Monitoring Internal Load with a Mobile App in Professional Soccer Players after the COVID-19 Lockdown (Readiness Soccer): Observational Study

Salvador Moreno-Gutierrez, Oresti Banos, Miguel Damas, Hector Pomares, Paula Postigo-Martin, Irene Cantarero-Villanueva, Manuel Arroyo-Morales

Submitted to: JMIR mHealth and uHealth
on: May 26, 2021

Table of Contents

Monitoring Internal Load with a Mobile App in Professional Soccer Players after the COVID-19 Lockdown (Readiness Soccer): Observational Study

Salvador Moreno-Gutierrez[1] MSc; Oresti Banos[1] PhD; Miguel Damas[1] PhD; Hector Pomares[1] PhD; Paula Postigo-Martin[2] MSc; Irene Cantarero-Villanueva[2] PhD; Manuel Arroyo-Morales[2] MD, PhD

[1]Department of Computer Architecture and Technology Research Centre for Information and Communication Technologies (CITIC-UGR) University of Granada Granada ES
[2]Department of Physical Therapy University of Granada Granada ES

Corresponding Author:
Manuel Arroyo-Morales MD, PhD
Department of Physical Therapy
University of Granada
Avenida de la Ilustración, 60
Granada
ES

Abstract

Background: Heavy physical and mental loads are typical for professional soccer players during the competitive season. COVID-19 lockdowns had recently forced competitions to be interrupted and later disputed in a shrunken calendar. Wearable sensors and mobile phones could be potentially useful in monitoring players' training load in such highly demanding environments.

Objective: The aim of this study was to explore whether remote heart rate variability (HRV) monitoring and self-reported wellness of professional soccer players could be useful to monitor players' internal training load and to estimate their performance during the continuation of the 2020 season after the COVID-19 lockdown in Spain.

Methods: A total of 21 professional soccer players participated in a 6-week study. Participants used an Android or iOS-based smartphone and a Polar H10 wearable ECG monitor for the duration of the study. Every morning they recorded their HRV and answered a questionnaire about their perceived recovery, muscle soreness, stress and sleep satisfaction. Smallest worthwhile change (SWC) and coefficient of variation (CV) were calculated for the logarithm of the root mean square of the successives differences (LnRMSSD) of the HRV. Players' in-game performance was evaluated subjectively by independent observers and classified as high, normal and low. In order to find which variables could be potentially linked to performance, we studied their correlation and tested for significant differences among distributions. We also trained random forest models with cross-validation and bootstrapping to find the wellness and HRV features with best predictive ability for performance.

Results: We found the usability of Readiness Soccer in a real scenario to be very good, with 81.36 points in the System Usability Scale. A total of 241 measurements of HRV and self-reported wellness were recorded. For a entire training microcycle (ie, time in between matches), self-reported high recovery (Mann-Whitney U, P=.003), low muscle soreness (P=.002), high sleep satisfaction (P=.02), low stress (Anderson-Darling, P=.03), and not needing more than 30 minutes to sleep since going to bed (Chi-Squared, P=.02), were found significant to differentiate high from normal match performance. Performance estimation models achieved the highest accuracy (73.4%) when combining self-reported wellness and HRV features.

Conclusions: HRV and self-reported wellness data were useful to monitor the evolution of professional soccer players' internal load and to predict match performance levels out of measures in a training microcycle. Despite the limitations, these findings highlight opportunities for long-term monitoring of soccer players during the competitive season as well as real-time interventions aimed at early management of overtraining and boosting individual performance.

(JMIR Preprints 26/05/2021:30728)
DOI: https://doi.org/10.2196/preprints.30728

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

- **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.
- No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

- **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
- Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
- Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Monitoring Internal Load with a Mobile App in Professional Soccer Players after the COVID-19 Lockdown (Readiness Soccer): Observational Study

Abstract

Background: Heavy physical and mental loads are typical for professional soccer players during the competitive season. COVID-19 lockdowns had recently forced competitions to be interrupted and later disputed in a shrunken calendar. Wearable sensors and mobile phones could be potentially useful in monitoring players' training load in such highly demanding environments.

Objective: The aim of this study was to explore whether remote heart rate variability (HRV) monitoring and self-reported wellness of professional soccer players could be useful to monitor players' internal training load and to estimate their performance during the continuation of the 2020 season after the COVID-19 lockdown in Spain.

Methods: A total of 21 professional soccer players participated in a 6-week study. Participants used an Android or iOS-based smartphone and a Polar H10 wearable ECG monitor for the duration of the study. Every morning they recorded their HRV and answered a questionnaire about their perceived recovery, muscle soreness, stress and sleep satisfaction. Smallest worthwhile change (SWC) and coefficient of variation (CV) were calculated for the logarithm of the root mean square of the successives differences (LnRMSSD) of the HRV. Players' in-game performance was evaluated subjectively by independent observers and classified as *high*, *normal* and *low*. In order to find which variables could be potentially linked to performance, we studied their correlation and tested for significant differences among distributions. We also trained random forest models with cross-validation and bootstrapping to find the wellness and HRV features with best predictive ability for performance.

Results: We found the usability of Readiness Soccer in a real scenario to be very good, with 81.36 points in the System Usability Scale. A total of 241 measurements of HRV and self-reported wellness were recorded. For a entire training microcycle (ie, time in between matches), self-reported high recovery (Mann-Whitney U, P=.003), low muscle soreness (P=.002), high sleep satisfaction (P=.02), low stress (Anderson-Darling, P=.03), and not needing more than 30 minutes to sleep since going to bed (Chi-Squared, P=.02), were found significant to differentiate high from normal match performance. Performance estimation models achieved the highest accuracy (73.4%) when combining self-reported wellness and HRV features.

Conclusions: HRV and self-reported wellness data were useful to monitor the evolution of professional soccer players' internal load and to predict match performance levels out of measures in a training microcycle. Despite the limitations, these findings highlight opportunities for long-term monitoring of soccer players during the competitive season as well as real-time interventions aimed at early management of overtraining and boosting individual performance.

Keywords: mHealth; mobile health; digital health; eHealth; soccer; app; internal load; heart rate variability; wellness; performance; covid-19

Introduction

The highly demanding environment of professional soccer competitions has recently joined the uncertainty of the new coronavirus disease (COVID-19) pandemic. This situation has forced players into lockdowns, causing national and international tournaments to be interrupted and delayed during the 2020 season. Resuming competition with a shrunken schedule hampered the preparation of

players [1], leading to lower physical performance [2] and increased mental stress [3] in the season continuation after the COVID-19 lockdown. In order to minimize overtraining and injury risks in such a period, soccer players need the best individual assessment of their health.
Monitoring training load is of great interest to the medical services of soccer teams looking forward to preventing increased levels of fatigue and higher risks of illness and injury [4]. Load measures can be categorized as internal or external [5], depending on whether such measures refer to aspects occurring internally or externally to the player. The external load is monitored as objective comparable measures of the exercise done [5], such as overall distances, training and match participation, accelerations and decelerations, impacts and metabolic power [6]. However, external load measures of the exercise done do not consider the internal processes of training assimilation [7]. The internal load is the individual and relative physiological and psychological stress felt by soccer players as a result of the training and competitive sessions and the rest of demands present in their daily lives [5,8].

A recent systematic review [6] gathered the most common methods to measure external and internal load, finding heart rate analysis and questionnaires as the two most used measures of internal load. On the one hand, wellness questionnaires as the Hooper Index [9] have been proven to be useful to measure internal load by taking into account stress, fatigue, sleep quality and muscle soreness after training. In fact, the Hooper Index has been associated with external training load [10,11], fatigue [12,13] and heart rate variability [14] (HRV) in soccer players. On the other hand, the subjective nature of the Hooper Index may be complemented with the physiological approach of HRV monitoring to monitor internal load [14,15]. HRV monitoring is known to be a reliable tool to measure physiological indicators of stress and recovery against variations of the external load in athletes [15,16]. Moreover, HRV has been found to be a significant parameter to relate internal load with external load [17]. Both HRV [16] and external training load indicators [18,19] have shown a relationship with match-day performance, even beyond linearity by leveraging data mining techniques [20].

Objective measurements of external and internal training load are specially needed during lockdown and intensive competition periods in order to avoid injuries and maintain performance. Such objective measurements should be quick and reliable to be used in practical sport settings. To the best of our knowledge there are no available solutions to monitor internal load and HRV to objectively monitor the recovery processes of professional soccer players. In light of this opportunity, we present the smartphone app *Readiness Soccer*, a novel approach to internal load assessment in soccer players. In combination with a Bluetooth wearable ECG sensor, Readiness Soccer collects daily measures of short HRV and self-reported wellness with questions regarding sleep, stress, fatigue and muscle soreness to assess internal load.

The aim of this study was to explore whether remote HRV monitoring and self-reported wellness of professional soccer players could be useful to monitor internal load, especially during a very challenging scenario such as the season continuation after the first COVID-19 lockdown. First, we evaluated the feasibility of using the Readiness Soccer app in a real scenario during competition. Next, we explored the relationship of match performance with HRV parameters and self-reported wellness. For this exploration, we first described the HRV and self-reported measures gathered, then we studied if there were significant differences among performance levels and the measures in a training microcycle, and, finally, we estimated match-day performance out of the same measures while studying the relevance of the features selected.

