PERFORMANCE MONITORING OF SOCCER PLAYERS USING PHYSIOLOGICAL SIGNALS AND PREDICTIVE ANALYTICS

Sampurna Ravindranathan, Hossein Davari Ardakani, Andrew Pimental, and Jay Lee
NSF I/UCRC for Intelligent Maintenance Systems (IMS)
University of Cincinnati
560 Baldwin Hall
Cincinnati, OH 45221-0072
Telephone: (513)556-3412
hossein.ardakani@uc.edu

Robert Mangine. PT, ATC
Assistant AD, Head Athletic Trainer University of Cincinnati
Nova Care
Cincinnati Oh

Joseph F. Clark
Professor of Neurology
University of Cincinnati
MSB 7110
Cincinnati, OH 45267-0536
Telephone: (513)556-7085

Abstract: With the rapid advancement in wearable sensor technologies and predictive analytics, college and professional sports teams are facing new opportunities of leveraging such technologies and at the same time challenges of maintaining their competency. In professional sports today, the margin between winning and losing is narrow although its impact can be massive. Effective use of predictive analytics along with the implementation of advanced sensors can bring about a deeper understanding of players' physical condition and performance potential throughout individual games or the entire season, and help sports medicine personnel get the most from available resources while keeping players healthy and minimizing their risks of injury. This paper presents a predictive analytics framework for analyzing and predicting soccer players' performance data. The data consists of GPS and physiological measurements, collected from female soccer players during both practices and games using Zephyr Bioharness device. The proposed framework consists of data cleaning, filtering, visualizations and analytics modules to provide deeper insights into the data. The preprocessing modules automatically remove outliers using intelligent tools and determine first half, second half and potential overtime RISKS based on data patterns. Furthermore, comparison-based metrics have been developed to analyze the performance of players from different aspects including their activity level, fitness and consistency. For instance, Kolmogorov-Smirnov (KS) test was utilized to extract performance metrics based on players' Heart Rate and Speed, or a Neural Network-based approach was utilized to analyze the Heart Rate recovery rate of the players and quantify their recovery rate, which is important for effective play. At the end, different visualization tools were used to combine players' running patterns and speed profiles, along with various metrics. Potentially clinically relevant trends related to objective performance parameters could be observed for players during individual games and the entire season which can provide the athletic training staff with a better understanding of player's performance and inclines or declines in their performance. Future work could validate the methodology for return to play, safe play and pull from play decision tools for the coaching and training staff. The tools used for analyzing biological signals can also be expanded for other applications involving human performance monitoring.

Key words: Sports analytics; physiological signal analysis; human performance monitoring; soccer; wearable sensors;

1. Introduction

Wearable sensor technologies are facilitating the collection of high-quality data from the human body in a variety of applications including sports, workers in harsh and remote environments and military. Not only monitoring the critical physiological signals can help protect workers, soldiers or athletes, but the implementation of predictive analytics on such parameters can bring further insight and vision into the performance of individuals, which can be used for improving their performance and minimize risks associated with injuries, and physical and mental fatigue. Soccer is one the sports in which sports analytics plays a significant role and there is still a substantial opportunity for implementing predictive analytics for improving the outcome. A great example of such implementation is Leicester City, which is one of the most advanced in the English premier league soccer to use data analytics, and has been using wearable technology and analytic tools in the past 10 years [1]. There is a growing trend in professional soccer towards using metrics and analytics.

Despite the development of analytic tools for analyzing physiological signals, performance analysis in Soccer has been done with mostly descriptive game statistics like the number of goals, passes, shots [2, 3] or event-based data [4]. For example, the uniqueness and consistency in player in-game movements is quantified to identify potential replacements to a given player [4]. Ball pass origins, destinations and spatiotemporal data has been used extensively to develop a conceptual field map that help predict the likelihood of the passes leading to shots [5]. Another dimension to this analysis is to find a relation between a player's physiological performance and on field performance. The fitness and conditioning of players is of great importance during a game as well as for the duration of the season. A comparison of professional and amateur players indicates that elite players have higher measures of maximum volume of oxygen (VO_2) and lower heart rates [6]. Other works, indicate the significance of heart rate (HR) along with VO_2 levels, vertical jump height, weight etc., to classify players based on those parameters as well as their proficiency and field position [7,8,9]. Apart from raw

HR, monitoring heart rate recovery (HRR) after a sub maximal running test has been used to predict changes in physical performance of players [10].

