UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Arian Skoki

Data-Driven Assessment of Player Performance and Recovery in Soccer

DOCTORAL DISSERTATION

Rijeka, 2025.

UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Arian Skoki

Data-Driven Assessment of Player Performance and Recovery in Soccer

DOCTORAL DISSERTATION

Thesis Supervisor: prof. Ivan Štajduhar, PhD
Thesis Co-supervisor: prof. Jonatan Lerga, PhD

Rijeka, 2025.

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET

Arian Skoki

Procjena karakteristika i oporavka igrača u nogometu zasnovana na podatcima

DOKTORSKI RAD

Mentor: prof. dr. sc. Ivan Štajduhar
Komentor: prof. dr. sc. Jonatan Lerga

Rijeka, 2025.

Abstract

Data collection in soccer encompasses a variety of methods, ranging from simple daily questionnaires that monitor subjective well-being to more sophisticated approaches such as wearable sensors that track external load during training and matches. While subjective measures like wellness questionnaires and Rate of Perceived Exertion (RPE) offer insights into players' internal states, they fall short of providing comprehensive assessments of players' physical fitness and performance. Consequently, the use of wearable sensors has become prevalent, offering detailed metrics such as distance covered, sprints, accelerations, and energy expenditure. These metrics enable coaches to better manage training loads, plan sessions and monitor the return-to-play (RTP) process after injuries.

However, traditional methods that aggregate data across entire matches often hide the variations in players' performance throughout the game. This dissertation addresses this issue by implementing a minute-by-minute analysis of player physical intensity using wearable data, revealing more details about players' fitness levels. By examining how physical performance fluctuates within different game contexts, the study provides a more accurate reflection of a player's condition and readiness.

The relationship between cognitive load (CL) and physical performance is another critical aspect explored in this work. While CL's impact on tactical and technical aspects is well-studied, its effects on physical performance, particularly in real-world settings, remain unexplored. Using data from the NASA-TLX questionnaire, this research investigates how varying levels of CL influence physical metrics such as total distance, high-speed running, and deceleration. Moreover, the study applies machine learning (ML) algorithms to classify players into motivation clusters, offering insights into how different motivational factors might affect performance under cognitive stress.

Injury recovery is a crucial area in sports, and the RTP process is crucial in ensuring players return to competition safely. This dissertation advances the understanding of injury management by developing ML models to predict recovery duration, using data collected by medical staff. These models are compared with traditional expert-based predictions, demonstrating that the integration of expert input with ML techniques enhances the accuracy of recovery time estimates.

In addition to injury recovery, the research focuses on developing individualized fatigue and recovery profiles. These profiles are essential for conditioning coaches to tailor training

regimens and ensure that players return to their pre-injury fitness levels. By providing a detailed, data-driven approach to monitoring physical recovery, this dissertation contributes valuable tools for optimizing player performance and reducing the risk of re-injury.

This work provides a comprehensive framework for soccer performance and recovery analysis, integrating wearable sensor data, cognitive load assessments, and ML techniques. The findings offer practical applications for coaches and sports scientists, helping them to better understand and manage the complex dynamics of player fitness and performance throughout the season.

Keywords: soccer performance analysis, wearable sensors, player intensity monitoring, cognitive load, machine learning, injury recovery prediction, muscle fatigue modelling, return-to-play, physical performance assessment, sports analytics

Sažetak

Prikupljanje podataka u nogometu obuhvaća različite metode, od jednostavnih dnevnih upitnika koji prate subjektivan dojam opterećenja pojedinca, do skupljih tehnoloških pristupa poput nosivih senzora koji prate vanjsko opterećenje sportaša tijekom treninga i utakmica. Dok subjektivne mjere poput upitnika za praćenje dobrobiti i ocjene doživljenog napora (RPE) pružaju uvid u unutarnje stanje igrača, one nisu dovoljne za sveobuhvatnu sliku karakteristika igrača. Iz tog razloga je upotreba nosivih senzora postala raširena, pružajući podatke o prijeđenoj udaljenosti, broju sprintova, ubrzanjima i potrošnji energije. Ove informacije omogućavaju trenerima bolju kontrolu opterećenja tijekom treninga, planiranje istih i praćenje procesa povratka na teren nakon ozljeda.

Međutim, tradicionalne metode prikupljaju podatke na razini cijele utakmice čime se onemogućava pregled izvedbe igrača tijekom samog trajanja utakmice. Ova disertacija nastoji riješiti taj problem primjenom podatkovne analize intenziteta igrača minutu-kroz-minutu koristeći podatke nosivih senzora. Pregledom promjena u fizičkim izvedbama tijekom različnih trenutaka u igri, pristup pruža precizniji odraz stanja i spremnosti igrača.

Odnos između kognitivnog opterećenja i fizičkog opterećenja igrača također je jedan od fokusa ovog rada. Dok je utjecaj kognitivnog opterećenja na taktičke i tehničke izvedbe igrača dobro istražen, njegov utjecaj na njihovo fizičko opterećenje, osobito tijekom utakmica, ostaje neistražen. Koristeći podatke iz NASA-TLX upitnika, ispituje se kako različite razine kognitivnog opterećenja utječu na fizičke mjere poput ukupne prijeđene udaljenosti, sprinta, ubrzavanja i usporavanja. Nadalje, primjenjuju se algoritmi strojnog učenja za klasifikaciju igrača u motivacijske grupe, pružajući uvid u to kako različiti motivacijski čimbenici mogu utjecati na izvedbe pod mentalnim opterećenjem.

Oporavak od ozljeda je jedan od ključnih čimbenika u sportu, a dobro planiranje procesa oporavka je presudno za siguran povratak igrača na teren. Razvijanjem modela strojnog učenja koji predviđaju trajanje oporavka nakon mišićnih ozljeda, poboljšava se točnost u samom planiranju. Modeli strojnog učenja uspoređeni su s procjenama stručnjaka, gdje se pokazalo kako je upravo kombinacija modela i znanja stručnjaka najbolja za procjenu vremena oporavka.

Kondicijski treneri moraju dobro poznavati igrače kako bi iz svakog izvukli maksimum i pripremili ih za sezonu. Iz tog razloga, vrlo je korisno imati uvid u individualni profil fizičke spreme igrača kako bi treneri mogli prilagoditi program treninga. Podatkovni pristup doprinosi lakšem praćenju opterećenja i smanjivanju rizika igrača od ozljede.

Korištenjem podataka iz različitih izvora, kreirane su metode koje treneri, analitičari, i sve osobe koje rade u nogometu mogu koristiti kako bi dobili detaljnije informacije o svojim igračima. Podatci su dobiveni korištenjem nosivih senzora i upitnika za procjenu kognitivnog opterećenja te su isti iskorišteni na različiti načine uz primjenu tehnika strojnog učenja i optimizacije. Cilj ovakvog pristupa je razvijanje objektivnih metoda koje se mogu automatizirati i olakšati posao ljudima koji rade u nogometu.

Ključne riječi: nogomet, podatkovna analiza, senzori, praćenje intenziteta igrača, kognitivno opterećenje, strojno učenje, oporavak od ozljede, matematičko modeliranje, ocjena fizičke izvedbe, analiza podataka u sportu

Contents

Abstract I
Sažetak III

I Introduction **1**

1 Introduction **3**

1.1 Overview of Data Collection Techniques in Soccer 3
1.2 Metrics and Methods for Evaluating Physical Performance 4
1.3 Approaches to Monitoring Injury and Physical Recovery 5
1.4 Dissertation Structure 6

2 Physical Performance Assessment **7**

2.1 Minute-by-Minute Intensity Evaluation in Soccer Matches 7
2.2 Developing Player Profiles to Inform Starting vs. Substitute Decisions 10
2.2.1 Data Collection and Preprocessing 11
2.2.2 Mathematical Modeling for Player Energy Expenditure 11
2.3 Impact of Cognitive Load on Physical Performance 13

3 Return-to-Play Data-Aided Systems **17**

3.1 Estimation of Muscle Injury Duration Using machine learning (ML) 17
3.1.1 Dataset Overview and Injury Classification 18
3.1.2 Assessment of Model Variants and Expert-Integrated Approaches . . . 20
3.2 Modelling Muscle Fatigue and Recovery with Optimization Methods 22

4 Conclusion **27**

4.1 Main Contributions 27
4.2 Future Work 28

5 Summary of Papers **33**

A Extended Energy-Expenditure Model in Soccer: Evaluating Player Performance in the Context of the Game 33

B Revolutionizing Soccer Injury Management: Predicting Muscle Injury Recovery Time Using ML . . . 34
C Building Individual Player Performance Profiles According to Pre-Game Expectations and Goal Difference in Soccer . . . 35
D Exploring the Impact of the Perceived Cognitive Load on the Physical Performance in Soccer . . . 36
E Enhancing Biophysical Muscle Fatigue Model in the Dynamic Context of Soccer 37
Bibliography . . . 48
List of Figures . . . 50
List of Publications . . . 51
Curriculum Vitae . . . 53

II Included Publications 55

Part I

Introduction

Chapter 1

Introduction

1.1 Overview of Data Collection Techniques in Soccer

There are a couple of methods that are used to collect the data in soccer. The simplest to implement is daily questionnaires [80]. These usually include wellness questionnaires which provide information about an individual's levels of sleep, stress, fatigue, and muscle soreness. Coaches use these to track players' subjective well-being and adjust training sessions accordingly. Another widely used method is the rate of perceived exertion (RPE), where players rate session intensity on a scale from 1 to 10 (or 6 to 20, based on the Borg scale) [6]. RPE provides a subjective measure of internal training load. A more objective but more costly approach is by using heart rate (HR) sensors [13].

However, this internal load data alone is insufficient for assessing players' fitness. Coaches also need information on the external load placed on athletes during training. A common method for measuring external load involves global positioning system (GPS) wearable sensors. These devices integrate accelerometers, gyroscopes, magnetometers, and GPS to track metrics such as total distance covered, number of accelerations and decelerations, sprints, and energy expenditure, etc. [41]. This data enables more precise session planning, better management of training loads, and facilitates the return-to-play (RTP) process [41].

During the preseason, teams focus on building physical capacity and conducting standardized tests to identify potential vulnerabilities that could lead to future problems. Tests such as repeated sprinting, the Yo-Yo test, vertical jump tests, and functional movement screenings (FMS) provide valuable information to conditioning coaches and medical staff about players' capabilities and limitations [86].

Another valuable data source frequently utilized in soccer is video tracking systems. Installed within stadiums, these systems can substitute GPS sensors and provide detailed data on player and ball positions throughout the match [17]. This allows for tailored performance analysis with additional contextual insights. However, a limitation is that these systems are fixed within the stadium and cannot be used elsewhere. Therefore, GPS sensors remain predominantly favored as they offer flexibility for use at training grounds, stadiums, and during away matches. Many teams with sufficient resources employ both systems concurrently to maximize data capture and analysis capabilities.

A widely used data source is event data extracted using TV broadcast cameras during soccer matches. This data includes details such as the position of events like tackles, passes, shots, and the players involved, as well as whether these actions were successful [42]. While utilizing this data requires additional resources from clubs, it provides valuable insights for evaluating team progress and scouting players to strengthen the squad. Notably, this type of data was not utilized in the dissertation, which focused instead on questionnaires and GPS data.

1.2 Metrics and Methods for Evaluating Physical Performance

During each preseason, conditioning coaches assess players' fitness using standardized tests. These include repeated sprinting, the Yo-Yo test, the 30-15 intermittent test, and similar evaluations [86, 15]. Additionally, functional movement screen (FMS) tests such as the deep squat, hurdle step, shoulder mobility, rotary stability, and trunk stability are conducted to identify potential vulnerabilities that could harm players in the future [54]. This dissertation focuses on the first type of test, evaluating players' aerobic and anaerobic capacities. Data from these standardized tests is used to gain insights into players' ability to recover after intense physical tasks.

The physical performance of soccer players can be also assessed through standardized training sessions using GPS and HR wearable sensors [81]. GPS serves as a calibration tool, ensuring that similar sessions yield comparable external load values over time. The fitness status of a soccer player is then evaluated by changes in internal load, such as HR response. These sessions act as soccer-specific exercise protocols for assessment. However, due to congested schedules, there is often limited opportunity to execute these protocols [60]. Consequently, physical performance is more frequently evaluated during matches using GPS data.

Coaching staff prioritize metrics such as total distance (TD), which indicates running volume, sprinting distance, and high-speed running (HSR) distance to assess high-intensity running [79]. Additionally, they consider high-intensity actions (HIA) not captured by speed alone, including the number of accelerations, decelerations, and metabolic power events (MPE). Other parameters, tailored to specific needs, include MPE average recovery power (W/kg), MPE average recovery time (s), average metabolic power (W/kg), energy expenditure (J/kg), anaerobic energy (J/kg), running distance (m), walking distance (m), running energy (J/kg), and walking energy (J/kg). A significant issue with using these metrics is the aggregation of data over the full duration of a game. This approach hides variations in a player's physical intensity throughout the match. For instance, a player might perform intensely in the first half but decline in the second, resulting in an above-average overall performance that misleads coaches about the player's fitness [58]. This dissertation addresses this problem by evaluating players' intensity on a minute-by-minute basis, offering soccer practitioners a clearer understanding of their players' fitness levels in the context of an actual match.

Lastly, cognitive load (CL) is a crucial factor that significantly impacts players' performance. While numerous studies have explored the effects of CL on decision-making and technical skills [3, 8, 21, 44, 75, 76], there is limited research on its impact on physical performance, especially in real-world settings outside controlled environments. Additionally, players respond differently to coaching instructions; some thrive on negative motivation, while most prefer positive motivation [51]. This dissertation aims to investigate the relationship between CL and physical performance using NASA Task Load Index (NASA-TLX) questionnaire data [40]. Furthermore, it explores whether ML algorithms can classify players into positive and negative motivation clusters based on their responses.

1.3 Approaches to Monitoring Injury and Physical Recovery

The ability of the coaching staff and medical team to keep key players fit and minimize injuries is crucial for a team's success. While injuries are inevitable, the RTP procedure must be carefully planned to avoid complications and reduce the risk of re-injury [27]. The medical staff typically establishes a recovery plan immediately after diagnosing the player's injury. Serious injuries requiring longer recovery times require more extensive adjustments to the plan [83, 20].