Methods

Participants

The selected participants were 21 male professional soccer players (age: mean 26.7, std 4.2; height: mean 182.4 cm, std 5.0; weight: mean 74.3 kg, std 5.0). Participants were provided with a Polar H10 ECG monitor (Polar Electro Ltd.) and with the Readiness Soccer app, installed in their personal smartphones. A total number of 241 measurements were recorded during the competition period. All participants provided their voluntary written informed consent, and the study conformed with the Declaration of Helsinki [21].

Design

To give response to the aim of the study we used a prospective observational cohort study. All of the 241 measurements were obtained during 6 weeks (10 matches), in the resume of the 2020 season after the COVID-19 lockdown. Participants were instructed to measure their HRV and answer wellness questionnaires with the app *Readiness Soccer* every day in the morning at home, between 8 and 11 am to avoid influences of circadian rhythm. Measurements started 8 days before the first match to obtain solid HRV baselines. The individual match performance of players was evaluated throughout the entire study period.

Readiness Soccer

Readiness Soccer is a cross-platform (Android and iOS) smartphone app that is designed and developed to provide an ubiquitous, simple and intuitive interface for daily recordings of HRV and self-reported wellness in professional soccer players. Players are able to use this app right after waking up and from their homes, only needing access to a Bluetooth wearable ECG that is paired to Readiness Soccer. The storage of data is supported by a centralized secure server. An overview of the system supporting the data collection can be found in Figure 1.

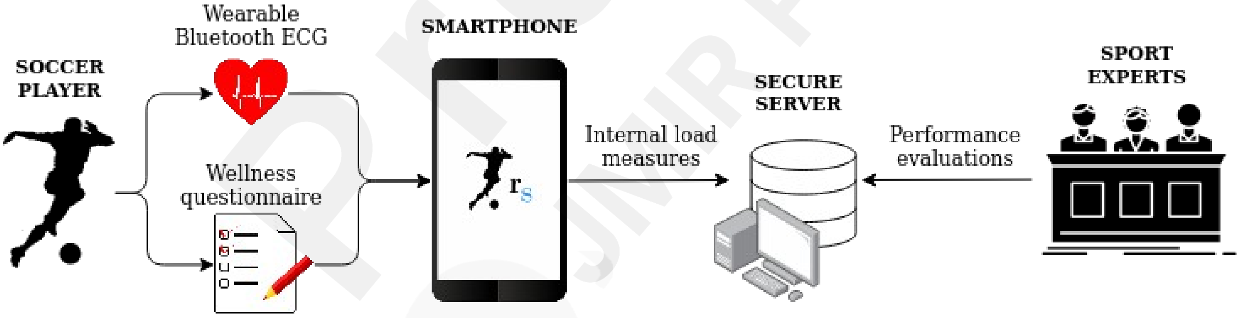
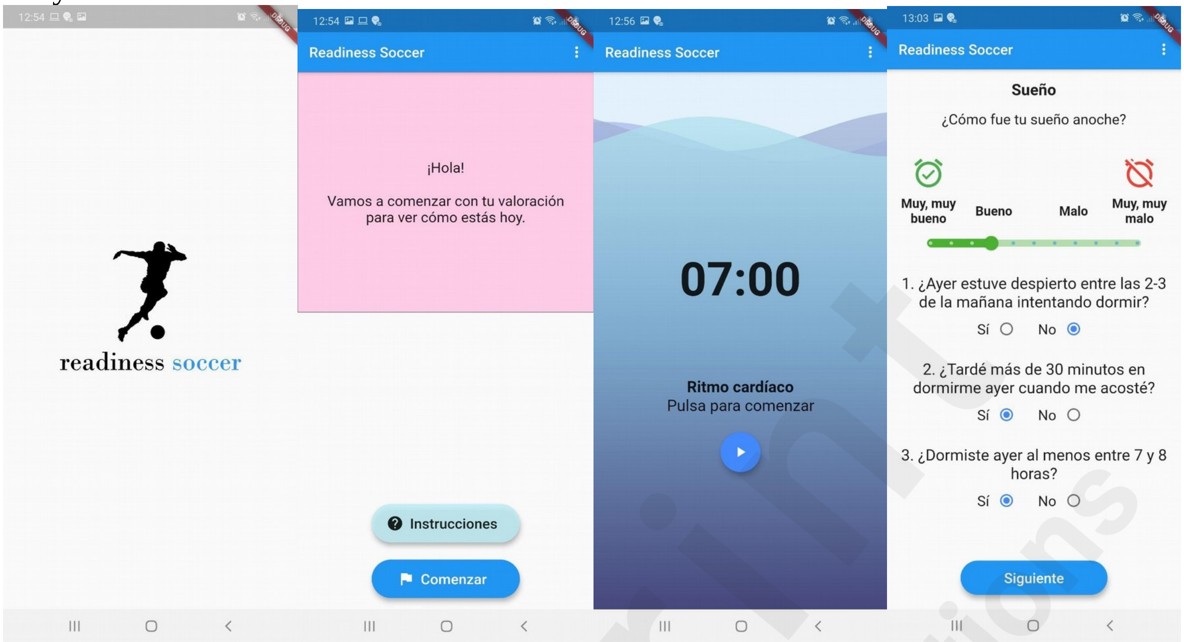


Figure 1: Top level view of the system.

Screenshots of the main views in Readiness Soccer are shown in Figure 2. First, the app welcomes the user showing the app icon while loading (Figure 2.a) and then presents a welcome message, giving the options to *Start* the measurement protocol through the *Start* button, or to check the *Instructions* on how to do so (Figure 2.b). Next, if the Bluetooth ECG monitor is connected, the app allows the user to start recording HRV with the *Play* button (Figure 2.c). Finally, the user is given each question defined for the self-reported wellness (see sleep satisfaction question in Figure 2.d). After recording HRV and self-reported wellness, the app confirms to the player the successful recording of data in a similar view as the welcome message in Figure 2.b. The specific requirements to monitor HRV and self-reported wellness are thoroughly described in their own sections ahead.

Since the purpose of this study was observational, the app provided no feedback beyond the confirmations of data being correctly recorded and sent to the secure server. Regarding data security and privacy, data were gathered and stored meeting the European General Data Protection

Regulation (GDPR). All online communications were secured via HTTPS connections with SSL/TLS encryption. The app was developed using Flutter (Google LLC) to enable cross-platform availability in both Android and iOS devices.



a) Loading screen b) Welcome view c) HRV d) Sleep quality

Figure 2: Readiness Soccer mobile app.

Self-Reported Wellness

We used a modification of the Hooper Index Questionnaire [14] to measure perceived *sleep satisfaction*, *recovery*, *stress* and *muscle soreness* levels. Each of these items was reported using 0-10 Likert scales with the Readiness Soccer app. All wellness factors were asked using a similar view as the sleep quality questionnaire shown in Figure 2.d. Besides the slider, the sleep quality questionnaire also contained three yes/no (*True/False*) insomnia-related questions: (*1) Was I awake between 2.00 and 3.00 AM last night trying to sleep?; (2) Did it take more than 30 minutes to sleep since I got in bed?;* and *(3) Did I sleep at least between 7 and 8 hours?*

Likert scales were labelled with two positive and two negative oriented stamps, all of them depending on the question asked, but following the same left-to-right scheme of *very, very positive; positive; negative;* and *very, very negative*. Hence, the sleep satisfaction scale was labelled as *very, very good; good; bad; and very, very bad*. The stress scale was labelled as *very, very low; low; high; and very, very high*. The recovery scale was labelled as *very, very high; high; low; and very, very low*. The muscle soreness scale was labelled as *very, very low; low; high; and very, very high*.
In order to ensure that the directions to these answers were given correctly, sliders were made color-responsive to the choice selected, associating green to a positive answer (eg, high sleep satisfaction, as shown in Figure 2.d) and red to a negative one (eg, high muscle soreness), both in a continuous gradient passing through yellow in the mid values. Moreover, the initial position of the slider was randomized for every question to mitigate anchoring or learning effects in the responses [22].
Participants were required to actively answer every item in the questionnaire before advancing to the next question. This was done by enabling the *Next* button only once the participant had actively selected a score on the slider.

HRV Monitoring

Participants used the chest-strapped Polar H10 ECG monitor (Polar Electro Ltd.) for HRV recording connected to the Readiness Soccer app via Bluetooth (Figure 2.c). Players were instructed to gather their measurements in the morning between 8 and 11 am, after emptying the bladder, and before starting the day (eg, drinking coffee, having breakfast, checking the phone) [23].

The root mean square of the successive differences (RMSSD) between consecutive R-R intervals is the most accepted and reliable resting HRV-related parameter to measure the adaptation of an athlete in response to training [15,16,24]. HRV measurements consisted of 7 minutes of R-R intervals. Outliers and ectopic beats of HRV signal were replaced with linear interpolation [25,26]. We removed the first and last minute of the recording, using the 5 minutes in the middle of the R-R signal for short term HRV analysis [27]. We extracted the daily logarithm of the RMSSD (LnRMSSD) for every recording.

We computed the smallest worthwhile change (SWC) [28] and the coefficient of variation (CV) [29] of the LnRMSSD. Both SWC and CV used a minimum 5-days-length and a maximum 7-days-length rolling window to establish a baseline for LnRMSSD comparison [30]. We used a normalized version of the LnRMSSD that refers to the magnitude of change in LnRMSSD in a continuous and comparable manner for every player at the moment of the recording, as well as the SWC, but opposed to the *True* or *False* possible values. Therefore, hereon we refer to this variable as SWC_{norm}. CV and SWC_{norm} were the continuous HRV features defined for each player as:

$$CV = 100 \cdot \frac{LnRMSSD_{rolling\,std}}{LnRMSSD_{rolling\,mean}} \qquad (1)$$

$$SWC_{norm} = \frac{LnRMSSD - LnRMSSD_{rolling\,mean}}{LnRMSSD_{rolling\,std}} \qquad (2)$$

where $LnRMSSD_{rolling\ mean}$ and $LnRMSSD_{rolling\ std}$ are the mean and standard deviation computed for the available LnRMSSD of the previous days.