Various studies have briefly analyzed and established the importance of parameters like HR, HRR, speed profiles, player running patters and fatigue levels during a game. It is challenging to incorporate all these measures to quantify or monitor players' performance. Our work proposes a novel approach to both visualize and quantify these measures to assist sports medicine personnel to make decisions.

2. Data

The data was collected using Zephyr Bioharness from 21 players of a soccer team during both practices and games for an entire season. The key variables used in this study were GPS and heart rate. The data was collected at the sampling rate of 1 Hz. Table 1 shows a complete list of variables used in this study.

Table 1. List of variables used in the study with units

Parameters	**Units**
Time	Seconds
Heart Rate	Beats per minute
Breath Rate	Breaths per minute
Speed	Mile per hour
Peak Acceleration	G Force
GPS (Latitude, Longitude Co-ordinates)	Degrees Minutes Seconds

3. Methods

The methodology for analyzing the data was developed with the oversight of the sports medical staff to ensure un-encumbered performance during games and practice, to ensure the applicability of such analysis. The methodology consisted of data preprocessing, player activity profile extraction and visualization, similarity-based player performance metric development, and heart rate recovery modeling. This section briefly explains the methodologies used for this study.

3.1. Data Preprocessing: The data collect from bio sensors tend to have large amounts of noisy data. This noise comes from down time, players going out of range for the sensors, local EM parameters etc. The players also typically wear the sensors well before the game begins and when they aren't actively playing on field. To ensure that such extraneous data doesn't bias the models we developed preprocessing methods to account for the non-game noise. This process consisted of removing outliers and out-of-field GPS samples. A simple k-means algorithm was implemented to estimate game start and end times of each game based on player speed profiles and time.

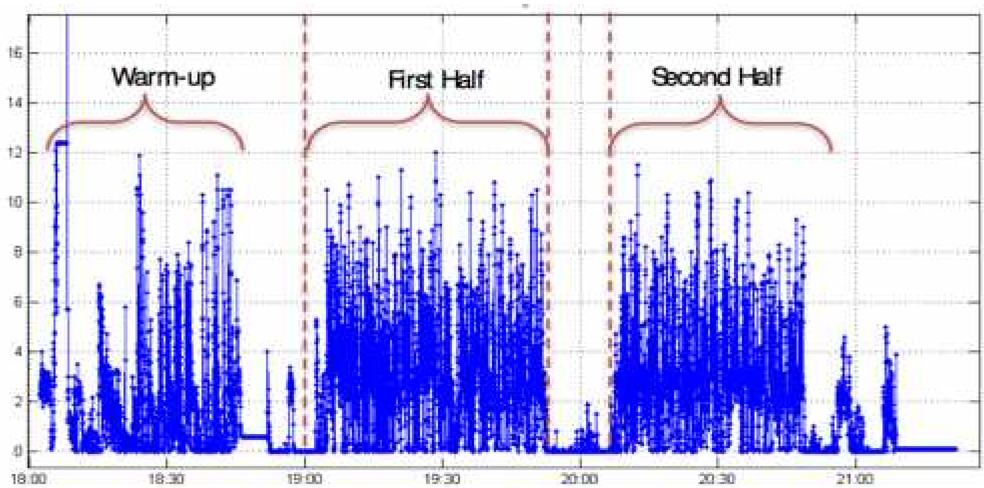

Figure 1. An example of a player's speed profile during one game; horizontal axis shows the time and vertical axes shows the speed in miles per hour

3.2. Player Activity Profile: Players' activity profiles were established based on analyzing the GPS measurements from two different perspectives: their speed profiles and its consistency throughout the game, and from game to game, and the movement vectors and movement patterns during the game. This was further extended to observe the relationship between player HR and distance profiles. Speed and HR are two indicators of how active a player is on the field and how easily they get fatigued. Figure 2 shows the movement pattern during one half of a game with intensity being represented by the color gradient white (low speed) to red (high speed).