Most injuries are non-contact, caused by overuse rather than tackles from other players, with the majority being muscle-related [26, 25, 52]. Recovery from such injuries can range from a few days to several months. An essential aspect of planning the recovery process and facilitating the player's return to the team is the medical team's initial assessment of injury duration. This assessment can be further refined using ML algorithms with data collected by the medical staff which is another focus of this dissertation [82]. Similar computer-aided diagnosis (CAD) systems have been successfully employed in medicine to enhance radiologists' decision-making [19, 70].

When a player is returning from an injury or engaging in regular training, it is beneficial to know his fitness capacity [16]. Understanding the rate at which player fatigue and recovers is crucial for conditioning coaches to identify areas for improvement [69]. This knowledge becomes even more important when a player has been sidelined due to injury and is gradually increasing workload to regain fitness. Pre-injury values serve as a control point to determine when a player is fully healed. While a player might be physically healed, the entire process concludes only when he is physically ready to participate at full capacity, which may take longer if the injury is severe. This dissertation focuses on building individual player fatigue rates and recovery profiles, providing a data-driven method for assessing fitness.

1.4 Dissertation Structure

This dissertation integrates multiple data sources to explore the physical performance of soccer players. It is organized into two primary sections: Physical Performance Assessment Methods, consisting of three distinct areas, and Return-to-Play Data-Aided Systems, comprising two focused topics. The flowchart in Figure 2.3 illustrates the connections between the different data sources and the specific goals of each section of the dissertation.

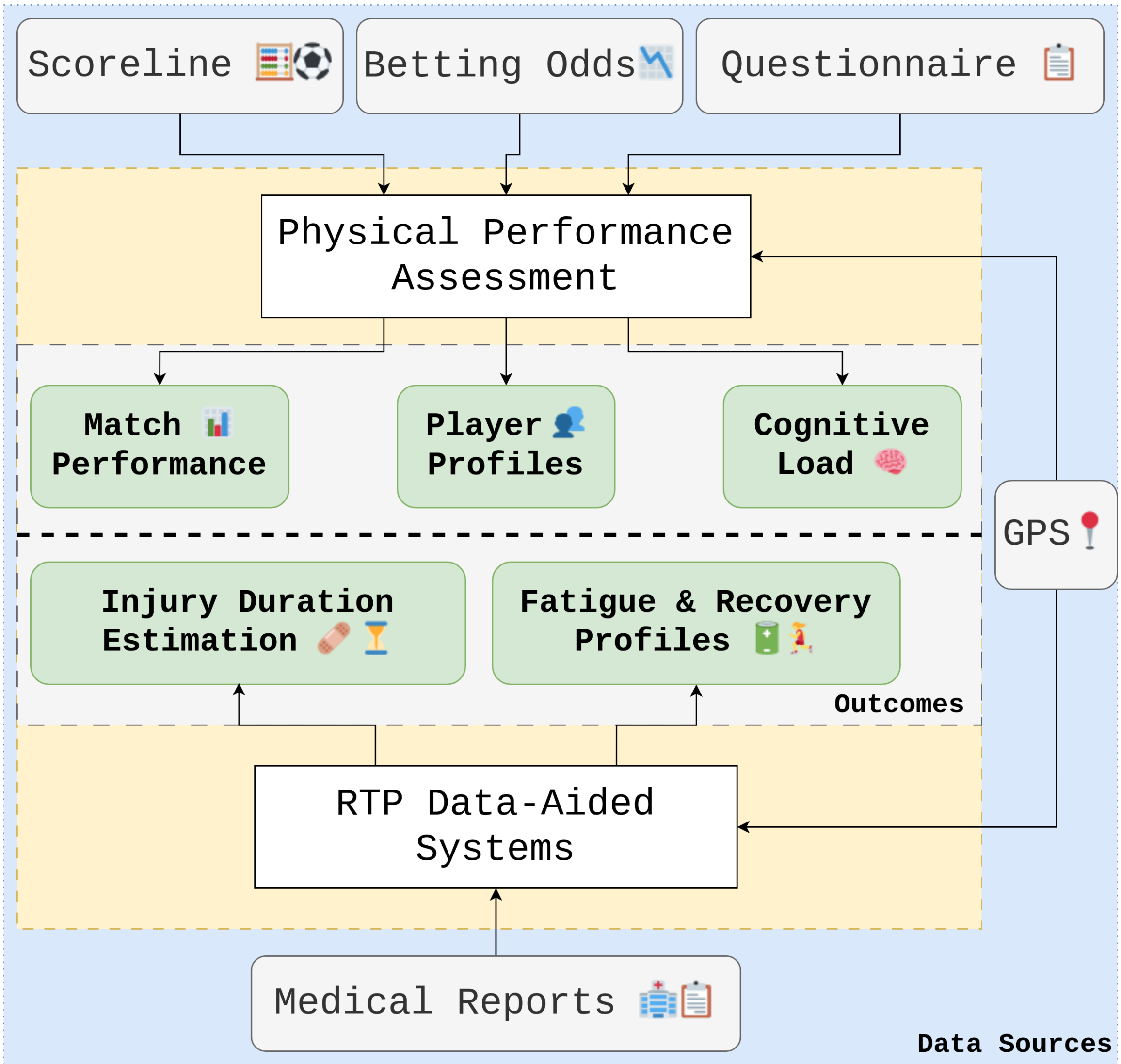

Figure 1.1: Flowchart illustrating the various data sources utilized in the dissertation and the corresponding outcomes derived from different components of the study. Blue areas indicate the data sources, yellow areas highlight the two primary topics of the dissertation, and green rectangles encapsulate all dissertation outcomes.

Chapter 2

Physical Performance Assessment

2.1 Minute-by-Minute Intensity Evaluation in Soccer Matches

As described in Section 1.2, GPS sensors provide a variety of metrics that capture different physical characteristics. Practitioners often analyze these metrics at an aggregate level across the entire session or match; for the purposes of this dissertation, this method is referred to as GM-GAME. This approach raises questions about how performance is distributed over time, as many factors influence how players behave and invest their energy during a match. One dominant factor is the quality of the opposition; higher-quality opponents typically require greater physical exertion from players [5, 65]. Analyzing differences between 1st and 2nd half of the match revealed a significant reduction in TD covered by the players against lower-level and medium-level teams [32]. Another significant issue in modern soccer is congested periods, which have been shown to reduce the number of HIA [61]. The workload of soccer players also depends on the playing position [59], and the game outcome (win, draw, lose) [50, 58, 64].

While many studies have examined various contextual factors and their effects on physical metrics derived from wearable sensors, none have analyzed changes on a more granular, minute-by-minute level. Additionally, the use of different GPS providers, each with unique thresholds for speed and acceleration, sampling rates, and proprietary metrics, complicates the transfer of knowledge between providers.

One goal of this dissertation was to develop a framework for evaluating player intensity throughout a match on a more granular level. However, providers often limit the granularity of data extraction. The dataset used included 38 male soccer players across two half-seasons, including 80 games. The playing positions included 11 center backs, 7 wing backs, 8 midfielders, 3 wide forwards, and 9 forwards. The GPS device used was the GPexe pro[2], operating at a sampling rate of 18Hz. Ideally, practitioners would like metrics split into 1-minute segments to evaluate differences across 90 minutes. However, GPS devices have limitations, and processing time for shorter intervals increases exponentially as the duration of the interval decreases. This prevents data extraction in a reasonable amount of time, so the minimum available time frame was set to 5-minute intervals.

The GPexe device allows for the additional extraction of MPE, which represents short-duration, HIA performed by a player during a session or match. This data is highly valuable as it provides information about a player's peak physical exertion. If a player can quickly recover after such events that makes him more effective, increasing the likelihood of capitalizing on opponents' mistakes.

However, MPE data should not be analyzed in isolation. The context of a player's recovery, whether through walking or moderate-intensity running, significantly impacts the interpretation of the data. Therefore, MPE data must be combined with general expenditure data for a comprehensive analysis. An example illustrating the integration of one GPS metric - Energy (J/kg) - with MPE energy and general energy over 5-minute periods (GM-5MIN) is shown in Figure 2.1. It is evident that MPE energy accounts for only about one-third of overall energy expenditure. To fully understand the intensity of a particular minute, a new dataset was created that combines both MPE and GM-5MIN datasets.

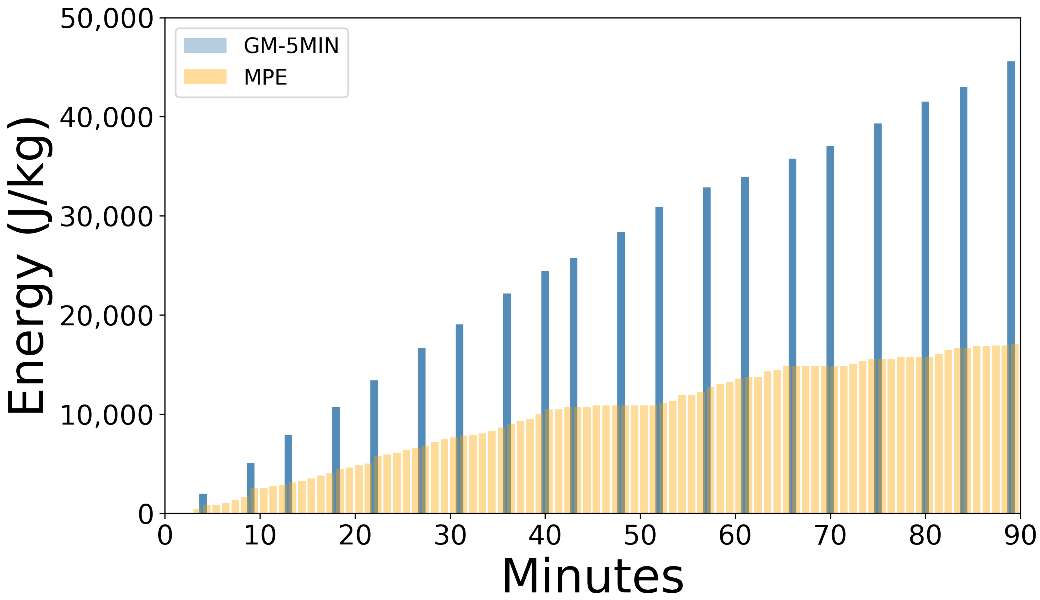

Figure 2.1: Comparison of GM-5MIN and MPE energy expenditure over 90 minutes, with blue bars representing GM-5MIN values throughout the match and orange bars indicating the corresponding MPE energy expenditure. Notably, MPE contribute to approximately one-third of a soccer players' total energy consumption.

For each player and game, the process involved merging three main data sources: 1) GM-5MIN attributes in the 5 min period prior to the observed minute (GM-5MIN-PRIOR), 2) MPE attributes in the preceding 3 and 5 minute periods (MPE-PRIOR), and 3) MPE attributes in the observed minute (MPE-CURRENT). This merging aimed to capture the overall player exertion just before the observed minute, the peak intensity preceding the observed minute, and the specifics of the observed minute itself.

The resulting dataset included 25 parameters, which is too much data for a practitioner to comprehend effectively. To address this, an ML model needed to be built to cluster the dataset efficiently. However, due to the limited number of games and players, and to prevent overfitting, the features in the dataset had to be reduced without losing significant information. Principal Component Analysis (PCA) was chosen for feature reduction, as it effectively reduces the number of dimensions in a dataset while preserving most of the original information.

Before applying PCA, the dataset was normalized using min-max normalization to prepare it appropriately. PCA managed to preserve 92.8% of the data variance and reduced the number of features from 25 to just 7 components. Following PCA, an unsupervised ML algorithm was applied to cluster the players' minutes and identify similarities. Several algorithms were considered, including K-means, Hierarchical Clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), t-stochastic neighbor embedding (SNE), Autoencoders, and Gaussian Mixture Models. K-means was chosen for its simplicity, ease of implementation, and speed. One drawback of the K-means algorithm is the requirement to specify the number of clusters. The optimal number of clusters was determined using the elbow method, which indicated that three clusters were ideal. Further analysis revealed distinct differences between these clusters, which can be described as low, medium, and high-intensity minutes. A detailed description of these differences is provided in Table 2.1.

The proposed clustering approach was evaluated by comparing its results to traditional GM-GAME data used for athlete load monitoring, which lacks a standard method for measuring intensity. The game intensity was categorized into high, middle, and low groups. These were equated to specific GM-GAME metrics: distance covered above 20 km/h for high intensity, running distance below 20 km/h (excluding walking) for middle intensity, and walking distance for low intensity. These metrics were selected due to their frequent use in intensity evaluation. The traditional GM-GAME approach was compared to the methodology presented in this dissertation by calculating the coefficient of variation (CoV) for each intensity cluster (high, middle, low). The clustering algorithm exhibited superior between-game variability, yielding CV scores of 45.91%, 30.66%, and 24.41% for high, middle, and low-intensity clusters respectively, compared to GM-GAME scores of 21.32%, 16.82%, and 19.12%. The analysis included all 80 available games, revealing a more natural distribution of intensity levels and highlighting limitations in GM-GAME data's ability to accurately capture game intensity.

Table 2.1: Clustering group results. The darker background color represents higher intensity.

Parameter Name	Low Group		Middle Group		High Group	
MPE features (1 min)						
	μ	σ	μ	σ	μ	σ
Energy (J)	0	0	180	150	230	280
Event count	0	0	1.8	0.9	2	1
Average recovery time (s)	60	0	20	7	18	8
MPE features (3 min before)						
	μ	σ	μ	σ	μ	σ
Energy (J/kg) 3 min	400	350	400	330	700	350
MPE count (3 min)	3.8	2.2	4.0	2.2	6	2
Recovery time (s) (3 min)	155	16	154	15	140	16
GM-5MIN features (5 min before)						
	μ	σ	μ	σ	μ	σ
Energy (J/kg)	2,900	1,000	1,300	1,000	2,800	500
MPE count	6	3.5	4	3.5	9	2.5
Anaerobic energy (J/kg)	750	400	500	400	1,000	150
Avg. MPE recovery time (s)	60	80	50	60	23	8
Running energy (J/kg)	1,400	800	1,000	800	2,250	450

2.2 Developing Player Profiles to Inform Starting vs. Substitute Decisions

Soccer player performance is influenced not only by technical and tactical abilities and the quality of the opponent but also by the context preceding each game. One crucial contextual factor is pre-game expectations, such as which team is favored to win or if the match is expected to be closely contested. These expectations influence players' motivation and exertion as the scoreline evolves during the game. Consequently, various studies have examined how the scoreline impacts player performance.