We also defined two categorical (*True/False)* features according to already available thresholds and criteria in the literature for the CV and the SWC [23]: CV_{ok} and SWC_{ok}. CV_{ok} was *True* if CV was below 4.5%; otherwise was *False*. SWC_{ok} was *False* if the LnRMSSD fell below the lower SWC interval. The lower SWC threshold was defined as:

$$SWC_{lower\ th} = LnRMSSD_{rolling\,mean} - 0.5 \cdot LnRMSSD_{rolling\,std} \qquad (3)$$

Performance

Performance evaluations were gathered during match days. The assessment was done by three independent sport experts for each player. Each expert labelled individual performance as *low*, *normal* or *high,* based on the impressions of in-match technical-tactic performance of players. Then, a unique value was assigned for the player after reaching an agreement. Performance was only evaluated if the player played a minimum of 45 minutes during the match according to the experts criteria, if not, it was labelled as *rest*.

In order to reflect the impact of the training sessions leading to a match, all measurements in a training microcycle were labelled with the performance of each individual. Considering any match day during the competitive season (MD_n), where *n* is the number of the match played, a training microcycle started a day after the previous match (MD_{n-1}+1) and ended the day the match was played (MD_n), that is, {MD_{n-1}+1, MD_{n-1}+2 , …, MD_n−1, MD_n}. The start of the first training microcycle was set three days before the first match day (MD_1), making it as similar as possible to the average duration of the rest of training microcycles.

Usability

To evaluate the feasibility of using the Readiness Soccer app in a real scenario during competition, we assessed its usability at the end of the study with the System Usability Scale (SUS) [31]. The SUS is a quick and reliable method that has been tested on a wide variety of systems in research and industry [32]. This scale is a 10-item questionnaire that gives a global view of subjective assessments of usability in which every item can be scored from 1, strongly agree, to 5, strongly disagree. The SUS questions are the following:

Q1. I think that I would like to use this system frequently.
Q2. I found the system unnecessarily complex.
Q3. I thought the system was easy to use.
Q4. I think that I would need the support of a technical person to be able to use this system.
Q5. I found the various functions in this system were well integrated.
Q6. I thought there was too much inconsistency in this system.
Q7. I would imagine that most people would learn to use this system very quickly.
Q8. I found the system very cumbersome to use.
Q9. I felt very confident using the system.
Q10. I needed to learn a lot of things before I could get going with this system.

A total score can be computed out of the ten items, ranging from 0 to 100 (the higher the score, the better the usability) as presented in [31]: (1) For odd-numbered items, the score is computed subtracting 1 to the user response; (2) for even-numbered items, the score is computed subtracting the user response to 5; and (3) all the scores obtained are added and multiplied by 2.5 to obtain the overall SUS score. A minimum score of 68 points necessary to assume good usability [32].

Data Analysis

First, we described the measures with descriptive statistics with the aim of giving an overview of the data collected. This description consisted of measure count and box plot distributions against time, participants and match-day performance levels.
Next, we studied the relationship of performance with self-reported wellness and HRV measures. We started computing bivariate correlations between each feature and performance in order to gain an understanding of the strength and directionality of the linear relationship between each feature and the performance. We used *Spearman* correlation, suitable for non-normal distributions. The magnitude of the correlations was classified according to [33]: $r \le 0.1$, trivial; $0.1 < r \le 0.3$, small; $0.3 < r \le 0.5$, moderate; $0.5 < r \le 0.7$, large; $0.7 < r \le 0.9$, very large; $r > 0.9$, nearly perfect .

We then looked for significant differences among performance levels against each one of the variables collected. The tests used for continuous variables were the nonparametric *Kolmogorov-Smirnoff* (KS), *Anderson-Darling* (AD) and *Mann-Whitney U* (MWU) [34,35], all suitable for non-normal distributions. The purpose of using these three tests was to reflect the differences among distributions from different perspectives. KS computes differences in the cumulative distributions functions of the distributions compared; its null hypothesis is that the samples are drawn from the same distribution, and it is sensitive to differences in the median and the shape of the cumulative distribution function of the distributions. AD is a modification of KS that gives more relevance to the tails of the distributions and shares the same null hypothesis as KS. MWU is the nonparametric alternative to ANOVA with the null hypothesis that, for randomly selected values X and Y from two distributions, the probability of X being greater than Y is equal to the probability of Y being greater than X. For categorical variables, we used the *chi-squared* (CS) test to determine if there is a significant difference between the frequencies in one or more categories of two distributions [36].

Finally, in order to explore beyond the univariate relationships between performance and the self-reported wellness and HRV features, we built a match performance estimation model. We defined the estimation of performance level as a classification problem where all measures in a training microcycle were labelled with the performance value during the match in it {−1, 0, 1}, for *low*, *normal*, and *high* performance, respectively. We used stratified 10-fold cross-validation on the entire dataset to build the match performance estimation model.

For the sake of interpretability, we used random forests with bootstrapping to train our models. Decision tree models predict the value of a target variable by learning linear decision rules inferred from the data features, using multiple variables to make the predictions. Random forests are an ensemble of decision tree classifiers in which randomness has explicitly been inserted into the building process of each decision tree. This is done by selecting different variables and bootstrapping the dataset at the moment of building each one of the trees composing the random forest. These mechanisms ensure low correlation between the different decision trees conforming the model, hence making random forests robust to errors, outliers and overfitting [[37]]. Once the random forest model is built, predictions are computed as the mode of the outputs delivered by every decision tree.

The process of building an estimation model also served the purpose of studying the relevance of each feature to estimate performance. In a random forest, feature importance represents the weight of each feature across all decision trees when classifying, using in our case the *entropy* criteria. We analyzed the weight of each variable across different selections of features while looking for a maximization of prediction accuracy.

Data were analyzed using Python (version 3.6.9); figures were plotted with Pandas (version 1.1.3) [38] and Seaborn (version 0.11.0) [39]; statistical analysis were done with SciPy (version 1.5.3) [40]; model training was done with Scikit-learn (version 0.23.2) [41] .

Results

Usability

A total of 11 participants voluntarily completed the 10-item SUS usability survey at the end of the study. Answers ranged 57.5 to 92.5, averaging an overall SUS score of 81.36 points. This score is over the minimum threshold of 68 points necessary to assume good usability [[32]]. Only one of the SUS measures scored below this minimum usability threshold. Overall SUS results are shown in Figure 3.

Figure 4 shows an aggregation to each one of the ten SUS questions for all players. For the odd-numbered questions, 7 players (64%) reported they would like to use Readiness Soccer frequently (Q1); all players thought the app was easy to use (Q3); only one player thought the functions provided by the app were not well integrated (Q5); and 10 players (91%) conceived that most people would learn to use Readiness Soccer easily (Q7) and felt very confident using it (Q9). For the even-numbered questions, 10 players (91%) did not found Readiness Soccer unnecessarily complex (Q2) neither thought they needed the support of a technical expert to use it (Q4); seven players (64%) did not found there was too much inconsistency in the app (Q6); and none of the players found Readiness Soccer cumbersome to use (Q8) neither considered they needed to learn a many things before being able to use it (Q10).

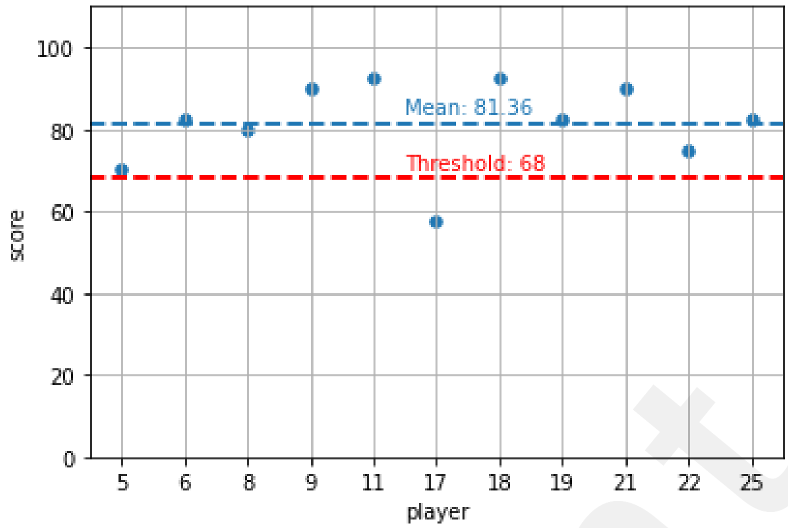


Figure 3: Overall SUS scores of Readiness Soccer. Individual scores are marked with dots; mean score in a blue dashed line, and minimum usability threshold in red.

Figure 4: SUS answers aggregated by question. Darker green tones depict strong and mild disagreement; lighter green and yellow tones represent mild and strong agreement.

Description of Data Collected

Figures 4 and 5 show the distribution of data collected across the study. Figure 5 shows how the number of measures collected by day oscillated in time (mean 3.0, std 2.4), descending especially by the end. Match days are indicated with vertical lines, hence depicting the training microcycles among matches. The duration of training microcycles ranged from 72h to 144h (mode 72 h, mean 90.7 h, std 29.45) for the 10 matches.