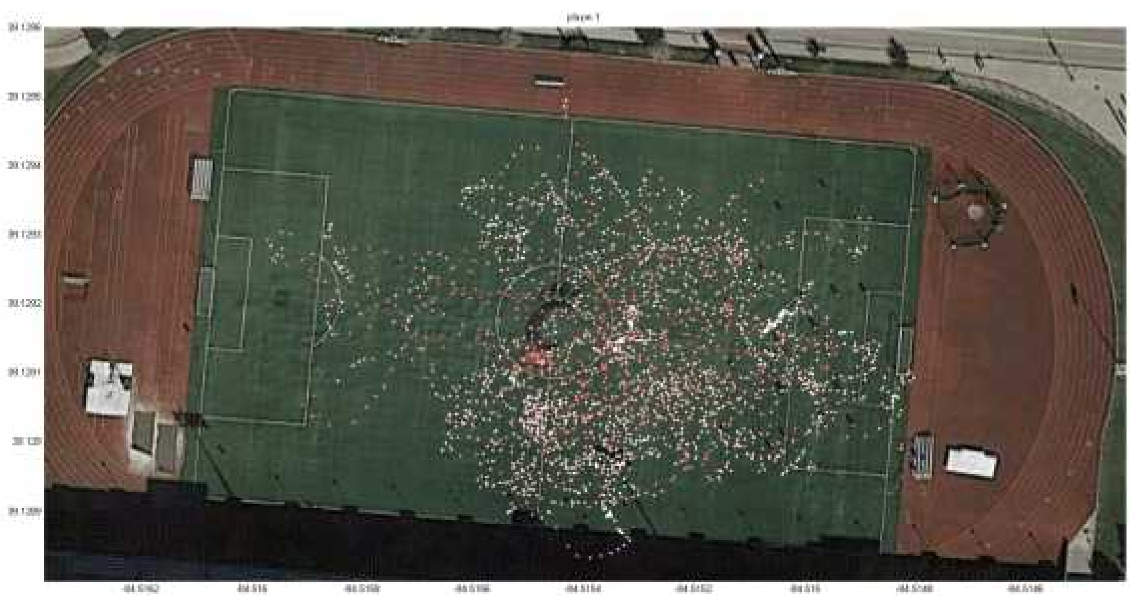

Figure 2. An example of a player's movement pattern during one half of a game; the color of each point represents the speed of the player at that point, red being the highest speed and white the lowest.

To quantify these speed data per game over the season, speed distributions were computed at 10-minute time windows during the game as shown in Figure 3. Three different speed zones were identified including 0-4 mph (low), 4-8 mph (medium) and 8-16 (high), and raw speeds were categorized into respective zones based on their values. The relationship between a players' percentage of maximum HR (%HRmax) exerted and distance traversed during a game has been visualized as shown in Figure 4.

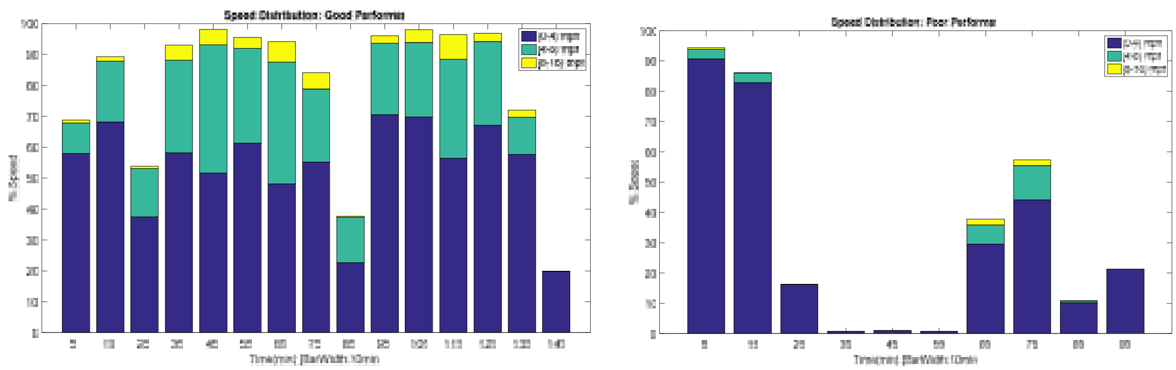

Figure 3. An example of a player's movement pattern during one half of a game; the color of each point represents the speed of the player at that point, red being the highest speed and white the lowest.