Research indicates that teams leading a game tend to be in a comfortable position and often conserve energy rather than exert extra effort [45, 73]. Conversely, some studies have found that TD covered is greater when a team is winning [18, 56, 63], and that teams with lower possession need to cover more distance [49]. Additionally, other studies suggest that the distance covered across various speed thresholds is higher when the game is tied [9, 68]. Many studies concluded that the scoreline significantly affects player performance [4, 9, 18, 45, 46, 56, 63, 68, 73]. However, Bloomfield et al. [11] found no such connection, arguing that any intensity increases following scoreline changes are short-lived.

The primary limitations of current studies include the lack of an individualized approach to account for player differences, the absence of publicly available datasets for replication and assessment, and insufficient methods for automating insights for coaching staff. This dissertation aims to address these limitations by developing a framework that provides personalized performance insights along with publicly available data and source code.

2.2.1 Data Collection and Preprocessing

The dataset was collected using three data sources: 1) betting coefficients before each match, 2) post-match data on the timing of scored and conceded goals, and 3) GPS wearable sensors, specifically the GPexe pro[2] (Exelio srl, Udine, Italy), to quantify player exertion. The dataset comprised 33 games played by 19 male soccer players, totaling $3,135$ minutes of game time. This served as the foundation for further individual player analysis.

The betting coefficients provided a way of assessing pre-match expectations, specifically determining whether a team was the favorite to win a match or not. These odds represented betting sites' estimations of the probabilities of a win, loss, or draw for the observed team against its opponent. The final expectation, denoted as e, was calculated by dividing the probability of winning by the probability of losing. If e was greater than or equal to 2, indicating that the chances of winning were at least double those of losing, the team was classified as a favorite. Conversely, if e was less than 0.5, indicating that the chances of losing were at least double those of winning, the team was classified as an underdog. Values of e between 0.5 and 2 were interpreted as indicative of closely contested matches. Given that only one team was observed, and it was considered an underdog in only two matches, the analysis was simplified into categories of favorite and non-favorite, with the latter defined as cases where e was less than 2.

Post-match goal timing data was used to categorize playing minutes into various goal difference (GD) scenarios. These scenarios were determined by calculating the difference between the score of the team under analysis and that of their opponents. The boundary values (-2 and 2) included all situations where the team was either trailing by 2 or more goals or leading by 2 or more goals, respectively. The distribution of minutes across each GD category is presented in Figure 2.2. The GD0 category is the most prominent, as every match begins with a 0-0 score. GD0, representing a draw, accounted for only 272 minutes where the GD differed from the initial score, meaning that only 19.4% of GD0 minutes reflected results other than 0-0. The GD distribution is negatively skewed, likely because the observed team was among the best in the league and a favorite in the majority of matches (21 out of 30).

2.2.2 Mathematical Modeling for Player Energy Expenditure

A mathematical model was constructed to account for each game j, defining expectation e as either favorite (f) or non-favorite (nf), denoted as $e_j \in f, nf$. GD was defined as a time-dependent variable t, with GD values spanning $d_j(t) \in -2, -1, 0, 1, 2$, used as a predetermined input to

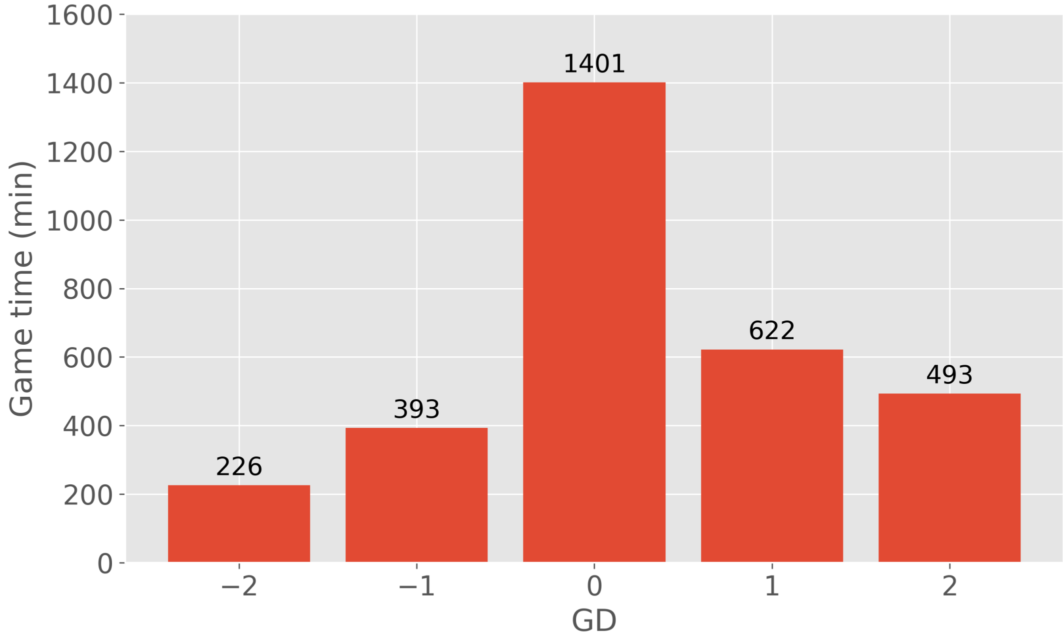

Figure 2.2: Distribution of game time across the observed GD states.

the model. For every player i, score performance parameters were defined as $P_{d,i}$. These parameters remain constant across a player's analyzed matches and need to be estimated from the available data. The final model is denoted as $P_i(d_j(t), e_j)$, incorporating both the expectation and the GD. Each player's energy expenditure $E_{i,j}(t)$, is governed by a differential equation, with the initial condition $E_{i,j}(0) = 0$:

$$\frac{dE_{i,j}}{d\tau} = \sum_{\tau=0}^{\tau=t_{e,i,j}-t_{s,i,j}} P_i(d_j(\tau + t_{s,i,j}), e_j) \cdot \tau \cdot \eta_i(\tau). \tag{2.1}$$

In this model, $\eta_i(\tau)$ is represented by a decaying exponential function $\eta_i(\tau) = e^{-\alpha_i \cdot \tau}$ with α_i defined as $-\ln(\eta_{90})/90$. Here, η serves as a player's endurance coefficient, and τ denotes the time the player has spent on the pitch. The playing time for player i in game j is computed by subtracting the time the player exited the game ($t_{e,i,j}$) from the time the player entered the field ($t_{s,i,j}$). The model is calculated using the Euler method with $\Delta t = 1$ minute, for the game time interval $t \in [0, t_{end,j}]$, where $t_{end,j}$ is the total duration of game j.

The power zone (P_i) limits were set between 200 and 800, which corresponds to the energy a player expends over a 5-minute duration. Endurance, denoted as η, is computed using a standard model of energy depletion with a fixed parameter, α. To account for typical intensity reduction over a 90-minute game, η_{90} is restricted to the range 0.5 to 1.0.

Equation (2.2) consolidates the parameters used in the optimization process, while Equation (2.3) defines the quality of the solution, where n represents the number of recorded games. The objective is to minimize the squared difference between measured and estimated values, normalized by the total number of minutes played. This is formalized as $\hat{\mathbf{x}} = \arg\min_x \varepsilon(\mathbf{x})$.

$$\mathbf{x} = \begin{pmatrix} P_{-2,f}, P_{-1,f}, P_{0,f}, P_{1,f}, P_{2,f}, \\ P_{-2,n}, P_{-1,n}, P_{0,n}, P_{1,n}, P_{2,n}, \\ \eta_{90} \end{pmatrix} , \tag{2.2}$$

$$\varepsilon_i(\mathbf{x}) = \frac{\sum_{j=1}^{j=n} \int_{t_{s,j}}^{t_{e,j}} (E_{i,j}(\mathbf{x}, t) - E_{measured,i,j}(t))^2 dt}{\sum_{j=1}^{j=n} (t_{e,i,j} - t_{s,i,j})}, \tag{2.3}$$

The process was conducted for each player across all matches in which they participated, with the procedure repeated ten times per player. This yielded power and endurance coefficients, enabling a comparison between the GPS sensor data and the algorithm's estimates for each match. An aggregated squared error was calculated for each player, and two baseline scenarios were established for comparison. The first baseline, B_1, used mean power values per GD without accounting for fatigue (i.e., η set to 1). The second baseline, B_2, used median power values per GD, also disregarding fatigue effects. The model's estimates were evaluated against B_1 and B_2 using standard regression metrics, including mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE).

This dissertation presents an objective tool for evaluating players' physical states, accounting for contextual parameters and individual endurance. The proposed method was validated against commonly used baselines that calculate mean and median energy expenditure per GD. This approach outperformed both baselines across all evaluated metrics, demonstrating its superior effectiveness and accuracy in assessing player performance and energy expenditure.

2.3 Impact of Cognitive Load on Physical Performance

Coaches play a crucial role in athletes' lives, heavily influencing their emotional and behavioral responses [23]. One of the reasons for this is that a coach's feedback directly communicates information about an athlete's competence, which in turn has a significant impact on their motivation [29]. Coaches often use either positive or negative motivational techniques to drive athletic improvement [51]. Positive feedback is generally preferred, as it tends to boost self-confidence and belief, whereas negative feedback can diminish these qualities [7] and may lead to adverse effects such as reduced self-esteem and performance [10, 43, 62]. Although motivation is inherently individual, certain patterns are observed across athletes' motivational profiles [22]. Selecting the appropriate motivational approach, particularly in high-pressure situations, can be critical in determining the outcome of a competition.

Physical fatigue, characterized by diminished muscle capacity, is directly associated with a decline in performance [33, 38]. Cognitive fatigue, on the other hand, arises from sustained engagement in mentally demanding tasks and involves both psychological and physiological components [53]. This form of fatigue intensifies feelings of tiredness and leads to reductions in performance, endurance, and motivation [12, 24, 37]. CL which athletes are experiencing often cause them to reduce their performance intensity [14], which can further impair endurance [53]. Additionally, cognitive fatigue adversely affects physical performance, particularly in motor skills, resistance training, and aerobic capacity, though anaerobic capacity appears less impacted [75]. Many studies have demonstrated the negative impact of cognitive fatigue on soccer players' technical performance [3, 76], and endurance [53, 75]. However, these effects have primarily been examined in highly controlled settings [8, 44, 76].

This dissertation seeks to address this limitation by examining the impact of CL during the final period of the soccer season. Current research typically assesses cognitive load through questionnaires, including tools such as:

- **NASA-TLX**: Measures perceived workload across mental demand, physical demand, temporal demand, performance, effort, and frustration on a 100-point scale.
- **dRPE**: Differentiates perceived exertion into overall exertion, cardiorespiratory effort, and muscular effort.
- **Competitive State Anxiety Inventory 2 (CSAI-2)**: Assesses competitive anxiety in athletes through cognitive anxiety, somatic anxiety, and self-confidence on a 4-point scale.
- **100 mm Visual Analog Scale (VAS)**: Measures subjective experiences by marking a position on a 100 mm line, providing a quantitative measure of attitudes or feelings.
- **Profile of Mood States (POMS)**: Evaluates mood states using 65 adjectives rated on a 5-point scale, grouped into tension-anxiety, depression-dejection, anger-hostility, vigor-activity, fatigue-inertia, and confusion-bewilderment.

Among the methods considered, practitioners determined that the NASA-TLX would be the most practical for daily implementation, leading to its selection for this research. After each session, players completed the NASA-TLX questionnaire, and the responses were integrated with GPS wearable sensor data. The NASA-TLX questionnaire, focusing on three key components–mental demand, temporal demand, and physical demand–was used to evaluate CL. The mental demand component assesses the cognitive effort required, while temporal demand addresses the perceived time pressure. Physical demand reflects the perceived physical effort, all of which contribute to an understanding of CL [57, 31].

The two combined data sources served as a base for building player cognitive clusters as shown in Figure 2.3. A new metric, the CL score, was introduced to evaluate cognitive load for each session. This score is derived from the average of three components of the NASA-TLX index: mental demand, temporal demand, and physical demand. The Z-score for each player is calculated to normalize individual data variations. The measurements are then divided into three clusters based on the CL Z-score. The threshold for distinguishing the clusters, denoted as θ, is determined through a grid search within the range $\theta \in [0.1, 1]$ with a step size of 0.01.

This range is chosen to ensure sufficient data samples in each of the three clusters. The clusters are categorized as low CL ($Z - score \leq -\theta$), moderate CL ($-\theta < Z - score < \theta$), and high CL ($Z - score \geq \theta$).