Figure 6 details the number of measures collected by each participant. Eleven players recorded less than 10 measures and twelve recorded more or equal than 10. Some players (#13, #27, #28) only recorded the initial measures before the first training microcycle (marked as *baseline)*. For *rest* measures (those not fulfilling the minimum time-played criteria), although they were distributed across all participants, they composed the majority of data for some of the players such as #4, #5, #9, #18, #19 and #32. *Low* performance measures were found only in two participants (#12, #17).

Normal and *high* measures were present in a more balanced distribution for the rest of participants, appearing both simultaneously in eight participants (#6, #7, #8, #11, #17, #21, #22, #25).

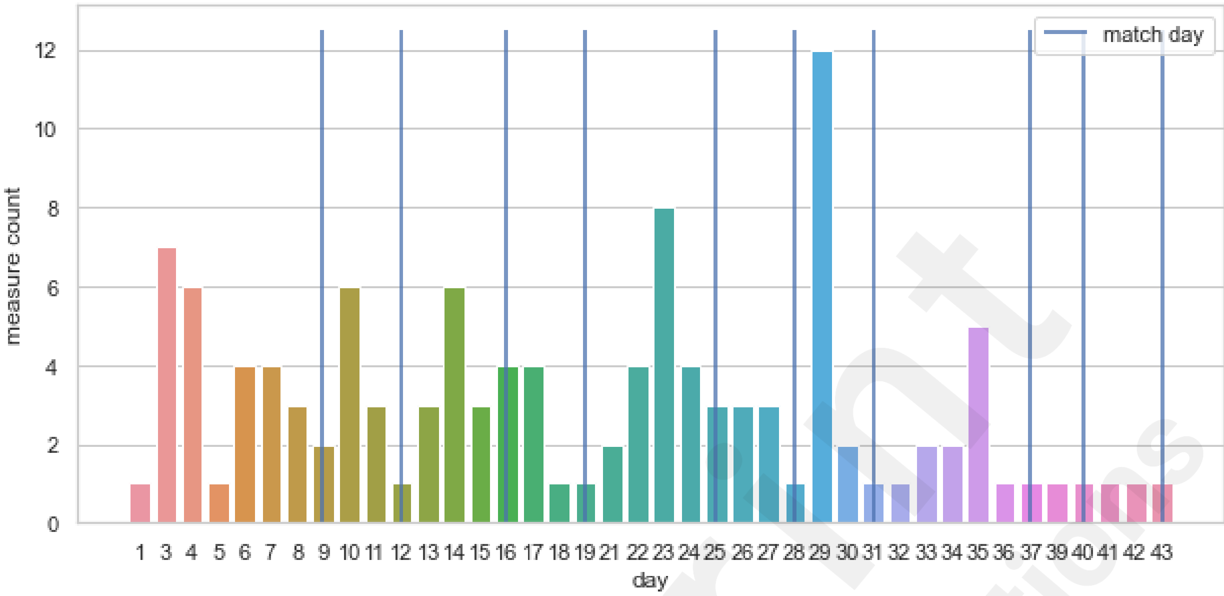


Figure 5: Measure count by day. Match days are labelled with blue vertical lines.

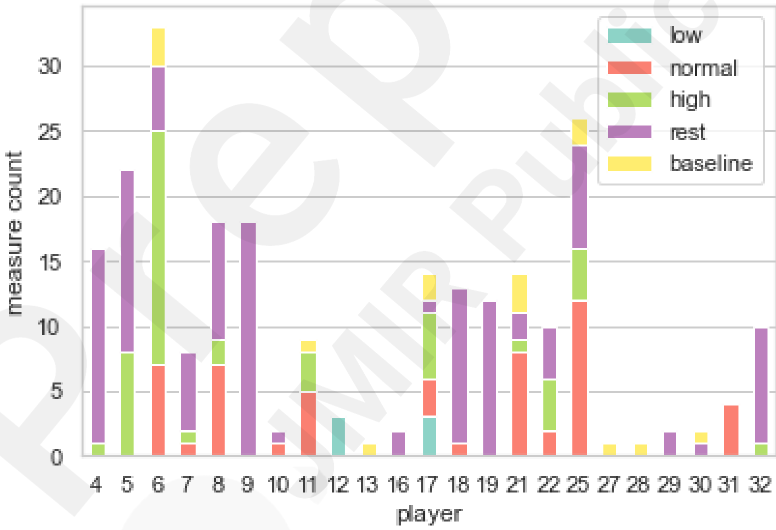


Figure 6: Measure count by player. Measures are stacked based on their labels for performance (low, normal and high), resting (rest), or setting an HRV baseline at the beginning of the study (baseline).

Self-Reported Wellness

Out of the 241 measurements gathered, a subset of 105 measures recorded by 15 players (age: mean 27.7, std 4.0; height: mean 181.6 cm, std 5.2; weight: mean 74.7 kg, std 4.4) was selected for self-reported wellness analysis. These measures matched the criteria of playing a minimum of 45 minutes on a match day. All measurements in a training microcycle were labelled with the match performance observed (high: 48; normal: 51; low: 6). The remaining 121 measures were from days before the first training microcycle (*baseline*) and from players who did not fulfill the minimum time played criteria (*rest*).

Figure 7 depicts the self-reported wellness measures for all participants and for the whole duration of the study in boxplots depending on performance. All measures seemed to follow an intuitive directionality (eg, the lower the stress or the muscle soreness, the higher the performance; or the higher the recovery or the sleep satisfaction, the higher the performance). Mean stress values decreased in scoring with 4.00, 2.51 and 2.08 for low, normal and high performance, respectively; for recovery, mean values increased in scoring with 5.33, 6.06 and 7.10 for the same performance labels; for sleep satisfaction, it increased with values 5.00, 6.20 and 7.00; and for muscle soreness, it decreased with values 4.17, 3.37 and 2.33.

Figure 8 shows the distributions of sleep quality categorical questions related to performance. The profiles or *True/False* ratio for each variable can be drawn intuitively at sight. The question “were you awake between 2 and 3 am last night trying to sleep?” kept regular ratios across all performance levels (17%, 24% and 21% of True values in low, normal and high performance, respectively). The question “did you sleep at least between 7 and 8 hours?” showed similar ratios for normal and high performance with 67% and 71% of True values, respectively, whilst only holding a 33% for low performance. The question “did it take more than 30 to sleep since you got in bed?” presented a descending trend of ratios across performance levels, with 83%, 45% and 31% of True values for low, normal and high performance.

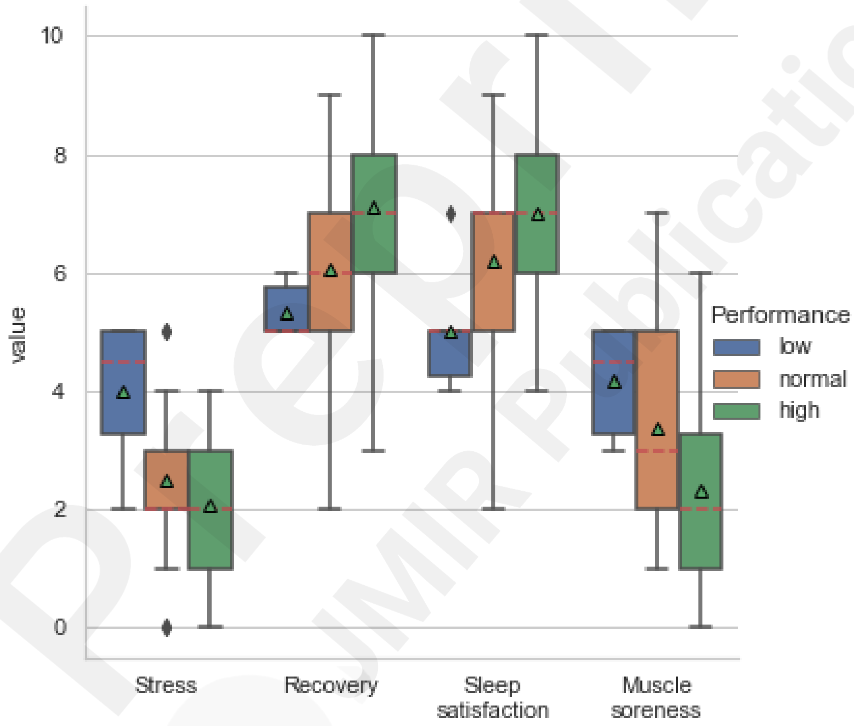


Figure 7: Boxplots of continuous self-reported measures by performance for all players during the complete study. Medians are marked with dashed red lines, means with green triangles.