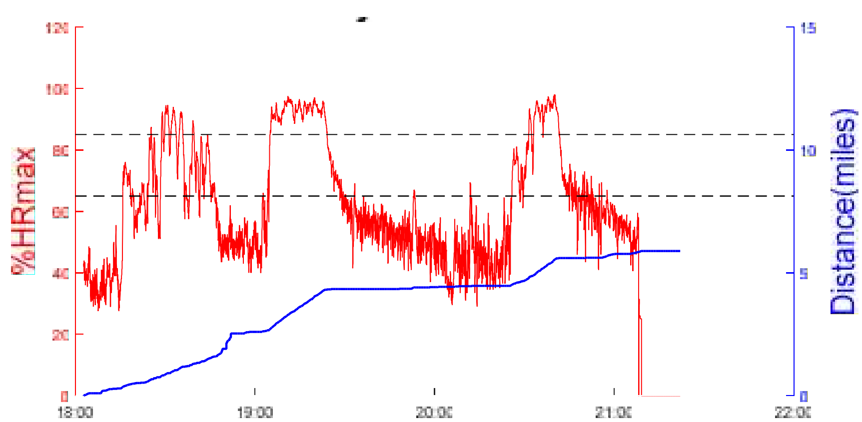

Figure 4. A player's %HRmax and distance versus time.

3.3. Similarity-based Player Performance Metric Development

The similarity-based performance metric development is based on comparing the measured Heart Rate and Speed parameters during the games. The ratio of Heart Rate to Speed (HR/S) at each instance was calculated. Kolmogorov-Smirnov (KS) test was used to benchmark each player's performance at a desired period of time with the rest of the data (or a specific player to benchmark against). KS test, explained in [11], utilizes the Cumulative Distribution Function (CDF) to determine whether a target distribution is statistically similar to a reference distribution. In this case, the data from the rest of the players is considered as the reference distribution. In order to quantify the similarity (or difference) between the two distributions, a distance metric D between the two distribution is defined. Mathematically, it is defined as:

$$D = \max_{-\infty < x < \infty} |T(x) - R(x)| \qquad (1)$$

where $T(x)$ is the target distribution and $R(x)$ is the reference distribution. As mentioned, the parameter HR/S was considered as the input for calculating the metric. Figure 5 shows an example of the distribution functions during 11 games for a player with higher performance than the team average. As it can be observed from Figure 5, the ECDF for Player 21 is at the left side of the ECDF of the rest of the team, meaning that this player has a lower heart rate compared to the rest of the team. Figure 6 shows the metric for HR/S which is equivalent to the maximum distance between the two distributions in each of the subplots in Figure 5. The performance consistency of Player 21 can be observed from the metric values. Figures 7 and 8 show the ECDF's and the HR/S metric for another player with less performance consistency.