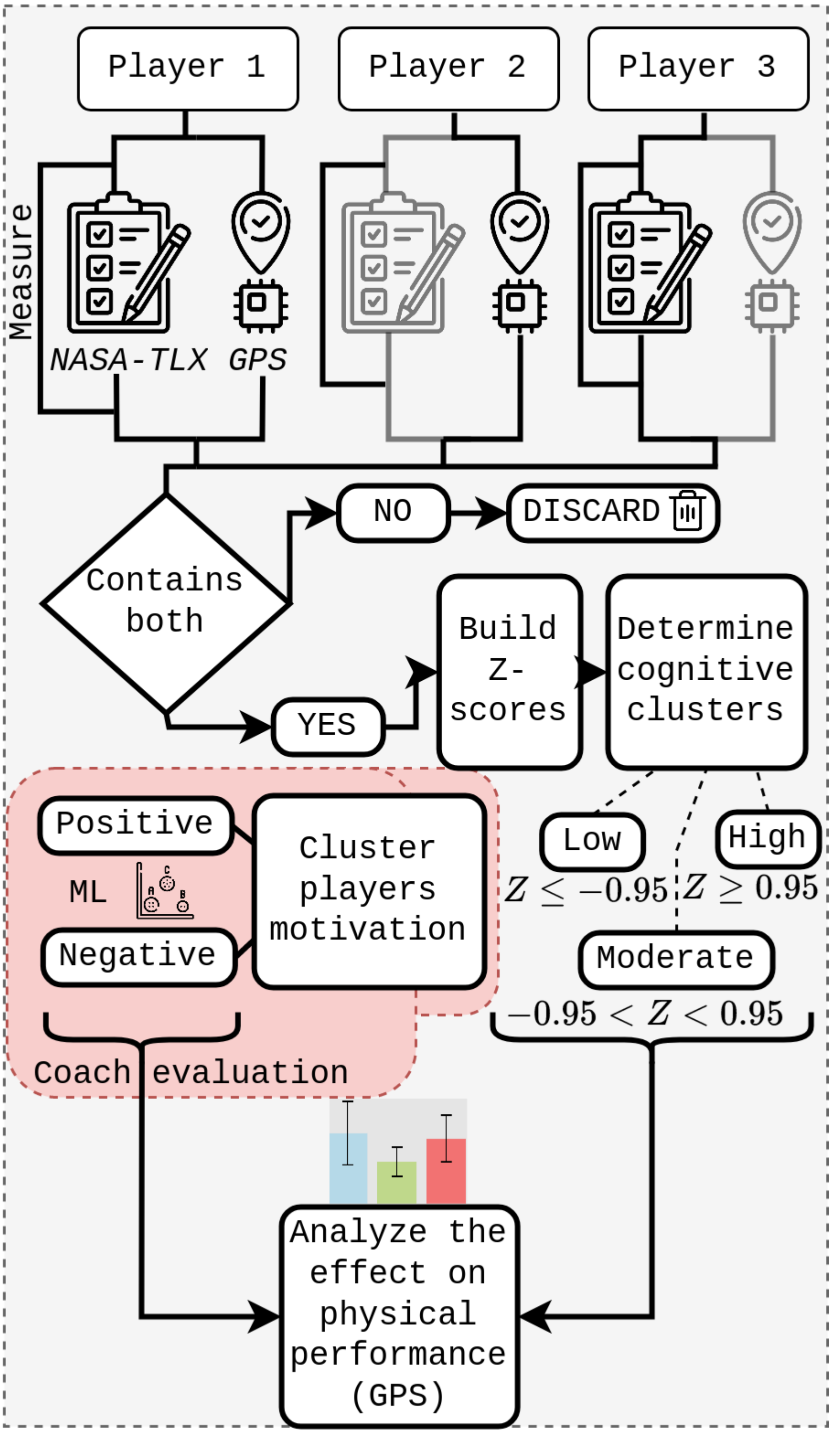

Figure 2.3: The flowchart illustrating the data collection and analysis workflow. The process involved the use of NASA-TLX questionnaires and GPS sensors. The research focused on two key objectives: examining the impact of cognitive load on physical performance, and exploring the effect of motivation profiles on the external load metric.

The quality of θ is evaluated using the function $f(\theta)$, defined as: $f(\theta) = max(|\mu_{low} - \mu_{high}|)$, where μ_{low} and μ_{high} are the mean CL scores for the low and high clusters, respectively. The resulting clusters are validated using a one-way Analysis of Variance (ANOVA), followed by the post-hoc Tukey honestly significant difference (HSD) test. To examine the impact of CL clusters on external load parameters Multivariate Analysis of Variance (MANOVA) test is conducted. If MANOVA indicates significant differences between the clusters, further post-hoc analysis with Tukey HSD test is conducted.

Another objective of this dissertation was to evaluate whether players can be clustered into positive or negative motivation profiles using a ML algorithm. For this purpose, data from the NASA-TLX questionnaire, including all six components, was utilized as input for the K-means clustering algorithm, a widely used unsupervised ML technique [2]. Using all of the components was essential to leverage the complete dataset, given that motivation is highly individual and context-dependent. The validity of the clustering results was assessed by comparing them with the observations of the coach who has been training the players. Further analysis was performed to examine whether motivation profiles vary significantly with the level of external workload and to assess how these variations in motivation might influence performance.

Chapter 3

Return-to-Play Data-Aided Systems

3.1 Estimation of Muscle Injury Duration Using ML

Injuries in soccer can significantly impact a team's final position in the league standings [28]. Coaches must adapt their tactics to opponents, and having a full roster available makes strategic adjustments easier. However, injuries are inevitable, and the responsibility for proper healing rests with the medical team. Developing an effective RTP plan is crucial; if a player returns before fully recovering, the risk of re-injury increases, potentially sidelining them for an extended period [27]. Various studies and guidelines provide recommendations for treating injured athletes based on injury type [55, 77, 87], all of which rely on expert opinions and athlete feedback [55]. Estimating injury duration involves considering the following factors:

- **Clinical Assessment and Imaging**: Initial evaluation through history, physical examination, and imaging studies (X-rays, magnetic resonance imaging (MRI), ultrasound) to diagnose the injury and assess its severity.
- **Diagnosis and Severity Classification**: Identification of the specific injury and its severity (e.g., Grade I-III for strains/sprains), providing an initial estimate of healing time based on medical knowledge and guidelines.
- **Individual Factors and Rehabilitation**: Consideration of the athlete's overall health, fitness level, and response to treatment, with regular monitoring and reassessment to adjust recovery timelines as needed.

This dissertation aims to enhance the process of determining when an athlete is ready to participate again in a soccer match. The main motivation lies in CAD systems, which have already proven their value in general medicine [84]. With the rise of data and ML in sports, there is an opportunity to improve the current RTP process [71]. CAD systems utilize deep learning techniques to assist radiologists in detecting diseases and abnormal patterns. However, medical records are often in a simpler tabular format, allowing the use of less data-intensive and simpler ML algorithms [36].

The most common injuries in soccer are non-contact injuries, which occur due to overuse or inadequate fitness. Among these, muscle injuries are predominant, including lesions, ruptures, partial ruptures, and strains, each with varying recovery durations [27, 39]. Valle et al. [82] attempted to predict recovery duration following hamstring injuries using ML algorithms on data extracted from MRI. Their system achieved an R^2 score of 0.48 and a MAE of 9.8 days. The accuracy of such models can be further improved by integrating injury attributes with physician predictions, similar to CAD systems[85].

3.1.1 Dataset Overview and Injury Classification

A dataset was collected to build an ML algorithm for injury duration estimation. Data was gathered from a professional soccer club over a two-year period. The club's medical team used a custom online platform to document each injury and track the recovery process. Injury reports were filled out by the club's physiotherapist after on-site examinations and, if necessary, by reaching a consensus with other medical team members. While records were updated throughout the recovery process, only the initial parameters and muscle injuries were considered for this dissertation.

The online platform for injury reporting involved three steps: 1) entering general information, 2) specifying injury details, and 3) providing recovery information. For a comprehensive list of injury parameters and their explanations, refer to Table 3.1. It is important to note that not all parameters listed were utilized as input for the ML algorithm. Parameters P4 (place of injury) and P5 (body side) were excluded, along with the recovery feature P17, due to their irrelevance to injury classification. Additionally, the features P1 (date of examination) and P19 (medical notes) were withheld by the club to protect player identities. The injury duration, which this dissertation aims to predict, was calculated by subtracting the date of the last examination from the date when the injury first occurred. Parameters P12 and P13 were also excluded, as they are not relevant to muscle injuries.

The classification system used for tracking muscle injuries was the British Athletic Muscle Injury Classification (BAMIC) system, which categorizes muscle injuries to standardize diagnosis and management:

- **Grade 0**: No visible muscle damage. Includes focal neuromuscular injury or generalized muscle soreness (0A and 0B).
- **Grade 1**: Small injuries with minimal muscle fiber disruption. Includes injuries confined to the muscle belly or extending into the muscle-tendon junction (1A and 1B).
- **Grade 2**: Moderate muscle fiber disruption. Includes injuries within the muscle belly or extending into the muscle-tendon junction (2A and 2B).
- **Grade 3**: Extensive muscle fiber disruption. Includes injuries within the muscle belly or extending into the muscle-tendon junction (3A and 3B).

The BAMIC system was used to standardize values provided by the online platform. Specifically, all contusions (C) were mapped to a value of 0; values 1A and 1B were mapped to 1; values 2A and 2B were mapped to 2; value 3A was mapped to 3; and value 3B was mapped to 4. This mapping corresponds directly to the injury severity grade.

Table 3.1: Parameters recorded through the online platform for injury reporting. Values separated by a slash “/” represent single-choice options, while those separated by a vertical bar “/” denote multiple-choice options.

<table>
<tr><th></th><th>Injury Parameter Description</th><th>Values</th></tr>
<tr><td></td><td>P1: Date of a clinical examination</td><td>Date</td></tr>
<tr><td></td><td>P2: Is the injury the result of a tackle?</td><td>Yes/No</td></tr>
<tr><td></td><td>P3: Has the player stopped playing?</td><td>Yes/No</td></tr>
<tr><td></td><td>P4: Where has it occurred?</td><td>Training/game/national team/other</td></tr>
<tr><td></td><td>P5: On which side of the body is it located?</td><td>Left/right/middle</td></tr>
<tr><td rowspan="11">Injury-specific</td><td>P6: Injury classification according to the BAMIC.</td><td>Numbers 0–4, suffix A/B/C</td></tr>
<tr><td>P7: What is the position according to muscle?</td><td>Proximal|distal|abdominal</td></tr>
<tr><td>P8: What is the depth?</td><td>Middle muscle|deep|superficial fascia</td></tr>
<tr><td>P9: Which body part is affected?</td><td>Hamstring|quadriceps|
adductors|abductors|calf</td></tr>
<tr><td>P10: What is the swelling level?</td><td>None/low/moderate/high</td></tr>
<tr><td>P11: What is the tone level?</td><td>None/low/moderate/high</td></tr>
<tr><td>P12: What is the crepitation level?</td><td>None/low/moderate/high</td></tr>
<tr><td>P13: What is the elasticity level?</td><td>None/low/moderate/high</td></tr>
<tr><td>P14: Is palpation painful?</td><td>Yes/no</td></tr>
<tr><td>P15: Is contraction painful?</td><td>Yes/no</td></tr>
<tr><td>P16: Is stretching painful?</td><td>Yes/no</td></tr>
<tr><td rowspan="3">Recovery</td><td>P17: What is the current phase of recovery?</td><td>Numbers 1 to 6</td></tr>
<tr><td>P18: Expected duration in days, weeks, and months.</td><td>Number</td></tr>
<tr><td>P19: Additional comments from the medical staff.</td><td>Text</td></tr>
</table>

Some parameters on the online platform offered multiple and single selection options. The parameter for injury according to the muscle (P7) sometimes included injuries affecting multiple muscle groups; thus, the available choices–proximal, abdominal, and distal muscles–were one-hot encoded. The depth feature (P8) was merged into a single feature by mapping the categories of no information, superficial fascia, middle muscle, and deep fascia to values ranging from 0 to 3, respectively. The parameter for body position (P9) was similarly one-hot encoded and grouped into the following categories: hamstring, quadriceps, adductors and abductors, calf, and abdominal wall. Clinical examination features, including swelling, tone, crepitation, and elasticity (P10–P13), were assessed using a scale from 0 to 3, reflecting their severity. Pain associated with palpation, contraction, and stretching (P14–P16) was classified as either painful or not painful. Finally, the expected recovery duration (P18) was taken into account and converted into days to facilitate comparison.

The final dataset consisted of 84 muscle injuries, with the distribution of muscle injury types according to the BAMIC system shown in Figure 3.1. Some injuries, however, exhibited significantly longer recovery times, which were considered outliers. These outliers posed a challenge for ML algorithms, particularly given the limited size of the dataset. To mitigate this issue, a cut-off threshold of 35 days (5 weeks) was applied, reducing the dataset to 80 muscle injuries. The excluded injuries had recovery duration of 39, 45, 52, and 67 days. The first two cases involved repetitive calf injuries and an abdominal wall rupture, while the latter two, with recovery periods of 52 and 67 days, were related to adductor and hamstring injuries in goalkeepers. Expanding the dataset in the future may allow for an extension of the cut-off threshold to 7 weeks, thereby including more injuries with extended recovery periods.

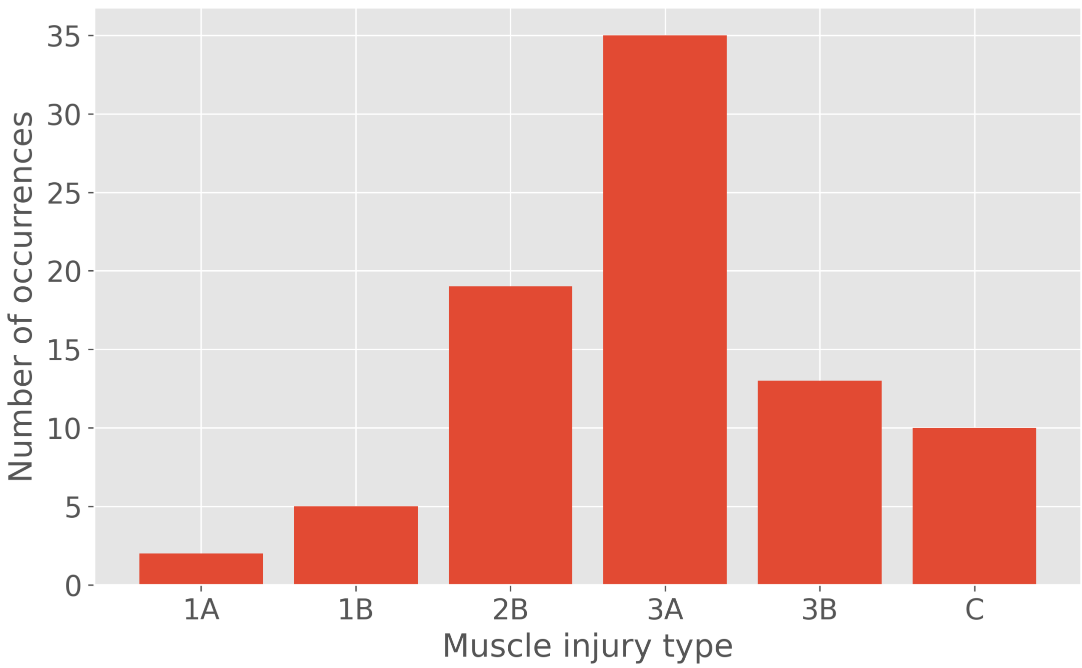

Figure 3.1: Muscle injuries distribution by type. Classification was performed using the BAMIC system. The x-axis denotes different injury types, as well as contusions denoted by C. The y-axis displays the occurrence count for each type of injury.