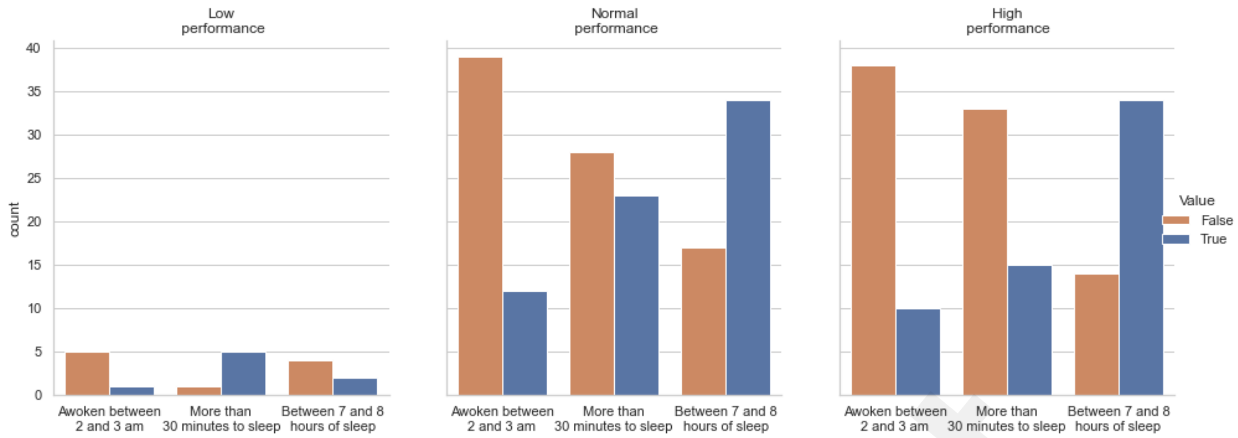


Figure 8: Value count of categorical sleep quality measures by performance for all players during the complete study. True values are in blue and False values in orange.

HRV Monitoring

Out of the 105 measures selected for self-reported wellness, 71 measures recorded by 9 players (age: mean 29.3, std 3.7; height: mean 181.6 cm, std 4.2; weight: mean 73.9 kg, std 2.7) were considered valid for HRV analysis due to having a minimum amount of five recordings in the previous 7-days period. Performance was also attached to these measures (high: 32; normal: 36; low: 3).

Figure 9 depicts the continuous HRV measures for all participants and for the whole duration of the study. Both CV (means 7.24%, 4.76% and 4.99% for low, normal and high performance) and SWC (means 1.17%, −0.22%, −0.46%) showed similar distributions for normal and high performance. Further analysis for low performance is discarded due to a low amount of measures.

Figure 10 depicts the categorical measures collected for HRV. Both CV_{ok} (*True/False* ratio 0%, 44.4% and 53.1% for low, normal and high performance) and SWC_{ok} (*True/False* ratio 100%, 61.1%, 46.9%) showed apparently different distributions depending on the performance level. For high performance, a balanced proportion among the True and False responses are shown for both CV_{ok} and SWC_{ok}. For normal performance, a CV value higher than 4.5% (CV_{ok}=*False*) was more present compared to high performance; also SWC was found ok more often (SWC_{ok}=*True*) than in high performance.

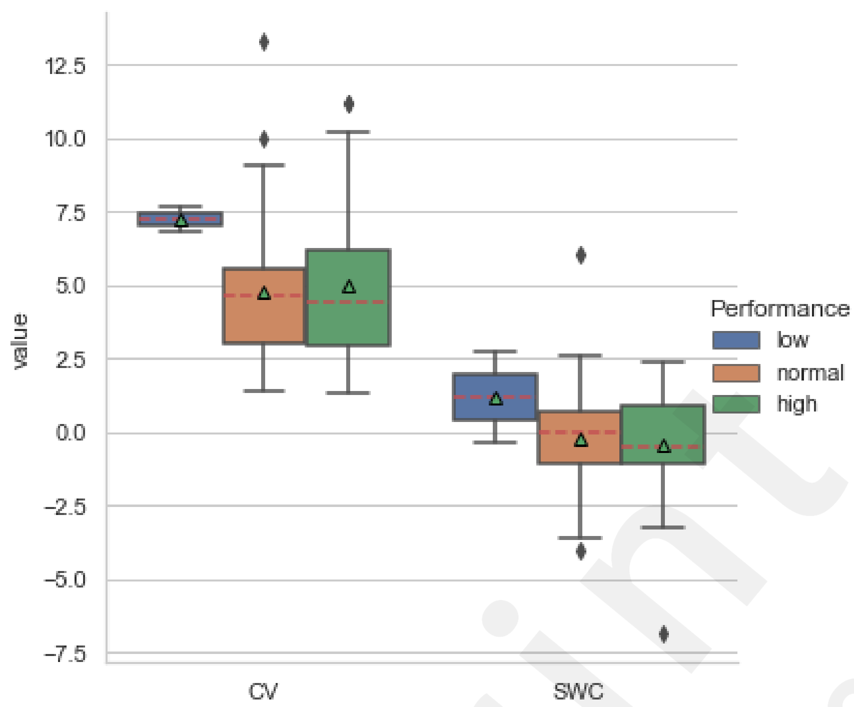


Figure 9: Boxplots of continuous HRV measures by performance. Medians are marked with dashed red lines; means are marked with green triangles.

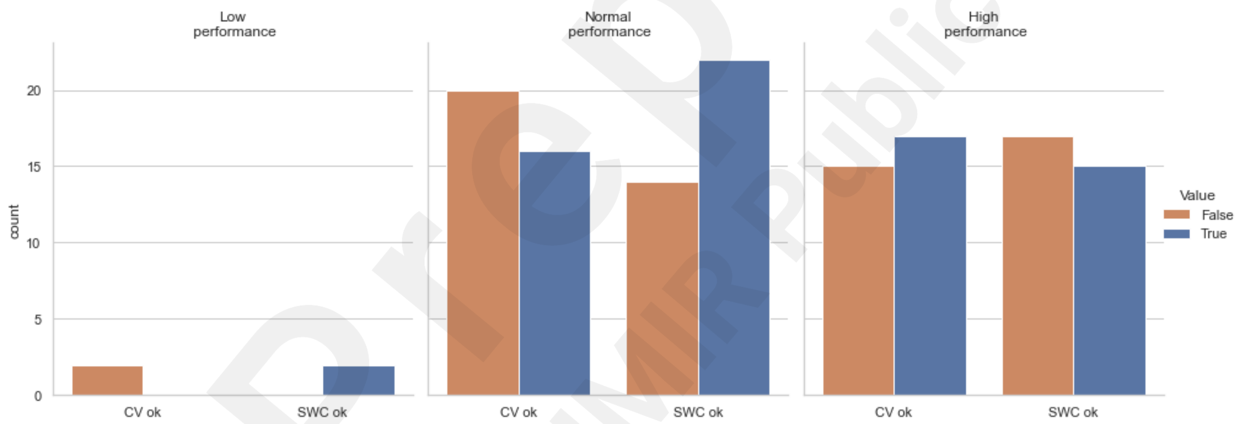


Figure 10: Value count of categorical HRV measures by performance.

Performance Relationship with Self-Reported Wellness and HRV Features

Correlation Analysis

We computed non-parametric Spearman correlation after not finding normality for the distributions of each feature. Table 1 shows the correlation of performance with the measures collected for the whole duration of the study. We identified small correlations for performance with sleep satisfaction, stress, and needing more than 30 minutes to sleep; and moderate with recovery and muscle soreness. Higher performance was associated with self-reported wellness (more recovery, less muscle soreness, less stress, more sleep satisfaction, not being awake between 2 and 3 am, not taking more than 30 minutes to sleep, and getting between 7 and 8 hours of sleep), and with HRV features (more times with CV_{ok}, less SWC, and less times with SWC_{ok}).

Table 1 also contains the two-sided *P* value for a hypothesis test whose null hypothesis is that two sets of data are uncorrelated. In other words, the *P* value roughly indicates the probability of an uncorrelated system producing datasets with a Spearman correlation at least as extreme as the one computed for our dataset. Recovery and muscle soreness showed the lowest and more significant *P* values ($P<.001$); next, sleep satisfaction ($P=.004$), stress ($P=.01$) and taking more than 30 minutes to sleep ($P=.03$).

Table 1: Spearman's correlation of performance with features. The P value (two-sided) is for a hypothesis test whose null hypothesis is that two sets of data are uncorrelated.

	Measures in entire training microcycles		
	r	**N**	***P* value**
Recovery	.33	105	<.001
Muscle soreness	−.34	105	<.001
Stress	−.24	105	.01
Sleep satisfaction	.28	105	.004
Awake between 2 and 3 am	−.01	105	.89
More than 30 minutes to sleep	−.22	105	.03
Between 7 and 8 hours of sleep	.12	105	.23
CV	−.04	71	.72
CV_{ok}	.10	71	.40
SWC	−.11	71	.35
SWC_{ok}	−.20	71	.09

Differences among performance levels

The low number of measures labelled with a *low* performance (6 measures, 3 of them with valid HRV data) hampered any comparison with the measures assigned to *normal* (51, 36) and *high* (48, 32). Hence, we compared all features distributions marked with *high* performance against *normal* performance. Table 2 shows the results for Kolmogorov-Smirnov, Anderson-Darling, and Mann-Whitney U tests. Table 3 shows the results for Anderson-Darling test.

We found significant differences between high and normal performance for most of the self-reported wellness features. These features were recovery (AD, $P=.002$; MWU, $P=.003$), muscle soreness (KS, $P=.04$; AD, $P=.001$; MWU, $P=.002$), stress (AD, $P=.03$), sleep satisfaction (AD, $P=.03$; MWU, $P=.02$), and needing more than 30 minutes to sleep (CS, $P=.02$). We found no significant differences for the HRV features.

Table 2: Comparison of high vs normal performance distributions across continuous features.

Statistics and P values for non-parametric tests Kolmogorov-Smirnov, Anderson-Darling, Mann-Whitney U.