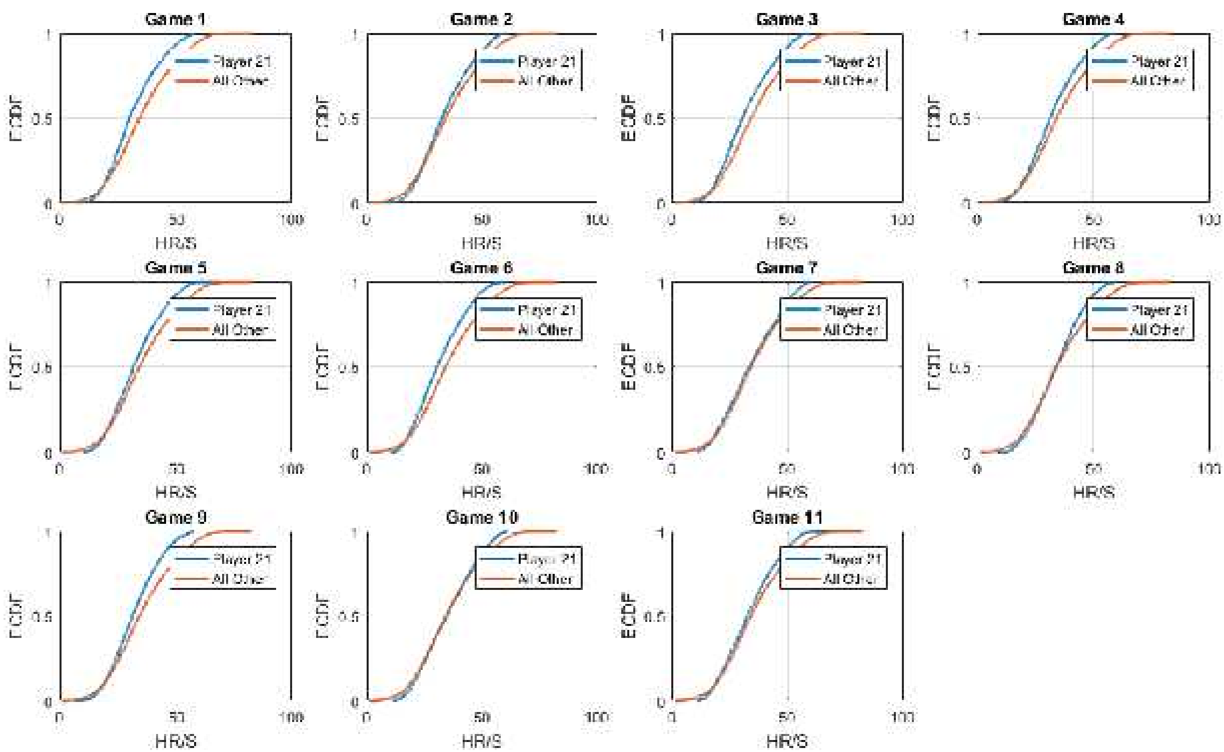

Figure 5. Empirical Cumulative Distribution Function of Heart Rate / Speed metric for a player during 11 games.

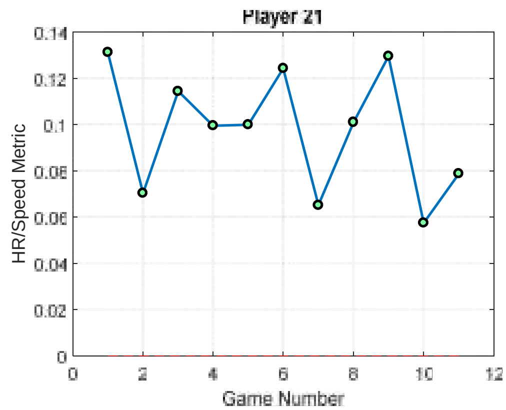

Figure 6. Similarity-based Heart Rate / Speed metric for a player during 11 games.