3.1.2 Assessment of Model Variants and Expert-Integrated Approaches

The previously described dataset enabled the development of ML models for estimating injury duration. Three distinct models were evaluated:

- **Linear Regression (LR)** is a simple model that assumes a linear relationship between input features and the target variable. It's efficient and effective for tasks where this linear assumption holds, making it a solid baseline model. However, its performance can suffer when dealing with non-linear data or multicollinearity, and it often requires feature scaling for optimal results.

- **Decision Trees (DT)** are non-parametric models that split the data based on feature values to predict outcomes, making them useful for handling complex, non-linear relationships. They excel in scenarios where model interpretability is crucial, as the decision paths are easily understandable. However, they can be prone to overfitting, especially with deep trees, and may struggle with datasets containing a high number of features without proper regularization.

- **Extreme Gradient Boosting (XGB)** is an advanced ensemble learning method that builds multiple decision trees sequentially to improve accuracy, making it highly effective for complex regression tasks. It often outperforms other models due to its ability to handle non-linear relationships and various data types, along with built-in regularization to prevent overfitting. However, XGB is less interpretable than simpler models like DT or LR, making it harder to understand the contribution of individual features to predictions.

Hyperparameters for each algorithm were fine-tuned to achieve optimal performance. This was accomplished using a five-fold Bayesian search cross-validation (CV), with MSE as the evaluation metric. The best-performing hyperparameters identified through this process were then used for training and evaluating the models. Due to the limited size of the dataset, model performance was assessed using the leave-one-out (LOO) method. Each experiment was repeated 10 times per model to account for potential variations caused by Bayesian search and differences in data within CV folds. This repetition ensured the stability of the models and the reliability of the results. For each iteration, the metrics of R^2, mean absolute percentage error (MAPE), MAE, MSE, and RMSE were calculated for both the Bayesian search CV and LOO evaluations. The model with the lowest average MSE in the five-fold Bayesian search CV was selected as the final model.

To evaluate whether incorporating an expert's estimation feature could enhance model performance, the process was executed in two phases. In the first phase, the feature containing the expert's estimated recovery time (in days) was excluded. In the second phase, this feature was included in the model. By comparing the performance between these two phases, the impact of integrating expert knowledge was assessed. Out of the initial 80 muscle injuries, expert input was available for 69, leaving 11 injuries without an expert-estimated recovery time, thereby reducing the usable dataset. To ensure a fair comparison between the algorithms and the expert input, the LOO performance evaluation was conducted exclusively on the subset of 69 injuries. However, hyperparameter selection was performed on the entire dataset to maximize the optimization of each algorithm's performance.

3.2 Modelling Muscle Fatigue and Recovery with Optimization Methods

Muscle fatigue models (MFMs) are valuable tools for assessing and enhancing understanding of athletes' performance and capabilities [67]. These models are categorized into theoretical and empirical types.

Theoretical MFMs offer the advantage of being grounded in physiological principles, which allows them to provide deeper mechanistic insights into the causes and processes of muscle fatigue [88]. This can make them broadly applicable across different scenarios and populations. However, these models can be complex, requiring a detailed understanding of physiology, and they may be difficult to apply in practical situations without extensive data or assumptions. In contrast, empirical MFMs are simpler and more directly applicable since they are built from experimental data and observations [47]. This makes them easier to use and often more accurate in specific, well-defined contexts. However, their reliance on observed data means they lack generalizability outside the conditions under which they were developed, and they do not provide insight into the underlying physiological mechanisms of fatigue.

The development of MFMs has progressed significantly over the years, starting with foundational work by Liu et al. who introduced a model using fatigue (F) and recovery (R) parameters to explain muscle behavior during voluntary effort [47]. This model was validated through hand-grip exercises, showing it could predict muscle force production and suggesting that under maximum effort, only 97% of true maximum force could be achieved. Xia et al. later refined this model into a three-compartment framework, which included resting, active, and fatigued states, making it more computationally efficient for predicting fatigue and recovery across various activation patterns [88]. However, it faced limitations, particularly with dynamic strength surfaces and central fatigue representation, especially in prolonged or high-intensity tasks. Other academics tried to adjust this model to specific tasks, though these changes proved effective only for the maximal effort tasks [34, 48, 78]. The Sih, Ng, and Stuhmiller [74] (SNS) model introduced a four-compartment approach, governed by differential equations, and offered a broader understanding of muscle activation [74]. Despite its high accuracy, it lacked constraints on motor unit activation speed, leading to potential inaccuracies. More recent research by Frey-Law examined the impact of different maximal voluntary contraction distributions on muscle fatigue predictions, highlighting the importance of proper data sampling to maintain model accuracy [30].

The goal of this dissertation is to refine the existing SNS approach to make it applicable to the dynamic demands of soccer. In the original model, the maximal motor unit capacity, M_0, is expressed as $M_0 = M_A + M_F + M_{UC}$, where M_A represents the number of active motor units, M_F denotes the number of fatigued units, and M_{UC} accounts for inactive yet unfatigued motor units. For soccer applications, this concept can be adapted to define M_P, representing a player's currently available motor units, as described by Equation 3.1.

$$M_P = M_0 - M_A - M_F \tag{3.1}$$

A significant limitation in the SNS model arises from the treatment of the rate of change of active motor units, denoted as $\frac{dM_A}{dt}$. The model assumes that when unfatigued motor units are available (i.e., $M_{AD} < M_A + M_{UC}$), the brain's commands maintain a consistent activation scheme, leading to $\frac{dM_A}{dt} = \frac{dM_{AD}}{dt}$. However, this assumption becomes problematic over extended periods with dynamic fluctuations in activation. Figure 3.2 illustrates these issues within the context of a soccer-specific 1-minute drill.

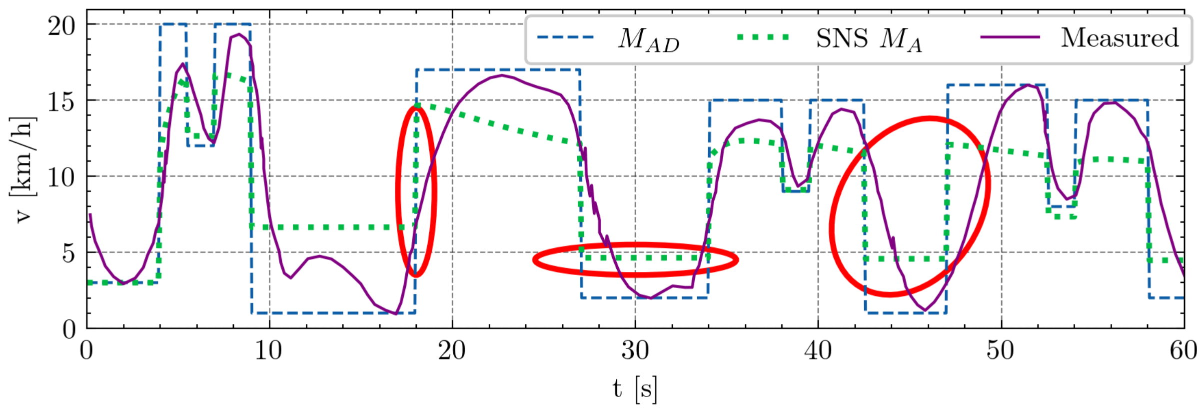

Figure 3.2: Soccer-specific 1-minute drill test [72]. The x-axis shows time in seconds, and the y-axis displays velocity in kilometers per hour (km/h). The green dotted line indicates the number of activated units (M_A), while the violet solid line represents the measured speed. The red ellipse highlights unrealistic activation rates, demonstrating challenges in numerically calculating the derivative, $\frac{dM_A}{dt} = \frac{dM_{AD}}{dt}$, as proposed by SNS.

The total number of motor units, M_0, available to a soccer player is calculated using the formula $M_0 = m \cdot v_{max}^2$, where m represents the player's mass, and v_{max} denotes the player's maximum velocity. This formula accounts for individual player characteristics, making it more tailored to each player's physical capabilities.

Unlike track events such as a 100-meter sprint, where the starting speed is 0 km/h and the goal is to reach maximum speed as quickly as possible, soccer involves complex and non-linear speed dynamics due to the unpredictable nature of the game. Consequently, the required change in active motor units is defined as $\Delta M_{AD} = M_{AD}(t) - M_A$, where ΔM_{AD} reflects the difference between the desired number of active motor units at time t and the current number of active units, M_A. This difference can represent either acceleration (positive ΔM_{AD}) or deceleration (negative ΔM_{AD}).

This refinement allows modification of the original approach for the change in active motor units, as shown in Equation (3.2). In this equation, α_A controls the rate of motor unit activation during acceleration, while β_D governs the rate of motor unit deactivation during deceleration. Despite these adjustments, the fatigue and recovery dynamics, governed by Equation (3.3), remain consistent with the original model.

$$\frac{dM_A}{dt} = max(0, \Delta M_{AD}) \cdot M_P \cdot \alpha_A + min(0, \Delta M_{AD}) \cdot M_A \cdot \beta_D \tag{3.2}$$

$$\frac{dM_F}{dt} = M_A \cdot F - M_F \cdot R \tag{3.3}$$

To test and validate the defined model, data was sourced from three key scientific studies and an additional dataset collected from a professional soccer club. The first study employed a pioneering approach in MFM using a maximal hand-grip test [47]. The other two studies were soccer-specific: one provided data from a 1-minute soccer-specific circuit performed by a representative player [72], and the other involved a 15-minute soccer-specific intermittent treadmill protocol [35].

In addition to these studies, data was collected from professional soccer players during the preseason. The players participated in a repeated sprint test, consisting of 10 maximal-effort sprints over an 80-meter distance. The second phase of data collection occurred during friendly and official matches over the following two months. During these matches, players were equipped with GPexe pro[2] devices, which recorded data at a sampling frequency of 18 Hz. Figure 3.3 presents a sample of the data collected from the repeated sprint test.

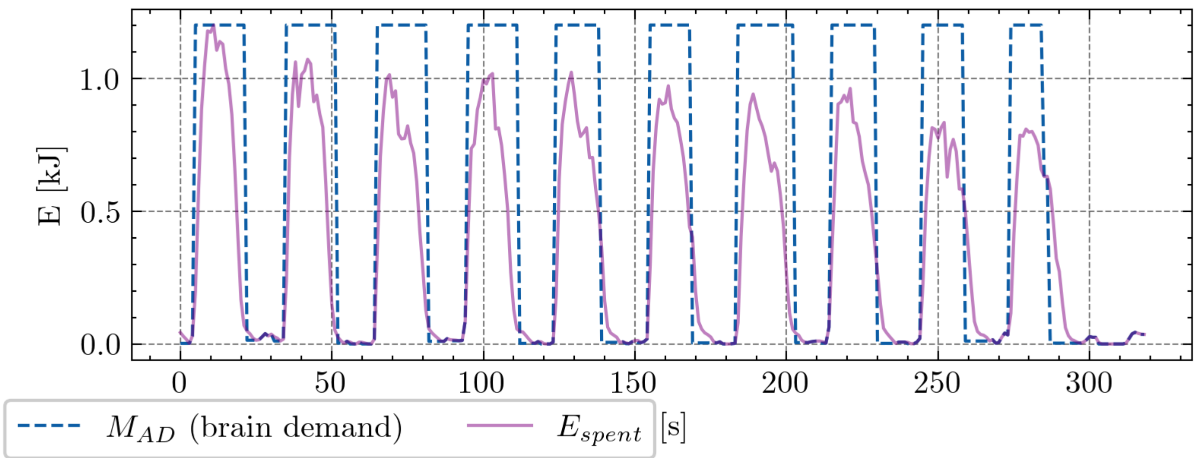

Figure 3.3: Visualization of the intended energy expenditure (M_{AD}) and actual energy spent (E_{spent}) during the 80-meter sprint test. Blue dotted lines represent the peak energy attained during the first sprint, used as a benchmark for the subsequent sprints. X-axis denotes time in seconds, and Y-axis indicates energy in kilojoules. Violet solid lines illustrate kinetic energy calculated from the measured velocity and athlete mass.

The optimal parameters for the proposed MFM model were determined using the Euler method with a time step of $\Delta t = 1$s. The model requires four parameters for each player: α_A, β_D, F, and R. Additionally, the constraint $M_0 - M_A - M_F \geq 0$ was applied to avoid the estimation of an unrealistic number of fatigued units M_F, which the model does not directly assess. Parameter values were bounded as specified in Equation (3.4).

$$\begin{aligned} 1 \cdot 10^{-5} \le \alpha_A \le 1 \cdot 10^{-3}, \\ 1 \cdot 10^{-5} \le \beta_D \le 1 \cdot 10^{-2}, \\ 1 \cdot 10^{-3} \le F \le 1 \cdot 10^{-1}, \\ 1 \cdot 10^{-3} \le R \le 1 \cdot 10^{-1}. \end{aligned} \tag{3.4}$$

These boundaries were established based on prior research and practical considerations of parameter ranges. The parameters were consolidated into the following optimization vector:

$$\mathbf{x} = \Big(\alpha_A, \beta_D, F, R\Big) \quad . \tag{3.5}$$

For each player i, the cost function was calculated as the squared error between the estimated and measured values, as shown in Equation (3.6). The model fitting process was conducted using the particle swarm optimization (PSO) method, known for its effectiveness in addressing various optimization problems [1]. To ensure the stability of the optimized parameters, the optimization process was repeated ten times for each player.

$$\varepsilon_i(\mathbf{x}) = \frac{\int_0^t (E_i(\mathbf{x}, t) - E_{measured,i}(t))^2 dt}{t}, \tag{3.6}$$

The modifications made to the model enhance its robustness against dynamic changes. However, there are some limitations related to the model's input assumptions. One such assumption is that a player always aims to achieve peak acceleration or deceleration immediately during matches. This may not accurately reflect real-world scenarios, where various contextual factors on the field can influence a player's actions. Another assumption is that the maximum motor unit capacity, M_0, or the maximum potential M_P, directly correlates with the player's kinetic energy. While this is physically valid, it remains uncertain whether the recorded maximum speed truly represents the player's absolute potential, as it may be affected by factors such as opposition pressure [66]. Despite these assumptions, they were retained to keep the model simple while still providing meaningful insights into player performance.