	Kolmogorov-Smirnov		**Anderson-Darling**		**Mann-Whitney U**	
	statistic	***P* value**	**statistic**	***P* value**	**statistic**	***P* value**
Recovery	0.24	.09	5.95	.002	835	.003
Muscle soreness	0.27	.04	7.30	.001	809	.002
Stress	0.24	.10	2.41	.03	1032	.08
Sleep satisfaction	0.19	.31	2.58	.03	946	.02
CV	0.19	.50	−0.43	.25	553	.39
SWC	0.17	.66	−0.48	.25	559	.42

Table 3: Comparison of high vs normal performance distributions across categorical features. Statistics and P values for the non-parametric Chi-Squared test.

	Chi-Squared	
	statistic	***P* value**
Awake between 2 and 3 am	0.43	.51
More than 30 minutes to sleep	5.02	.02
Between 7 and 8 hours of sleep	0.64	.42
CV_{ok}	0.60	.44
SWC_{ok}	1.54	.21

Estimation of performance with wellness and HRV data

We built and studied four match performance estimation models using different sets of the features available. We used complete measures with valid self-reported wellness and HRV data to build the models. The accuracy results are shown in Table 4, and the importance of the features used in each model in Table 5.

The first model (M1) used all variables and provided 67.7% accuracy. The decision trees in a random forests model make predictions out of linear combinations of the features provided (eg, if recovery is higher than 7 and stress is lower than 2, then performance is high). Hence we trained M1 with all features, with the aim of testing if features marked as non-significant in the previous sections could still be relevant in combination with the other variables. With this approach we allowed the model to combine physiological features from HRV monitoring with features from self-reported wellness.

To gain understanding of the M1 model and its use of variables, we extracted the feature importance for it (see Table 5). The most relevant features (with an accumulated importance higher than 80%) were CV, muscle soreness, recovery, SWC and stress, in this order. On the other hand, the sleep quality features ‘having between 7 and 8 hours of sleep’ and ‘being awake between 2 and 3 am trying to sleep’ scored very low importance (less than 3%). During the fine tuning of this model and in further iterations of the models trained, these two sleep quality features reduced the performance of the model trained. Therefore, and for the sake of clarity, such features are removed from the rest of models presented.

Features CV_{ok} and SWC_{ok} were also found with very little importance for M1, with less than 1%

weight. This happened due to two main reasons. First, such variables present very high collinearity (higher than 0.8) with CV and SWC, respectively; and second, since CV and SWC are continuous variables and different thresholds can be defined in them, they are most likely to be used by the model during its training. Therefore, and in order to leverage such continuous property in CV and SWC, we built the M2 model using self-reported wellness and continuous HRV features. However, and taking into account that CV_{ok} and SWC_{ok} were defined according to already validated criteria [23], we also built the M3 model using self-reported wellness and categorical (*True/False*) HRV features.

The M2 model scored 64.6%, with a decreased performance compared to M1. Continuous HRV features were the most important features for M2. Compared to M1, SWC increased 8.1% of relevance, whilst CV decreased 2.5%. Therefore, SWC, CV, recovery, muscle soreness, and stress, were the most important variables for M2, in this order.

Conversely, M3 scored 73.4%, outperforming M1 and M2. Self-reported wellness features were the most important for M3. Compared to M1 and M2, each one of them increased its relevance, finding recovery, muscle soreness, stress, sleep satisfaction and ‘taking more than 30 minutes to sleep’ as the most important features for M3, in this order. The relevance for SWC_{ok} (7.6%) and CV_{ok} (4.7%) is lower compared to the continuous SWC (22.2%) and CV (18.0%) in M2, but higher compared to SWC_{ok} (0.6) and CV_{ok} (0.4) in M1.

Finally, we built the M4 model only with self-reported wellness features. M4 scored 71.6%, outperforming M1 and M2, but below M3. The distribution of feature importance was very similar to the one found in M3, only with a very clear difference for sleep satisfaction. Sleep satisfaction increased its relevance in 15%, reaching up to 26.5%, when compared to the one found in M3 (11.5%). The most important features for M4 were recovery, sleep satisfaction, muscle soreness and stress.

Table 4: Average performance for each model using 10-fold-validation.

	M1: All features	**M2: Self-reported wellness and continuous HRV features**	**M3: Self-reported wellness and categorical HRV features**	**M4: Self-reported wellness features only**
Accuracy (%)	67.7	64.6	73.4	71.6

Table 5: Feature importance across models. A dash means the feature was not used in that model.

	M1: All features (%)	**M2: Self-reported wellness and continuous HRV features (%)**	**M3: Self-reported wellness and categorical HRV features (%)**	**M4: Self-reported wellness features only (%)**
Recovery	18.1	18.9	31.3	30.6
Muscle soreness	19.4	18.3	20.0	21.0
Sleep satisfaction	5.6	7.1	11.5	26.5
Stress	10.1	10.0	15.4	14.7
More than	6.1	5.5	9.4	7.2

30 minutes to sleep				
Between 7 and 8 hours of sleep	2.6	-	-	-
Awake between 2 and 3 am	2.4	-	-	-
CV	20.5	18.0	-	-
SWC	14.1	22.2	-	-
CV_{ok}	0.4	-	4.7	-
SWC_{ok}	0.6	-	7.6	-

Discussion

In this work, we presented the mobile app Readiness Soccer as a useful tool to monitor internal load and HRV in professional soccer players. We successfully conducted a pilot study in a very challenging real scenario: the season continuation after the COVID-19 lockdown in Spain. We found that self-reported wellness and HRV features were useful to monitor internal load and relate it to match performance. To the best of our knowledge, Readiness Soccer is the first attempt to provide a mobile HRV and wellness monitoring solution to professional soccer players.

Principal Findings

Usability. Usability results showed that Readiness Soccer has a high level of user-friendliness, acceptability and ease of use from the perspective of the professional soccer players asked. The overall SUS score of Readiness Soccer was found to be, not only over the minimum required of 68 points [32] but also over the industrial goal of 80 points to provide an above average usability experience [42]. This result also stands out when exclusively compared to the health domain standards, for which the SUS was found as the most commonly used usability evaluation in a literature review [43], as well as in recent work [44,45]. It is of particular interest that such results were achieved during the very demanding context of the post-COVID-19 lockdown, a scenario in which the athletic performance of soccer players was reduced [46] and their injury risk raised [1]. Hence, a context in which the daily monitoring of players' health status is essential.

Description of data collected. Despite the good usability results, the players gathered a relatively low number of measures during the study. In fact, they had to be reminded to use the app periodically throughout the study (hence the oscillations). We argue that this responds, not only to the difficulties presented by the creation of new habits [47], such as when implementing new nutrition strategies in elite athletes [48], but also to the impact of COVID-19 lockdown in the daily habits of the population, such as dietary [49(p)] and consuming habits [50,51].

Measure counts also varied by player. The demographics of the data collected suggest that younger players were less likely to follow the protocol consistently. This could be influenced by a negative emotional state with respect to the COVID-19 pandemia and its lockdowns [3], which could be more acute in younger players with low social support.

A small amount of data was related to low performance, therefore we focused the analysis on normal and high performance. This may suggest that the overall training/recovery cycles were handled correctly by the team, and that the coach selected the players better prepared for high match performance. Moreover, the medians and means of the CV distributions were very close to the upper threshold of 4.5% defined in [23], hence implying a good overall state of the players. In fact, out of three low performance observations in the 10 matches, two were monitored and presented in the results. This may discard the possibility of players avoiding to register their status due to low overall wellness.

The 10 matches played in the post-lockdown continuation of the season were characterized by very short training microcycles, with 72h as the most common scenario. Monitoring internal load in such intense periods, in which the risk for overtraining and injuries rise [52,53], it is crucial to understand how well players are managing their recovery processes.

Performance relationship with self-reported wellness and HRV features. It is when professional soccer players are under optimal training conditions (ie, adequate adaptation and control of training loads) that they are most likely to perform at their best [6]. This is why the literature focuses on studying the relevance of tools like HRV and self-reported wellness to assess internal load [10,11] and match fatigue [12–15]. Our work assumes the validity of these tools, thus our approach links HRV and self-reported wellness directly with performance. The complexities of defining performance in professional soccer are widely addressed in the literature, being addressed from physical, psychological and cognitive perspectives [54–56]. However, we consider that challenge outside the scope of this article, and that our simple evaluations of match performance (based on impressions of in-match technical-tactic performance of players) were sufficient to start exploring the association of performance with HRV and self-reported wellness. To the best of our knowledge, this is the first work that addresses the relation of HRV and wellness directly with performance instead of training load and fatigue.

We found five of the seven self-reported wellness features to be significantly associated with performance from statistical and machine learning perspectives. Recovery and muscle soreness were moderately correlated with performance, and found as the best predictors of high performance. This matches the results of previous studies focusing on training load and fatigue [10,11,57], that is why the medical staff of the teams often focus on alleviating muscle soreness and improving overall recovery [58]. Sleep satisfaction aspects and stress were lowly correlated with performance, yet they were also relevant predictors of high performance. This relationship is aligned with previous research [10,11,59,60], hence the efforts on providing sleep hygiene and recovery strategies [60,61].

Small correlations were found for performance with three of the four available HRV parameters. However, the sample of HRV measures was not high enough to obtain significant correlations. The only HRV parameter distribution close to significance was for the categorical SWC_{ok} (P=0.09). Hence, results involving HRV should be interpreted carefully, especially regarding its role in modeling , which may suffer from overfitting.