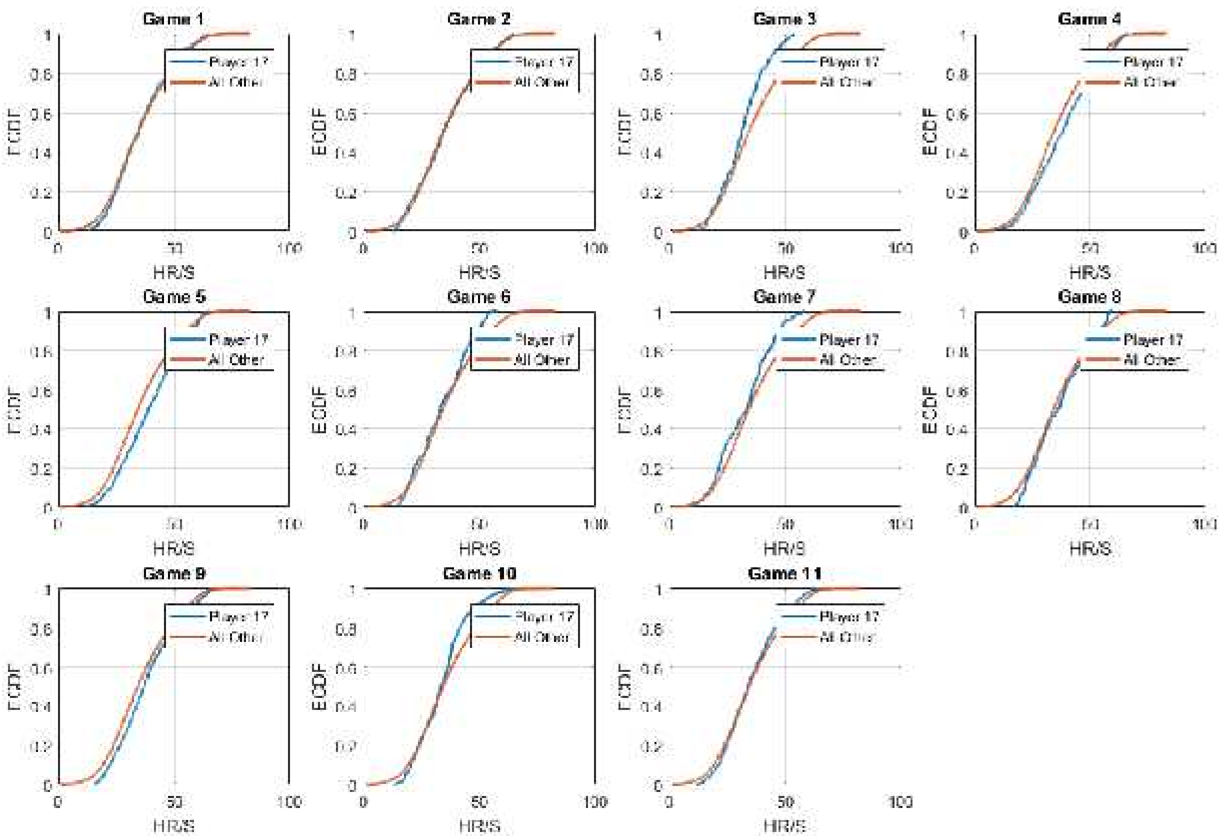

Figure 7. Empirical Cumulative Distribution Function of Heart Rate / Speed metric for a player during 11 games.

3.4. Heart Rate Recovery Modeling

The heart rate recovery model uses a neural network to understand how players compare under similar levels of strain in respect to their heart rate recovery. Through examining how player's actual data compares to a modeled player, a set of scores was then acquired. Figure 9 gives a description of the model detailed below.

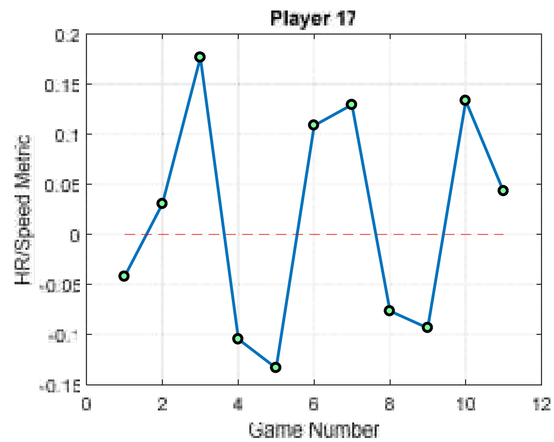

Figure 8. Similarity-based Heart Rate / Speed metric for a player during 11 games.