Chapter 4

Conclusion

4.1 Main Contributions

As outlined in the introduction, data analysis in soccer is becoming increasingly important. This dissertation seeks to deepen practitioners' understanding of key aspects of soccer players' performance, including physical capabilities, endurance, cognitive function, and recovery. The underlying motivation for this work is to gain more precise insights at the individual level, acknowledging that each athlete is unique and should be treated as such. Data analysis, when combined with ML and optimization techniques, has proven effective in uncovering insights that managers, conditioning coaches, and medical staff can use to enhance their day-to-day decision-making. The primary scientific contributions of this dissertation are as follows:

- **Framework for Processing GPS Wearable Data and Intensity Classification:** A framework was developed for processing GPS wearable data, enabling the analysis of player performance on a minute-by-minute basis. An unsupervised learning approach using K-means clustering was implemented to classify playing minutes into low, moderate, and high-intensity clusters. This method simplifies the evaluation of intensity distribution throughout a match by reducing complex metrics derived from wearable sensors into three distinct intensity categories.

- **Mathematical Model for Individualized Player Profiling:** A mathematical model was created to develop individualized player profiles, considering pre-game expectations, goal difference fluctuations during the match, and player endurance. Data for this model was collected over one season from a professional soccer club, with all data and code made publicly available for replication and further research. This approach aids coaches in making informed substitution decisions based on the match context and score dynamics.

- **Quantification of Cognitive Load and Physical Performance Relationship:** An approach was presented to quantify the CL of soccer players, examining its relationship with physical performance during the latter stages of a soccer season. Additionally, a clustering algorithm was developed to categorize players into positive and negative motivation groups. This approach was validated against the coach's assessments, demonstrating a strong alignment with their evaluations.

- **ML Model for Estimating Recovery Duration Post-Injury:** An ML model was developed to estimate recovery duration following muscle injuries. The approach demonstrated how data can be collected to provide valuable insights into injury recovery. The ML models, particularly the XGB algorithm, demonstrated improved accuracy and outperformed expert estimates in predicting recovery time.

- **Adjustment of the MFM for Dynamic Adaptation:** The MFM was adapted to accommodate dynamic changes in the number of active and unfatigued motor units demanded (M_{AD}), enabling the application of optimization methods such as PSO. This adjustment allows for the determination of individual player MFM parameters during repeated sprint tests, facilitating the evaluation of energy expenditure throughout a match and the detection of performance declines. The data was sourced from existing literature and recorded by a professional soccer club, with all data and code made open-source for future research.

Considering all the scientific contributions, this dissertation lays the groundwork for multiple areas of research within the soccer domain. As the field of sports continues to evolve rapidly, a multidisciplinary approach is essential for advancing knowledge and generating new insights. By harnessing data over an extended acquisition period, additional opportunities will arise to further refine and enhance the methodologies presented in this dissertation.

4.2 Future Work

The advancement of sensor technology, data collection methods, and the decreasing cost of such systems present significant opportunities for enhancing the existing research. Potential areas for future work include:

- The use of more advanced sensors and video-tracking systems capable of extracting data at higher sampling rates can reduce the complexity of preprocessing, thereby providing more reliable data. This would allow researchers to focus primarily on developing and optimizing ML models for match intensity classification. The collaboration between coaches, researchers, and data scientists within soccer clubs could lead to the automation of intensity classification, potentially eliminating the need to monitor multiple wearable metrics simultaneously, which is particularly challenging when managing an entire team.

- The model for individualized player profiling can be refined by extending the research to encompass data from an entire league or multiple leagues. With a larger dataset, it would be possible to track changes in score with greater granularity. For example, a team's response to narrowing a two-goal deficit differs from the impact of conceding a goal in a previously tied match. Such scenarios significantly influence player responses, especially as matches progress to their final stages. A more extensive dataset could provide insights into the mental aspects of performance and their effect on physical output, offering a more comprehensive and robust player profile.

- The relationship between CL and physical performance can be further validated by extending the data collection across an entire season and including multiple teams from different leagues. Additionally, the robustness of the automatic clustering method that categorizes players into positive and negative motivation groups should be tested on a larger dataset to confirm its generalizability and accuracy.

- The injury duration model should be tested in a real-world, production environment and expanded to include various types of injuries as more data becomes available. While model accuracy is crucial, it may need to be balanced with explainability, which is vital in decision-making processes that affect player health. Future research in this area should also address the challenges posed by the limited availability of medical records and the legal concerns that need to be resolved between research groups, soccer clubs, and the players themselves.

- The primary limitation of the MFM, specifically the determination of the desired activation profile (M_{AD}) during soccer matches, could be mitigated by incorporating contextual information from the soccer pitch. This would require access to tracking data for all players on the field, along with event data. Such information would provide a deeper understanding of player actions, such as sprinting, jogging, or jumping, in relation to their intentions and the broader match context.

Abbreviations

The following abbreviations are used in this dissertation:

SNS Sih, Ng, and Stuhmiller [74]

MFM muscle fatigue model

GPS global positioning system

RPE rate of perceived exertion

RTP return-to-play

PSO particle swarm optimization

MSE mean squared error

MAE mean absolute error

MAPE mean absolute percentage error

RMSE root mean squared error

LOO leave-one-out

CV cross-validation

CoV coefficient of variation

XGB Extreme Gradient Boosting

DT Decision Trees

LR Linear Regression

GD goal difference

FMS functional movement screenings

HR heart rate

TD total distance

HSR high-speed running

MPE metabolic power events

CL cognitive load

ML machine learning

CAD computer-aided diagnosis

HIA high-intensity actions

PCA Principal Component Analysis

DBSCAN Density-Based Spatial Clustering of Applications with Noise

SNE stochastic neighbor embedding

NASA-TLX NASA Task Load Index

ANOVA Analysis of Variance

MANOVA Multivariate Analysis of Variance

HSD honestly significant difference

MRI magnetic resonance imaging

BAMIC British Athletic Muscle Injury Classification

Chapter 5

Summary of Papers

A Extended Energy-Expenditure Model in Soccer: Evaluating Player Performance in the Context of the Game

Every soccer game influences each player's performance differently. Many studies have tried to explain the influence of different parameters on the game; however, none went deeper into the core and examined it minute-by-minute. The goal of this study is to use data derived from GPS wearable devices to present a new framework for performance analysis. A player's energy expenditure is analyzed using data analytics and K-means clustering of low-, middle-, and high-intensity periods distributed in 1 min segments. Our framework exhibits a higher explanatory power compared to usual game metrics (e.g., high-speed running and sprinting), explaining 45.91% of the coefficient of variation vs. 21.32% for high-, 30.66% vs. 16.82% for middle-, and 24.41% vs. 19.12% for low-intensity periods. The proposed methods enable deeper game analysis, which can help strength and conditioning coaches and managers in gaining better insights into the players' responses to various game situations.

Skoki A., Rossi A., Cintia P., Pappalardo L., Štajduhar I., Extended Energy-Expenditure Model in Soccer: Evaluating Player Performance in the Context of the Game. Sensors. 2022; 22(24):9842. https://doi.org/10.3390/s22249842

B Revolutionizing Soccer Injury Management: Predicting Muscle Injury Recovery Time Using ML

Predicting the optimal recovery time following a soccer player's injury is a complex task with heavy implications on team performance. While most current decision-based models rely on the physician's perspective, this study proposes a ML-based approach to predict recovery duration using three modeling techniques: linear regression, decision tree, and extreme gradient boosting (XGB). Performance is compared between the models, against the expert, and together with the expert. The results demonstrate that integrating the expert's predictions as a feature improves the performance of all models, with XGB performing best with a mean R^2 score of 0.72, outperforming the expert's predictions with an R^2 score of 0.62. This approach has significant implications for sports medicine, as it could help teams make better decisions on the return-to-play of their players, leading to improved performance and reduced risk of re-injury.

Skoki A., Napravnik M., Polonijo M., Štajduhar I., Lerga J., Revolutionizing Soccer Injury Management: Predicting Muscle Injury Recovery Time Using ML. Applied Sciences. 2023; 13(10):6222. https://doi.org/10.3390/app13106222

C Building Individual Player Performance Profiles According to Pre-Game Expectations and Goal Difference in Soccer

Soccer player performance is influenced by multiple unpredictable factors. During a game, score changes and pre-game expectations affect the effort exerted by players. This study used GPS wearable sensors to track players' energy expenditure in 5-min intervals, alongside recording the goal timings and the win and lose probabilities from betting sites. A mathematical model was developed that considers pre-game expectations (e.g., favorite, non-favorite), endurance, and goal difference (GD) dynamics on player effort. Particle Swarm and Nelder–Mead optimization methods were used to construct these models, both consistently converging to similar cost function values. The model outperformed baselines relying solely on mean and median power per GD. This improvement is underscored by the mean absolute error (MAE) of 396.87 ± 61.42 and root mean squared error (RMSE) of 520.69 ± 88.66 achieved by our model, as opposed to the B_1 MAE of 429.04 ± 84.87 and RMSE of 581.34 ± 185.84, and B_2 MAE of 421.57 ± 95.96 and RMSE of 613.47 ± 300.11 observed across all players in the dataset. This research offers an enhancement to the current approaches for assessing players' responses to contextual factors, particularly GD. By utilizing wearable data and contextual factors, the proposed methods have the potential to improve decision-making and deepen the understanding of individual player characteristics.

Skoki A., Gašparović B., Ivić S., Lerga J., Štajduhar I., Building Individual Player Performance Profiles According to Pre-Game Expectations and Goal Difference in Soccer. Sensors. 2024; 24(5):1700. https://doi.org/10.3390/s24051700

D Exploring the Impact of the Perceived Cognitive Load on the Physical Performance in Soccer

The impact of an increased cognitive load (CL) on the players performance, encompassing tactical and technical aspects, is well-documented, yet its impact on the physical performance remains uncertain. Unlike prior research focusing on controlled settings, our study uses GPS wearable sensors and the NASA-TLX questionnaire to assess the physical and cognitive relationship during the national championship final round. The collected data serves a dual purpose: firstly, to quantify the overall perceived CL by averaging the NASA-TLX components and calculating the Z-scores for individual players, and secondly, to investigate the feasibility of constructing cognitive clusters based on positive and negative motivation profiles derived from the questionnaire responses. Data analysis reveals a significant difference between the low and moderate CL clusters in terms of the total distance (TD), high-speed running distance (HSR), and deceleration (DEC) parameters. Further differences are observed between the moderate and high CL clusters for TD and distance per minute (DPM) metrics. To refine our understanding, the players are systematically categorized into positive-aided and negative-aided motivation categories using the K-means clustering. The resulting clusters exhibit a strong F1 score of 0.8696 closely aligning with the coach evaluations. However, no significant differences are observed between the motivation groups and physical performance. The findings underscore the adverse impact of CL on the soccer players physical performance.

Skoki A., Petra A., Ljubić S., Naglić A., Lerga J., Štajduhar I., Exploring the Impact of the Perceived Cognitive Load on the Physical Performance in Soccer. Journal of Electrical Engineering and Computer Science. 2024; 91(3), p.108-116.

E Enhancing Biophysical Muscle Fatigue Model in the Dynamic Context of Soccer

In the field of muscle fatigue models (MFMs), the prior research has demonstrated success in fitting data in specific contexts, but it falls short in addressing the diverse efforts and rapid changes in exertion typical of soccer matches. This study builds upon the existing model, aiming to enhance its applicability and robustness to dynamic demand shifts. The objective is to encapsulate the complexities of soccer dynamics with a streamlined set of parameters. Our refined model achieved a slight improvement in the R^2 score in the maximum hand-grip test, increasing from 0.87 to 0.89 compared to the existing model. It also demonstrated dynamic change robustness in a soccer-specific 1 min drill and 15 min treadmill protocol extracted from the literature. Through individualized fitting on a 10-repetition 80 m sprint test for a soccer player, the model exhibited R^2 scores between 0.62 and 0.80. Furthermore, when tested with actual soccer match data, it maintained a robust performance, with the average R^2 scores ranging from 0.70 to 0.72. The proposed approach holds the potential to advance the understanding of tactical decisions by correlating them with real-time physical performance, offering opportunities for more informed strategies and ultimately enhancing team performance.

Skoki A., Ivić S., Ljubić S., Lerga J., Štajduhar I., Enhancing Biophysical Muscle Fatigue Model in the Dynamic Context of Soccer. Sensors, 24(24), 8128. https://doi.org/10.3390/s24248128

Bibliography

[1] Ahila, R., Sadasivam, V., and Manimala, K. Particle Swarm Optimization-Based Feature Selection and Parameter Optimization for Power System Disturbances Classification. *Applied Artificial Intelligence 26*, 9 (2012), 832–861.

[2] Ahmed, M., Seraj, R., and Islam, S. M. S. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics 9*, 8 (Aug. 2020), 1295. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

[3] Alarcón, F., Castillo-Díaz, A., Madinabeitia, I., and Castillo-Rodríguez, A. La carga mental deteriora la precisión del pase en jugadores de fútbol. *Journal of Sport Psychology* (2018).

[4] Andrzejewski, M., Konefał, M., Chmura, P., Kowalczuk, E., and Chmura, J. Match outcome and distances covered at various speeds in match play by elite german soccer players. *International Journal of Performance Analysis in Sport 16*, 3 (2016), 817–828.

[5] Aquino, R., Munhoz Martins, G. H., Palucci Vieira, L. H., and Menezes, R. P. Influence of Match Location, Quality of Opponents, and Match Status on Movement Patterns in Brazilian Professional Football Players. *The Journal of Strength & Conditioning Research 31*, 8 (Aug. 2017), 2155–2161.