The continuous HRV parameters SWC and CV presented very similar distributions for normal and high performance, suggesting that there are no differences among such performance levels. In fact, the medians and means of the CV distributions were very close to the upper threshold of 4.5% defined in [23], hence implying a good overall state of the players during the competitive season. Conversely, continuous SWC and CV were found as very relevant predictors of high performance when modeling. This matches the results of using day-by-day analysis of HRV of previous research [14,16], which random forests may be able to leverage taking into account the distance of the measures to the match day.

The categorical HRV measures SWC_{ok} and CV_{ok} pictured different ratios for normal and high performance, with a more balanced *True/False* ratio for high performance. This may indicate faster and better suited workload-recovery ratios than for normal performance. However, the power of categorical HRV features is ultimately reflected in the modeling process. SWC_{ok} and CV_{ok} scored an accumulated importance of 12.3% in the model with highest accuracy (M3), similar to wellness parameters like stress and sleep satisfaction. This lower relevance, compared to the continuous SWC and CV, may be influenced by the threshold itself, which is taken from cycling [24,30], a sport involving more aerobic activity.

Despite this limitation, we argue that, due to their simplicity, this categorical HRV features may be more useful in this assessment and performance prediction context. Since HRV may be easily affected by several lifestyle and external factors [62], HRV parameters need to be combined with other wellness factors to be interpreted correctly. This is why the decision trees building the random forests models may be one of the best choices to leverage such relationships.

The discussed role of wellness and HRV features across models support that the combination of selected self-reported wellness (recovery, muscle soreness, stress, sleep satisfaction, not needing more than 30 minutes to sleep) and categorical HRV features (SWC_{ok} and CV_{ok}) built the best model (M3). We believe this combination of features leverages the power of two complementary perspectives of the players' health status: perceived and subjective for self-reported wellness; and physiological and subjective for HRV. An alternative model, M4, only using selected self-reported features was also able to report a high degree of accuracy, highlighting the power of sole self-reported wellness and making them useful even in the absence of HRV data. This matches results already found in the literature exposing the advantages of Hooper Index over LnRMSSD [14].

Limitations

The results of this study should be considered preliminary. First, the study was conducted with a relatively small sample of professional male soccer players, and influenced by the COVID-19 pandemic and its lockdowns during the 6 weeks and 10 matches played. Future work should include more players from different teams and gender in a more normalized competition scenario to generalize results. Second, the amount of data available was relatively small due to inconsistent adherence to the protocol. Thus, future research should address reducing the necessary time to monitor HRV. In fact, recent work has shown that 1-minute measurements of RMSSD preceded by 1 or more minutes of pre-stabilization are valid to measure HRV in professional male soccer players [63]. Third, we used a subjective evaluation of the technical-tactic performance of the players as target variable to relate to internal load. Future research should also include objective parameters related to performance like external load measures in its definition. Finally, we reported the results obtained from only one classification method. Our choice was based on the interpretability of the random forests model and the high performance of the method in our dataset. However, the results may vary with a different data sample.

Conclusions

Our findings highlight the feasibility of using mobile phone and wearable sensors to assess the internal load of professional soccer players during the competitive season. We found the app Readiness Soccer to be useful to monitor internal load during a very intense competitive scenario such as the season continuation after the COVID-19 lockdown. We established preliminary patterns of HRV and self-reported wellness for higher and lower levels of subjective perceived match performance of players. HRV and self-reported wellness data were useful to monitor the evolution of professional soccer players' internal load and to predict match performance levels out of measures in a training microcycle.

Our work also highlights opportunities for long-term monitoring of soccer players during competition, as well as real-time interventions aimed at the early management of overtraining and boosting of individual performance. Information from Readiness Soccer could be integrated into the electronic health records of the teams to make more informed decisions when adjusting training types and intensity. Prediction models could provide the automatic detection of a worsening status, thus enabling just-in-time adaptive interventions aimed at improving (or, at least, maintaining) the player's condition. Such personalized interventions could improve player's performance by minimizing the risk of injuries and overtraining.

Acknowledgements

This research was funded by the Andalucian Ministry of Economic Transformation, Industry, Knowledge and Universities under project MONITOR-COVID (CV20-29556) and the Spanish Ministry of Science, Innovation, and Universities (MICINN) under Projects PGC2018-098813-B-C31 and RTI2018-101674-B-I00. Salvador Moreno-Gutierrez holds a FPU Spanish Grant FPU16/04201. Paula Postigo-Martin holds a FPU Spanish Grant FPU17/00939.

Conflicts of Interest

None declared.

Bibliography

1. Seshadri DR, Thom ML, Harlow ER, Drummond CK, Voos JE. Case Report: Return to Sport Following the COVID-19 Lockdown and Its Impact on Injury Rates in the German Soccer League. *Front Sports Act Living*. 2021;3. doi:10.3389/fspor.2021.604226
2. De Souza DB, González-García J, Campo RL-D, et al. Players' physical performance in LaLiga across the season: Insights for competition continuation after COVID-19. *Biol Sport*. 2021;38(1):3-7. doi:10.5114/BIOLSPORT.2020.96856
3. Leguizamo F, Olmedilla A, Núñez A, et al. Personality, Coping Strategies, and Mental Health in High-Performance Athletes During Confinement Derived From the COVID-19 Pandemic. *Front Public Health*. 2021;8. doi:10.3389/fpubh.2020.561198
4. Oliveira R, Brito JP, Martins A, et al. In-season internal and external training load quantification of an elite European soccer team. *PloS One*. 2019;14(4):e0209393. doi:10.1371/journal.pone.0209393
5. Bourdon PC, Cardinale M, Murray A, et al. Monitoring Athlete Training Loads: Consensus Statement. *Int J Sports Physiol Perform*. 2017;12(Suppl 2):S2161-S2170. doi:10.1123/IJSPP.2017-0208
6. Miguel M, Oliveira R, Loureiro N, García-Rubio J, Ibáñez SJ. Load Measures in Training/Match Monitoring in Soccer: A Systematic Review. *Int J Environ Res Public Health*. 2021;18(5):2721. doi:10.3390/ijerph18052721
7. Wallace LK, Slattery KM, Coutts AJ. The ecological validity and application of the session-RPE method for quantifying training loads in swimming. *J Strength Cond Res*. 2009;23(1):33-38. doi:10.1519/JSC.0b013e3181874512
8. Halson SL. Monitoring training load to understand fatigue in athletes. *Sports Med Auckl NZ*. 2014;44 Suppl 2:S139-147. doi:10.1007/s40279-014-0253-z
9. Hooper SL, Mackinnon LT. Monitoring overtraining in athletes. Recommendations. *Sports Med Auckl NZ*. 1995;20(5):321-327. doi:10.2165/00007256-199520050-00003
10. Moalla W, Fessi MS, Farhat F, Nouira S, Wong DP, Dupont G. Relationship between daily training load and psychometric status of professional soccer players. *Res Sports Med*. 2016;24(4):387-394. doi:10.1080/15438627.2016.1239579
11. Clemente FM, Mendes B, Nikolaidis PT, Calvete F, Carriço S, Owen AL. Internal training load and its longitudinal relationship with seasonal player wellness in elite professional soccer. *Physiol Behav*. 2017;179:262-267. doi:10.1016/j.physbeh.2017.06.021
12. Thorpe RT, Strudwick AJ, Buchheit M, Atkinson G, Drust B, Gregson W. Monitoring Fatigue During the In-Season Competitive Phase in Elite Soccer Players. *Int J Sports Physiol Perform*. 2015;10(8):958-964. doi:10.1123/ijspp.2015-0004
13. Fessi MS, Nouira S, Dellal A, Owen A, Elloumi M, Moalla W. Changes of the psychophysical state and feeling of wellness of professional soccer players during pre-season and in-season periods. *Res Sports Med*. 2016;24(4):375-386. doi:10.1080/15438627.2016.1222278
14. Rabbani A, Clemente FM, Kargarfard M, Chamari K. Match Fatigue Time-Course Assessment Over Four Days: Usefulness of the Hooper Index and Heart Rate Variability in Professional Soccer Players. *Front Physiol*. 2019;10. doi:10.3389/fphys.2019.00109