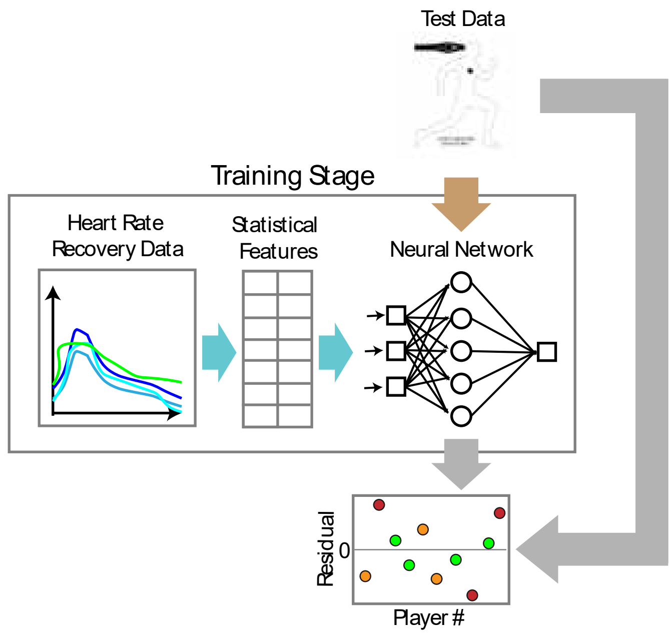

Figure 9. Heart Rate Recovery Model.

Subsets of preprocessed data were extracted to represent the moments in which players are in a state of recovery after a high speed moment. To do this, the speed data for a single player at a time was first grouped into sections of continuous time to remove issues with improper data acquisition. Then, for each section of a significant length, the data was passed into a windowed mean algorithm with a window size of 10 to smooth the data to show the pattern of play for the player. These windowed mean speeds were then passed into a filter that set all values less than a limit of 1.5 mile per hour to zero. This then created a set of speed spikes in which the players exhibited a higher speed for an amount time that then dropped below the threshold. The locations of these speed spikes were then recorded at the end of the spike to either the next speed spike or 150 seconds later depending on whether there was a nearby speed spike after the current one or not. This Heart Rate Recovery data was used to generate the statistical featured for the neural network.

The features fed into the neural network were the current heart rate, the mean speeds for 5-50 seconds at 5 second intervals, the current speed, the current distance traveled since the start of the game, the time since the start of the half on the field, the current activity and the respiration rate data. The most influential features in predicting the target value were mean speeds for the last 50 second and the last 5 seconds. The target value of the neural network is the change in heart rate per a second of time.

To train the neural network, the players labeled by the coach as top performers and bottom performers were used. Then, the final score was found by then examining the difference between the predicted values for the change in heart rate over a second compared to the actual change in heart rate over a second of time. This further allowed examining the rate of recovery of the players compared to a shown good player.

4. Results and Discussion

Data preprocessing: The data processing section was developed to be autonomous and versatile for different situations and data settings. It was established based on common sense and occasional errors observed in the recorded data. The beginning and end of the game, half time etc. were also determined by means of a k-means clustering based approach and proved to work successfully by considering the speed profile of all the players.

Metric development: Metrics can be very informative and indicative for coaches and training staff to get a better understanding of each individual player and the roster as a whole, along with increases or decreases in their abilities, strength and readiness. For this study, two different approaches are introduced to produce metrics mainly based on heart rate and speed. Although the metrics need improvements and standardization, they provided helpful insight for the training staff in terms of trends observed throughout the season, and relative performance of each player compared to the rest of the roster. In a number of players, a consistent trend was observed in HR/S metric throughout the season indicating either improvement or decline in performance. This metric is currently showing promise in terms of quantifying players' performance but as part of the future work, it needs to be standardized and validated using cases where the performance and readiness level of a player is known i.e. identified by the coaching and training staff for validation purposes.

Overall, distribution comparison techniques provide a simple but powerful tool to quantify the similarities and differences between players' physiological signature and performance.

A neural network-based approach was also proposed to develop a data-driven model which learns based on the observed heart rate recovery rates. In sports medicine, rapid heart rate recovery is considered a sign of good conditioning and lack of fatigue, so this parameter could be important or helpful to the sports medical professional. Further, fatigue in the fit individual is thought to be associated with an increased risk of injury, therefore this modeling system could be used to keep athletes safe by removing them from play and reduce risk of injury. The model was developed to predict the heart rate recovery patterns based on the given heart rate values from prior samples. Ideally, the model would need to be trained based on data from a collection of players who are fit and in their best condition so that the model can give each individual player a score based on comparing them to the best.