[6] Arney, B. E., Glover, R., Fusco, A., Cortis, C., Koning, J. J. d., Erp, T. v., Jaime, S., Mikat, R. P., Porcari, J. P., and Foster, C. Comparison of RPE (Rating of Perceived Exertion) Scales for Session RPE. *International Journal of Sports Physiology and Performance 14*, 7 (Aug. 2019), 994–996. Publisher: Human Kinetics Section: International Journal of Sports Physiology and Performance.

[7] Audia, P. G., and Locke, E. A. Benefiting from negative feedback. *Human Resource Management Review 13* (12 2003), 631–646.

[8] Badin, O. O., Smith, M. R., Conte, D., and Coutts, A. J. Mental Fatigue: Impairment of Technical Performance in Small-Sided Soccer Games. *International Journal of Sports Physiology and Performance 11*, 8 (Nov. 2016), 1100–1105.

[9] Barrera, J., Sarmento, H., Clemente, F. M., Field, A., and Figueiredo, A. J. The effect of contextual variables on match performance across different playing positions in professional portuguese soccer players. *International Journal of Environmental Research and Public Health 18*, 10 (2021), 5175.

[10] Bernichon, T., Cook, K. E., and Brown, J. D. Seeking self-evaluative feedback: The interactive role of global self-esteem and specific self-views. *Journal of Personality and Social Psychology 84* (2003), 194–204.

[11] Bloomfield, J., Polman, R., and Donoghue, P. Effects of score-line on intensity of play in midfield and forward players in fa premier league. *Journal of Sports Sciences 23*, 2 (2005), 191–192.

[12] Boksem, M. A., and Tops, M. Mental fatigue: Costs and benefits. *Brain Research Reviews 59*, 1 (2008), 125–139.

[13] Borresen, J., and Lambert, M. I. Quantifying Training Load: A Comparison of Subjective and Objective Methods. *International Journal of Sports Physiology and Performance 3*, 1 (Mar. 2008), 16–30. Publisher: Human Kinetics, Inc. Section: International Journal of Sports Physiology and Performance.

[14] Brownsberger, J., Edwards, A., Crowther, R., and Cottrell, D. Impact of mental fatigue on self-paced exercise. *International Journal of Sports Medicine 34*, 12 (2013), 1029–1036.

[15] Buchheit, M. The 30-15 Intermittent Fitness Test: Accuracy for Individualizing Interval Training of Young Intermittent Sport Players. *The Journal of Strength & Conditioning Research 22*, 2 (Mar. 2008), 365.

[16] Buchheit, M., Racinais, S., Bilsborough, J. C., Bourdon, P. C., Voss, S. C., Hocking, J., Cordy, J., Mendez-Villanueva, A., and Coutts, A. J. Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players. *Journal of Science and Medicine in Sport 16*, 6 (Nov. 2013), 550–555.

[17] Castellano, J., Alvarez-Pastor, D., and Bradley, P. S. Evaluation of Research Using Computerised Tracking Systems (Amisco® and Prozone®) to Analyse Physical Performance in Elite Soccer: A Systematic Review. *Sports Medicine 44*, 5 (May 2014), 701–712.

[18] Castellano, J., Blanco-Villaseñor, A., and Alvarez, D. Contextual variables and time-motion analysis in soccer. *International journal of sports medicine 32*, 06 (2011), 415–421.

[19] Chan, H.-P., Hadjiiski, L. M., and Samala, R. K. Computer-aided diagnosis in the era of deep learning. *Medical Physics 47*, 5 (2020), e218–e227. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mp.13764.

[20] Chomiak, J., Junge, A., Peterson, L., and Dvorak, J. Severe Injuries in Football Players. *The American Journal of Sports Medicine 28*, 5_suppl (Sept. 2000), 58–68. Publisher: SAGE Publications Inc STM.

[21] Coutinho, D., Gonçalves, B., Wong, D. P., Travassos, B., Coutts, A. J., and Sampaio, J. Exploring the effects of mental and muscular fatigue in soccer players' performance. *Human Movement Science 58* (Apr. 2018), 287–296.

[22] Cowden, R. G., Mascret, N., and Duckett, T. R. A person-centered approach to achievement goal orientations in competitive tennis players: Associations with motivation and mental toughness. *Journal of Sport and Health Science 10*, 1 (Jan. 2021), 73–81.

[23] Cumming, S. P., Smith, R. E., and Smoll, F. L. Athlete-perceived coaching behaviors: Relating two measurement traditions. *Journal of Sport and Exercise Psychology 28*, 2 (2006), 205–213.

[24] Dantzer, R., Heijnen, C. J., Kavelaars, A., Laye, S., and Capuron, L. The neuroimmune basis of fatigue. *Trends in Neurosciences 37*, 1 (2014), 39–46.

[25] Ekstrand, J., Hägglund, M., and Waldén, M. Epidemiology of Muscle Injuries in Professional Football (Soccer). *The American Journal of Sports Medicine 39*, 6 (June 2011), 1226–1232. Publisher: SAGE Publications Inc STM.

[26] Ekstrand, J., Hägglund, M., and Waldén, M. Injury incidence and injury patterns in professional football: the UEFA injury study. *British Journal of Sports Medicine 45*, 7 (June 2011), 553–558. Publisher: British Association of Sport and Excercise Medicine Section: Original articles.

[27] Ekstrand, J., Krutsch, W., Spreco, A., Zoest, W. v., Roberts, C., Meyer, T., and Bengtsson, H. Time before return to play for the most common injuries in professional football: a 16-year follow-up of the UEFA Elite Club Injury Study. *British Journal of Sports Medicine 54*, 7 (Apr. 2020), 421–426. Publisher: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine Section: Original research.

[28] Eliakim, E., Morgulev, E., Lidor, R., and Meckel, Y. Estimation of injury costs: financial damage of English Premier League teams' underachievement due to injuries. *BMJ Open Sport & Exercise Medicine 6*, 1 (May 2020), e000675. Publisher: BMJ Specialist Journals Section: Original research.

[29] Fransen, K., Boen, F., Vansteenkiste, M., Mertens, N., and Broek, G. V. The power of competence support: The impact of coaches and athlete leaders on intrinsic motivation and performance. *Scandinavian Journal of Medicine and Science in Sports 28* (2 2018), 725–745.

[30] Frey-Law, L. A., Schaffer, M., and Urban, F. K. Muscle fatigue modelling: Solving for fatigue and recovery parameter values using fewer maximum effort assessments. *International Journal of Industrial Ergonomics 82* (Mar. 2021), 103104.

[31] Galy, E., Cariou, M., and Mélan, C. What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology 83*, 3 (2012), 269–275.

[32] García-Unanue, J., Pérez-Gómez, J., Giménez, J.-V., Felipe, J. L., Gómez-Pomares, S., Gallardo, L., and Sánchez-Sánchez, J. Influence of contextual variables and the pressure to keep category on physical match performance in soccer players. *PLOS ONE 13*, 9 (Sept. 2018), e0204256. Publisher: Public Library of Science.

[33] Gawron, V. J., French, J., and Funke, D. An Overview of Fatigue. In *Human Factors in Transportation. Stress, Workload, and Fatigue*, &. P. A. D. P. A. Hancock, Ed., 1 ed. Lawrence Erlbaum Associates, Inc, Mahwah, NJ, 2001, ch. An Overvie, pp. 581–595.

[34] Gede, G., and Hubbard, M. A bioenergetic model for simulating athletic performance of intermediate duration. *Journal of Biomechanics 47*, 14 (Nov. 2014), 3448–3453.

[35] Greig, M., and Siegler, J. Soccer-Specific Fatigue and Eccentric Hamstrings Muscle Strength. *Journal of athletic training 44* (Mar. 2009), 180–4.

[36] Grinsztajn, L., Oyallon, E., and Varoquaux, G. Why do tree-based models still outperform deep learning on tabular data?, 2022.

[37] Habay, J., Uylenbroeck, R., Van Droogenbroeck, R., De Wachter, J., Proost, M., Tassignon, B., De Pauw, K., Meeusen, R., Pattyn, N., Van Cutsem, J., and Roelands, B. Interindividual Variability in Mental Fatigue-Related Impairments in Endurance Performance: A Systematic Review and Multiple Meta-regression. *Sports Medicine - Open 9*, 1 (2023).

[38] Hagberg, M. Muscular endurance and surface electromyogram in isometric and dynamic exercise. *Journal of Applied Physiology Respiratory Environmental and Exercise Physiology 51*, 1 (1981), 1–7.

[39] Hallén, A., and Ekstrand, J. Return to play following muscle injuries in professional footballers. *Journal of Sports Sciences 32*, 13 (Aug. 2014), 1229–1236.

[40] Hart, S. G., and Staveland, L. E. Development of NASA-TLX. *Human Mental Workload. Advances in Psychology*, 52 (1988), 139–183.

[41] Hennessy, L., and Jeffreys, I. The Current Use of GPS, Its Potential, and Limitations in Soccer. *Strength & Conditioning Journal 40*, 3 (June 2018), 83–94.

[42] Khaustov, V., and Mozgovoy, M. Recognizing Events in Spatiotemporal Soccer Data. *Applied Sciences 10*, 22 (Jan. 2020), 8046. Number: 22 Publisher: Multidisciplinary Digital Publishing Institute.

[43] Kluger, A. N., and Denisi, A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory we argue that a considerable body of evidence suggesting that feedback intervention (fi) effects on performance are quite vari. *Psychological Bulletin 9* (1996), 254–284.

[44] Kunrath, C. A., Nakamura, F. Y., Roca, A., Tessitore, A., and Teoldo Da Costa, I. How does mental fatigue affect soccer performance during small-sided games? A cognitive, tactical and physical approach. *Journal of Sports Sciences 38*, 15 (Aug. 2020), 1818–1828.

[45] Lago, C., Casais, L., Dominguez, E., and Sampaio, J. The effects of situational variables on distance covered at various speeds in elite soccer. *European journal of sport science 10*, 2 (2010), 103–109.

[46] Lago-Peñas, C., Kalén, A., Lorenzo-Martinez, M., López-Del Campo, R., Resta, R., and Rey, E. Do elite soccer players cover longer distance when losing? differences between attackers and defenders. *International Journal of Sports Science & Coaching 16*, 3 (2021), 840–847.

[47] Liu, J. Z., Brown, R. W., and Yue, G. H. A dynamical model of muscle activation, fatigue, and recovery. *Biophysical Journal 82*, 5 (May 2002), 2344–2359.

[48] Looft, J. M. *Adaptation and validation of an analytical localized muscle fatigue model for workplace tasks*. PhD thesis, University of Iowa, 2014.

[49] Lorenzo-Martinez, M., Kalén, A., Rey, E., López-Del Campo, R., Resta, R., and Lago-Peñas, C. Do elite soccer players cover less distance when their team spent more time in possession of the ball? *Science and Medicine in Football 5*, 4 (2021), 310–316.

[50] Lorenzo-Martinez, M., Kalén, A., Rey, E., López-Del Campo, R., Resta, R., and Lago-Peñas, C. Do elite soccer players cover less distance when their team spent more time in possession of the ball? *Science & Medicine in Football 5*, 4 (2021), 310–316.

[51] Lotfi, G., Rabavi, A., and Jafarzadeh, M. Positive and negative motivational self-talk affect learning of soccer kick in novice players, mediated by anxiety. *International Journal of Humanities and Cultural Studies 3* (Feb. 2016), 1945–1952.

[52] López-Valenciano, A., Ruiz-Pérez, I., Garcia-Gómez, A., Vera-Garcia, F. J., Croix, M. D. S., Myer, G. D., and Ayala, F. Epidemiology of injuries in professional football: a systematic review and meta-analysis. *British Journal of Sports Medicine 54*, 12 (June

2020), 711–718. Publisher: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine Section: Review.

[53] Marcora, S. M., and Staiano, W. The limit to exercise tolerance in humans: Mind over muscle? *European Journal of Applied Physiology 109*, 4 (2010), 763–770.

[54] Marques, V. B., Medeiros, T. M., de Souza Stigger, F., Nakamura, F. Y., and Baroni, B. M. The functional movement screen (fms™) in elite young soccer players between 14 and 20 years: Composite score, individual-test scores and asymmetries. *International Journal of Sports Physical Therapy 12*, 6 (Nov. 2017), 977–985.

[55] Matheson, G. O., Shultz, R., Bido, J., Mitten, M. J., Meeuwisse, W. H., and Shrier, I. Return-to-play decisions: are they the team physician's responsibility? *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine 21*, 1 (Jan. 2011), 25–30.

[56] Moalla, W., Fessi, M. S., Makni, E., Dellal, A., Filetti, C., Di Salvo, V., and Chamari, K. Association of physical and technical activities with partial match status in a soccer professional team. *The Journal of Strength & Conditioning Research 32*, 6 (2018), 1708–1714.

[57] Na, K. The effects of cognitive load on query reformulation: mental demand, temporal demand and frustration. *Aslib Journal of Information Management 73*, 3 (2021), 436–453. Publisher: Emerald Publishing Limited.

[58] Nobari, H., Oliveira, R., Brito, J. P., Pérez-Gómez, J., Clemente, F. M., and Ardigò, L. P. Comparison of Running Distance Variables and Body Load in Competitions Based on Their Results: A Full-Season Study of Professional Soccer Players. *International Journal of Environmental Research and Public Health 18*, 4 (Jan. 2021), 2077. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[59] Oliva-Lozano, J. M., Fortes, V., Krustrup, P., and Muyor, J. M. Acceleration and sprint profiles of professional male football players in relation to playing position. *PLOS ONE 15*, 8 (2020), e0236959. Publisher: Public Library of Science.

[60] Oliveira, R., Brito, J. P., Martins, A., Mendes, B., Marinho, D. A., Ferraz, R., and Marques, M. C. In-season internal and external training load quantification of an elite European soccer team. *PLOS ONE 14*, 4 (Apr. 2019), e0209393. Publisher: Public Library of Science.

[61] Palucci Vieira, L. H., Aquino, R., Lago-Peñas, C., Munhoz Martins, G. H., Puggina, E. F., and Barbieri, F. A. Running Performance in Brazilian Professional Football Players During a Congested Match Schedule. *The Journal of Strength & Conditioning Research 32*, 2 (2018), 313–325.

