wikipedia.org/wiki/Grading_systems_by_country

[75] Ian H. Witten, Eibe Frank, Leonard E. Trigg, Mark A. Hall, Geoffrey Holmes, and Sally Jo Cunningham. 1999. Weka: Practical machine learning tools and techniques with Java implementations. *Hamilton, New Zealand: University of Waikato, Department of Computer Science* (Nov. 1999).

[76] Daniel WT Wundersitz, Casey Josmanand Ritu Gupta, Kevin J Netto, Paul B Gastin, and Sam Robertson. 2015. Classification of team sport activities using a single wearable tracking device. *Journal of Biomechanics* 48, 15 (2015), 3975–3981. https://doi.org/10.1016/j.jbiomech.2015.09.015

[77] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. 2019. Deep Learning Based Recommender System - A Survey and New Perspectives. *ACM Comput. Surv.* 52, 1 (2019), 1–38. https://doi.org/10.1145/3285029

[78] Hongyang Zhao, Shuangquan Wang, Gang Zhou, and Woosub Jung. 2019. TennisEye - tennis ball speed estimation using a racket-mounted motion sensor. *IPSN* (2019), 241–252. https://doi.org/10.1145/3302506.3310404