Visualization: In sports predictive analytics, data visualization is key. The data and the information extracted from it need to be visualized in a way that makes it easier for the coach and training staff to make informed decisions. This may be more significant when coaches monitor real-time data and analytics to make decisions on substitutions or change tactics. In this study, the visualizations above were developed based on regular feedback from coaching staff, to provide the information in the most convenient way. For example, Visualizing the players' movement patterns during the game (Figure 2) helped in analyzing a player's positions during the game along with improving cardiovascular performance in training. Figure 3, captures the difference in speed profiles of a Good and Poor performer. It is observed that a Good performer has spent greater duration in mid/ high speed zones compared to a player with lower performance. In Figure 4 a strong correlation between distance traversed and %HRmax is observed, also good performers do not cross 80-90 %HRmax typically. The figure also helps in identifying warm-up, game halves and half-time zones. High activity zones are when %HRmax and distance increase.

Players' positions: As each position in the soccer field requires specific sets of skills and physical capabilities, considering the position of each player can significantly improve the outcome of the analysis and make the metrics more informative.

5. Conclusion and Future Work

As the wearable sensor technology is helping us capture more data from the field, there is a great opportunity for sports teams to develop advanced Predictive Analytics to monitor the performance of athletes, optimize their training programs for maximized physical readiness while minimizing the injury risks, and win more games. This work provides a collection of simple but effective tools for analyzing physiological and game data, by integrating the knowledge and experience of University of Cincinnati's Sports Medicine team, along with the experience of researchers at Center for Intelligent Maintenance Systems (IMS) at the University of Cincinnati in developing predictive analytic tools. The future work consists of enhancing the analysis and more particularly metrics, and adding more tools to this collection for analyzing other variables being measured. Moreover, as next step, these algorithms will be integrated into a web-based predictive analytics platform that provides the coaching team, training staff and the athletes' access to the results of the analysis at any time for making optimum decisions.

6. References

[1] Foxy Leicester City FC won Premiership with Data Analytics, Simon Creasey 16 May 2016 10:45. http://www.computerweekly.com/news/450296302/Foxy-Leicester-City-FC-won-Premiership-with-data-analytics

[2] Anderson, Chris, and David Sally. The numbers game: why everything you know about football is wrong. Penguin UK, 2013.

[3] Duch, Jordi, Joshua S. Waitzman, and Luís A. Nunes Amaral. "Quantifying the performance of individual players in a team activity." PloS one 5.6 (2010): e10937.

[4] Gyarmati, László, and Mohamed Hefeeda. "Analyzing In-Game Movements of Soccer Players at Scale." arXiv preprint arXiv:1603.05583 (2016).

[5] Brooks, Joel, Matthew Kerr, and John Guttag. "Developing a Data-Driven Player Ranking in Soccer Using Predictive Model Weights." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.

[6] OSTOJIC, SERGEJ M. "Elite and nonelite soccer players: preseasonal physical and physiological characteristics." Research in Sports Medicine 12.2 (2004): 143-150.

[7] Little, Thomas, and Alun G. Williams. "Measures of exercise intensity during soccer training drills with professional soccer players." The Journal of Strength & Conditioning Research 21.2 (2007): 367-371.

[8] Ostojić, Sergej M. "Physical and physiological characteristics of elite Serbian soccer players." Facta universitatis-series: Physical Education and Sport 1.7 (2000): 23-29.

[9] Sporis, Goran, et al. "Fitness profiling in soccer: physical and physiologic characteristics of elite players." The Journal of Strength & Conditioning Research 23.7 (2009): 1947-1953.

[10] Buchheit, M., et al. "Monitoring changes in physical performance with heart rate measures in young soccer players." European journal of applied physiology 112.2 (2012): 711-723.

[11] Cong, Feiyun, Jin Chen, and Yuna Pan. "Kolmogorov-Smirnov test for rolling bearing performance degradation assessment and prognosis." Journal of vibration and control 17.9 (2011): 1337-1347.