15. Tibana RA, Sousa NMF de, Prestes J, Feito Y, Ferreira CE, Voltarelli FA. Monitoring Training Load, Well-Being, Heart Rate Variability, and Competitive Performance of a Functional-Fitness Female Athlete: A Case Study. *Sports Basel Switz*. 2019;7(2). doi:10.3390/sports7020035
16. Sánchez-Sánchez J, Botella J, Felipe Hernández JL, et al. Heart rate variability and physical demands of in-season youth elite soccer players. *Int J Environ Res Public Health*. 2021;18(4):1-8. doi:10.3390/ijerph18041391
17. Sobolewski EJ. The Relationships between Internal and External Load Measures for Division I College Football Practice. *Sports Basel Switz*. 2020;8(12). doi:10.3390/sports8120165
18. Modric T, Versic S, Sekulic D. Relations of the weekly external training load indicators and running performances in professional soccer matches. *Sport Mont*. 2021;19(1):31-37. doi:10.26773/SMJ.210202
19. Swallow WE, Skidmore N, Page RM, Malone JJ. An examination of in-season external training load in semi-professional soccer players: considerations of one and two match weekly microcycles. *Int J Sports Sci Coach*. 2021;16(1):192-199. doi:10.1177/1747954120951762
20. Gamonales JM, León K, Rojas-Valverde D, Sánchez-Ureña B, Muñoz-Jiménez J. Data mining to select relevant variables influencing external and internal workload of elite blind 5-A-side soccer. *Int J Environ Res Public Health*. 2021;18(6):1-11. doi:10.3390/ijerph18063155
21. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*. 2013;310(20):2191-2194. doi:10.1001/jama.2013.281053
22. Gehlbach H, Barge S. Anchoring and Adjusting in Questionnaire Responses. *Basic Appl Soc Psychol*. Published online 2012. Accessed October 8, 2020. https://dash.harvard.edu/handle/1/11393840
23. Javaloyes A, Sarabia JM, Lamberts RP, Moya-Ramon M. Training Prescription Guided by Heart-Rate Variability in Cycling. *Int J Sports Physiol Perform*. 2019;14(1):23-32. doi:10.1123/ijspp.2018-0122
24. Plews DJ, Laursen PB, Stanley J, Kilding AE, Buchheit M. Training Adaptation and Heart Rate Variability in Elite Endurance Athletes: Opening the Door to Effective Monitoring. *Sports Med*. 2013;43(9):773-781. doi:10.1007/s40279-013-0071-8
25. Peltola MA. Role of Editing of R–R Intervals in the Analysis of Heart Rate Variability. *Front Physiol*. 2012;3. doi:10.3389/fphys.2012.00148
26. Giles DA, Draper N. Heart Rate Variability During Exercise: A Comparison of Artefact Correction Methods. *J Strength Cond Res*. 2018;32(3):726-735. doi:10.1519/JSC.0000000000001800
27. Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. *Front Public Health*. 2017;5:258. doi:10.3389/fpubh.2017.00258
28. Hopkins WG. How to Interpret Changes in an Athletic Performance Test. *Sportscience*. 2004;8:1-7.
29. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med Auckl NZ*. 2000;30(1):1-15. doi:10.2165/00007256-200030010-00001
30. Plews DJ, Laursen PB, Meur YL, Hausswirth C, Kilding AE, Buchheit M. Monitoring Training With Heart-Rate Variability: How Much Compliance Is Needed for Valid Assessment? *Int J Sports Physiol Perform*. 2014;9(5):783-790. doi:10.1123/ijspp.2013-0455
31. Brooke J. SUS: a “quick and dirty” usability scale. In: *P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.). Usability Evaluation in Industry. London: Taylor and Francis*.
32. Sauro J. *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. CreateSpace Independent Publishing Platform; 2011.
33. Hopkins KD, Glass GV, Hopkins BR. *Basic Statistics for the Behavioral Sciences, 2nd Ed*. Prentice-Hall, Inc; 1987:x, 307.
34. Harris JE, Boushey C, Bruemmer B, Archer SL. Publishing Nutrition Research: A Review of Nonparametric Methods, Part 3. *J Am Diet Assoc*. 2008;108(9):1488-1496. doi:10.1016/j.jada.2008.06.426
35. NIST/SEMATECH e-Handbook of Statistical Methods. Accessed March 2, 2021. https://www.itl.nist.gov/div898/handbook/index.htm
36. Tests for Nominal Scale Data: Chi-Square and Fisher Exact Test. In: *Nonparametric Statistics for Non-Statisticians*. John Wiley & Sons, Inc.; 2011:155-191. doi:10.1002/9781118165881.ch8
37. Aggarwal CC. *Data Mining*. Springer International Publishing; 2015. doi:10.1007/978-3-319-14142-8
38. McKinney W. Data Structures for Statistical Computing in Python. In: Walt S van der, Millman J, eds. *Proceedings of the 9th Python in Science Conference*. ; 2010:51-56.

39. Waskom M, team the seaborn development. *Mwaskom/Seaborn*. Zenodo; 2020. doi:10.5281/zenodo.592845
40. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods*. 2020;17:261-272. doi:10.1038/s41592-019-0686-2
41. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011;12(85):2825-2830.
42. Lewis JR, Sauro J. Item Benchmarks for the System Usability Scale. *J Usability Stud*. 2018;13(3):158-167.
43. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. *Disabil Rehabil Assist Technol*. Published online January 8, 2020:1-10. doi:10.1080/17483107.2019.1701103
44. Bailon C, Damas M, Pomares H, Banos O. SPIRA: an automatic system to support lower limb injury assessment. *J Ambient Intell Humaniz Comput*. Published online February 23, 2018:1-13. doi:10.1007/s12652-018-0722-6
45. Moreno-Gutierrez S, Postigo-Martin P, Damas M, et al. ATOPE+: An mHealth System to Support Personalized Therapeutic Exercise Interventions in Patients With Cancer. *IEEE Access*. 2021;9:16878-16898. doi:10.1109/ACCESS.2021.3049398
46. de Albuquerque Freire L, Tannure M, Sampaio M, et al. COVID-19-Related Restrictions and Quarantine COVID-19: Effects on Cardiovascular and Yo-Yo Test Performance in Professional Soccer Players. *Front Psychol*. 2020;11:589543. doi:10.3389/fpsyg.2020.589543
47. Arlinghaus KR, Johnston CA. The Importance of Creating Habits and Routine. *Am J Lifestyle Med*. 2019;13(2):142-144. doi:10.1177/1559827618818044
48. Heaney S, O'Connor H, Naughton G, Gifford J. Towards an Understanding of the Barriers to Good Nutrition for Elite Athletes. *Int J Sports Sci Coach*. 2008;3(3):391-401. doi:10.1260/174795408786238542
49. Rodríguez-Pérez C, Molina-Montes E, Verardo V, et al. Changes in Dietary Behaviours during the COVID-19 Outbreak Confinement in the Spanish COVIDiet Study. *Nutrients*. 2020;12(6):1730. doi:10.3390/nu12061730
50. Salmerón L, Arfé B, Avila V, et al. READ-COGvid: A Database From Reading and Media Habits During COVID-19 Confinement in Spain and Italy. *Front Psychol*. 2020;11. doi:10.3389/fpsyg.2020.575241
51. Sheth J. Impact of Covid-19 on consumer behavior: Will the old habits return or die? *J Bus Res*. 2020;117:280-283. doi:10.1016/j.jbusres.2020.05.059
52. Pedersen L, Lindberg J, Lind RR, Rasmusen H. Reopening elite sport during the COVID-19 pandemic: Experiences from a controlled return to elite football in Denmark. *Scand J Med Sci Sports*. 2021;31(4):936-939. doi:10.1111/sms.13915
53. García-Aliaga A, Marquina M, Cordón-Carmona A, Sillero-Quintana M, de la Rubia A, Refoyo Román I. Comparative Analysis of Soccer Performance Intensity of the Pre-Post-Lockdown COVID-19 in LaLiga™. *Int J Environ Res Public Health*. 2021;18(7). doi:10.3390/ijerph18073685
54. Slimani M, Chamari K, Miarka B, Del Vecchio FB, Chéour F. Effects of Plyometric Training on Physical Fitness in Team Sport Athletes: A Systematic Review. *J Hum Kinet*. 2016;53:231-247. doi:10.1515/hukin-2016-0026
55. Mielgo-Ayuso J, Calleja-Gonzalez J, Del Coso J, Urdampilleta A, León-Guereño P, Fernández-Lázaro D. Caffeine Supplementation and Physical Performance, Muscle Damage and Perception of Fatigue in Soccer Players: A Systematic Review. *Nutrients*. 2019;11(2). doi:10.3390/nu11020440
56. Rowat O, Fenner J, Unnithan V. Technical and physical determinants of soccer match-play performance in elite youth soccer players. *J Sports Med Phys Fitness*. 2017;57(4):369-379. doi:10.23736/S0022-4707.16.06093-X
57. Silva J, Rumpf M, Hertzog M, et al. Acute and Residual Soccer Match-Related Fatigue: A Systematic Review and Meta-analysis. *Sports Med Auckl NZ*. 2018;48(3):539-583. doi:10.1007/s40279-017-0798-8
58. Gimenes SV, Marocolo M, Pavin LN, et al. Compression Stockings Used During Two Soccer Matches Improve Perceived Muscle Soreness and High-Intensity Performance. *J Strength Cond Res*. Published online February 14, 2019. doi:10.1519/JSC.0000000000003048
59. Nédélec M, Halson S, Abaidia A-E, Ahmaidi S, Dupont G. Stress, Sleep and Recovery in Elite Soccer: A Critical Review of the Literature. *Sports Med Auckl NZ*. 2015;45(10):1387-1400. doi:10.1007/s40279-

015-0358-z

60. Nédélec M, Halson S, Delecroix B, Abaidia A-E, Ahmaidi S, Dupont G. Sleep Hygiene and Recovery Strategies in Elite Soccer Players. *Sports Med Auckl NZ*. 2015;45(11):1547-1559. doi:10.1007/s40279-015-0377-9
61. Ajjimaporn A, Ramyarangsi P, Siripornpanich V. Effects of a 20-min Nap after Sleep Deprivation on Brain Activity and Soccer Performance. *Int J Sports Med*. 2020;41(14):1009-1016. doi:10.1055/a-1192-6187
62. Sammito S, Böckelmann I. Factors influencing heart rate variability. *Int Cardiovasc Forum J*. 2016;6(0):18-22. doi:10.17987/icfj.v6i0.242
63. Forner-Llacer F, Aranda-Malaves R, Aranda Malavés R, Calleja-Gonzalez J, Perez-Turpin JA, Gonzalez-Rodenas J. Minimal Stabilization Time for Ultra-short Heart Rate Variability Measurements in Professional Soccer. *Int J Sports Med*. 2020;41(14):1032-1038. doi:10.1055/a-1186-1316