[62] Peterson, R. S., and Behfar, K. J. The dynamic relationship between performance feedback, trust, and conflict in groups: A longitudinal study. *Organizational Behavior and Human Decision Processes 92* (2003), 102–112.

[63] Ponce-Bordón, J. C., Díaz-García, J., López-Gajardo, M. A., Lobo-Triviño, D., López del Campo, R., Resta, R., and García-Calvo, T. The influence of time winning and time losing on position-specific match physical demands in the top one spanish soccer league. *Sensors 21*, 20 (2021), 6843.

[64] Ponce-Bordón, J. C., Díaz-García, J., López-Gajardo, M. A., Lobo-Triviño, D., López del Campo, R., Resta, R., and García-Calvo, T. The Influence of Time Winning and Time Losing on Position-Specific Match Physical Demands in the Top One Spanish Soccer League. *Sensors 21*, 20 (Jan. 2021), 6843. Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

[65] Rago, V., Silva, J., Mohr, M., Randers, M., Barreira, D., Krustrup, P., and Rebelo, A. Influence of opponent standard on activity profile and fatigue development during preseasonal friendly soccer matches: a team study. *Research in Sports Medicine 26*, 4 (2018), 413–424. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15438627.2018.1492400.

[66] Rampinini, E., Coutts, A. J., Castagna, C., Sassi, R., and Impellizzeri, F. M. Variation in Top Level Soccer Match Performance. *International Journal of Sports Medicine 28*, 12 (Dec. 2007), 1018–1024. Publisher: © Georg Thieme Verlag KG Stuttgart · New York.

[67] Rashedi, E., and Nussbaum, M. A. A review of occupationally-relevant models of localised muscle fatigue. *International Journal of Human Factors Modelling and Simulation 5*, 1 (2015), 61.

[68] Redwood-Brown, A. J., O’Donoghue, P. G., Nevill, A. M., Saward, C., Dyer, N., and Sunderland, C. Effects of situational variables on the physical activity profiles of elite soccer players in different score line states. *Scandinavian journal of medicine & science in sports 28*, 12 (2018), 2515–2526.

[69] Reilly, T., Drust, B., and Clarke, N. Muscle Fatigue during Football Match-Play. *Sports Medicine 38*, 5 (May 2008), 357–367.

[70] Richter, C., O’Reilly, M., and Delahunt, E. Machine learning in sports science: challenges and opportunities. *Sports Biomechanics 0*, 0 (2021), 1–7. Publisher: Routledge _eprint: https://doi.org/10.1080/14763141.2021.1910334.

[71] Richter, C., O’Reilly, M., and Delahunt, E. Machine learning in sports science: challenges and opportunities. *Sports Biomechanics 23*, 8 (Aug. 2024), 961–967. Publisher: Routledge _eprint: https://doi.org/10.1080/14763141.2021.1910334.

[72] Savoia, C., Padulo, J., Colli, R., Marra, E., McRobert, A., Chester, N., Azzone, V., Pullinger, S. A., and Doran, D. A. The Validity of an Updated Metabolic Power Algorithm Based upon di Prampero's Theoretical Model in Elite Soccer Players. *International Journal of Environmental Research and Public Health 17*, 24 (2020), 9554.

[73] Shaw, J., and O'Donoghue, P. The effect of scoreline on work rate in amateur soccer. *Notational analysis of sport VI* (2004), 84–91.

[74] Sih, B., Ng, L., and Stuhmiller, J. Generalization of a Model Based on Biophysical Concepts of Muscle Activation, Fatigue and Recovery that Explains Exercise Performance. *International Journal of Sports Medicine 33*, 04 (Apr. 2012), 258–267.

[75] Smith, M. R., Thompson, C., Marcora, S. M., Skorski, S., Meyer, T., and Coutts, A. J. Mental Fatigue and Soccer: Current Knowledge and Future Directions. *Sports Medicine (Auckland, N.Z.) 48*, 7 (July 2018), 1525–1532.

[76] Smith, M. R., Zeuwts, L., Lenoir, M., Hens, N., De Jong, L. M. S., and Coutts, A. J. Mental fatigue impairs soccer-specific decision-making skill. *Journal of Sports Sciences 34*, 14 (July 2016), 1297–1304.

[77] Snyders, C., Pyne, D. B., Sewry, N., Hull, J. H., Kaulback, K., and Schwellnus, M. Acute respiratory illness and return to sport: a systematic review and meta-analysis by a subgroup of the IOC consensus on 'acute respiratory illness in the athlete'. *British Journal of Sports Medicine 56*, 4 (Feb. 2022), 223–231.

[78] Sonne, M. W., and Potvin, J. R. A modified version of the three-compartment model to predict fatigue during submaximal tasks with complex force-time histories. *Ergonomics 59*, 1 (Jan. 2016), 85–98. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00140139.2015.1051597.

[79] Theodoropoulos, J. S., Bettle, J., and Kosy, J. D. The use of GPS and inertial devices for player monitoring in team sports: A review of current and future applications. *Orthopedic Reviews 12*, 1 (Apr. 2020), 7863.

[80] Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., and Gregson, W. Tracking Morning Fatigue Status Across In-Season Training Weeks in Elite Soccer Players. *International Journal of Sports Physiology and Performance 11*, 7 (Oct. 2016), 947–952. Publisher: Human Kinetics, Inc. Section: International Journal of Sports Physiology and Performance.

[81] Torreño, N., Munguía-Izquierdo, D., Coutts, A., Villarreal, E. S. d., Asian-Clemente, J., and Suarez-Arrones, L. Relationship Between External and Internal Loads of Professional Soccer Players During Full Matches in Official Games Using Global Positioning Systems

and Heart-Rate Technology. *International Journal of Sports Physiology and Performance 11*, 7 (Oct. 2016), 940–946. Publisher: Human Kinetics, Inc. Section: International Journal of Sports Physiology and Performance.

[82] Valle, X., Mechó, S., Alentorn-Geli, E., Järvinen, T. A. H., Lempainen, L., Pruna, R., Monllau, J. C., Rodas, G., Isern-Kebschull, J., Ghrairi, M., Yanguas, X., Balius, R., and la Torre, A. M.-D. Return to Play Prediction Accuracy of the MLG-R Classification System for Hamstring Injuries in Football Players: A Machine Learning Approach. *Sports Medicine 52*, 9 (Sept. 2022), 2271–2282.

[83] Waldén, M., Hägglund, M., and Ekstrand, J. UEFA Champions League study: a prospective study of injuries in professional football during the 2001–2002 season. *British Journal of Sports Medicine 39*, 8 (Aug. 2005), 542–546. Publisher: British Association of Sport and Excercise Medicine Section: Original article.

[84] Winkel, D. J., Tong, A., Lou, B., Kamen, A., Comaniciu, D., Disselhorst, J. A., Rodríguez-Ruiz, A., Huisman, H., Szolar, D., Shabunin, I., Choi, M. H., Xing, P., Penzkofer, T., Grimm, R., von Busch, H., and Boll, D. T. A Novel Deep Learning Based Computer-Aided Diagnosis System Improves the Accuracy and Efficiency of Radiologists in Reading Biparametric Magnetic Resonance Images of the Prostate: Results of a Multireader, Multicase Study. *Investigative Radiology 56*, 10 (Oct. 2021), 605.

[85] Winkel, D. J., Tong, A., Lou, B., Kamen, A., Comaniciu, D., Disselhorst, J. A., Rodríguez-Ruiz, A., Huisman, H., Szolar, D., Shabunin, I., Choi, M. H., Xing, P., Penzkofer, T., Grimm, R., von Busch, H., and Boll, D. T. A Novel Deep Learning Based Computer-Aided Diagnosis System Improves the Accuracy and Efficiency of Radiologists in Reading Biparametric Magnetic Resonance Images of the Prostate: Results of a Multireader, Multicase Study. *Investigative Radiology 56*, 10 (Oct. 2021), 605.

[86] Wong, P.-l., Chaouachi, A., Chamari, K., Dellal, A., and Wisloff, U. Effect of Preseason Concurrent Muscular Strength and High-Intensity Interval Training in Professional Soccer Players. *The Journal of Strength & Conditioning Research 24*, 3 (Mar. 2010), 653.

[87] Writing Committee, Gluckman, T. J., Bhave, N. M., Allen, L. A., Chung, E. H., Spatz, E. S., Ammirati, E., Baggish, A. L., Bozkurt, B., Cornwell, W. K., Harmon, K. G., Kim, J. H., Lala, A., Levine, B. D., Martinez, M. W., Onuma, O., Phelan, D., Puntmann, V. O., Rajpal, S., Taub, P. R., and Verma, A. K. 2022 ACC Expert Consensus Decision Pathway on Cardiovascular Sequelae of COVID-19 in Adults: Myocarditis and Other Myocardial Involvement, Post-Acute Sequelae of SARS-CoV-2 Infection, and Return to Play: A Report of the American College of Cardiology Solution Set Oversight Committee. *Journal of the American College of Cardiology 79*, 17 (May 2022), 1717–1756.

[88] Xia, T., and Frey Law, L. A. A theoretical approach for modeling peripheral muscle fatigue and recovery. *Journal of Biomechanics 41*, 14 (Oct. 2008), 3046–3052.

List of Figures

1.1 Flowchart illustrating the various data sources utilized in the dissertation and the corresponding outcomes derived from different components of the study. Blue areas indicate the data sources, yellow areas highlight the two primary topics of the dissertation, and green rectangles encapsulate all dissertation outcomes. . . 6

2.1 Comparison of GM-5MIN and MPE energy expenditure over 90 minutes, with blue bars representing GM-5MIN values throughout the match and orange bars indicating the corresponding MPE energy expenditure. Notably, MPE contribute to approximately one-third of a soccer players' total energy consumption. 8

2.2 Distribution of game time across the observed GD states. 12

2.3 The flowchart illustrating the data collection and analysis workflow. The process involved the use of NASA-TLX questionnaires and GPS sensors. The research focused on two key objectives: examining the impact of cognitive load on physical performance, and exploring the effect of motivation profiles on the external load metric. 15

3.1 Muscle injuries distribution by type. Classification was performed using the BAMIC system. The x-axis denotes different injury types, as well as contusions denoted by C. The y-axis displays the occurrence count for each type of injury. 20

3.2 Soccer-specific 1-minute drill test [72]. The x-axis shows time in seconds, and the y-axis displays velocity in kilometers per hour (km/h). The green dotted line indicates the number of activated units (M_A), while the violet solid line represents the measured speed. The red ellipse highlights unrealistic activation rates, demonstrating challenges in numerically calculating the derivative, $\frac{dM_A}{dt} = \frac{dM_{AD}}{dt}$, as proposed by SNS. 23

3.3 Visualization of the intended energy expenditure (M_{AD}) and actual energy spent (E_{spent}) during the 80-meter sprint test. Blue dotted lines represent the peak energy attained during the first sprint, used as a benchmark for the subsequent sprints. X-axis denotes time in seconds, and Y-axis indicates energy in kilojoules. Violet solid lines illustrate kinetic energy calculated from the measured velocity and athlete mass. 24

List of Publications

- Skoki A., Ivić S., Ljubić S., Lerga J., Štajduhar I. Enhancing Biophysical Muscle Fatigue Model in the Dynamic Context of Soccer. Sports Engineering. 2024;.

- Skoki A., Gašparović B., Ivić S., Lerga J., Štajduhar I. Building Individual Player Performance Profiles According to Pre-Game Expectations and Goal Difference in Soccer. Sensors. 2024; 24(5):1700. https://doi.org/10.3390/s24051700

- Skoki A., Anić P., Ljubić S., Naglić A., Lerga J., Štajduhar I. Exploring the Impact of the Perceived Cognitive Load on the Physical Performance in Soccer. Elektrotehniški vestnik. 2024; 91(3):108-116.

- Sala R., Maffei A,, Pirola F., Enoksson F., Ljubić S., Skoki A., Zammit J. P., Bonello A., Podržaj P., Žužek T., Priarone P. C., Antonelli D., Pezzotta G. Blended learning in the engineering field: A systematic literature review. Computer Applications in Engineering Education. 2024; 32(3):e22712. doi:10.1002/cae.22712

- Skoki A., Napravnik M., Polonijo M., Štajduhar I., Lerga J. Revolutionizing Soccer Injury Management: Predicting Muscle Injury Recovery Time Using ML. Applied Sciences. 2023; 13(10):6222. https://doi.org/10.3390/app13106222

- Bonello A., Francalanza E., Zammit J. P., Pirola F., Pezzotta G., Sala R., Skoki A., Ljubić S., Podržaj P., Antonelli D., Minetola P., Enoksson F., Maffei A. Beyond the pandemic: How has Covid-19 shaped the capability to adopt an Agile Blended Learning in HEI? 5th International Conference on Higher Education Learning Methodologies and Technologies Online. 2023; pp. 29-30.

- Ljubić S., Skoki A., Hržić F., Salkanović A. Empowering Computer Engineering Education: Leveraging Cloud-Based Programming Platforms and Online Assessment Tools. 5th International Conference on Higher Education Learning Methodologies and Technologies Online. 2023; pp. 43-45.

- Sala R., Pirola F., Pezzotta G., Enoksson F., Ljubić S., Skoki A., Francalanza E., Zammit J. P., Bonello A., Podržaj P., Priarone P., Antonelli D., Minetola P., Maffei A., Boffa

E. Examining the implementation of Blended Learning in the Engineering field. 5th International Conference on Higher Education Learning Methodologies and Technologies Online. 2023; pp. 83-84.

- Skoki A., Rossi A., Cintia P., Pappalardo L., Štajduhar I. Extended Energy-Expenditure Model in Soccer: Evaluating Player Performance in the Context of the Game. Sensors. 2022; 22(24):9842. https://doi.org/10.3390/s22249842
- Skoki A., Lerga J. Štajduhar I., ML-Based Approach for NFL Defensive Pass Interference Prediction Using GPS Tracking Data. 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO). 2021; pp. 1038-1043. doi: 10.23919/MIPRO52101.2021.9596877
- Skoki A., Ljubić S., Lerga J., Štajduhar I. Automatic music transcription for traditional woodwind instruments sopele. Pattern Recognition Letters. 2019; 128:340-347. https://doi.org/10.1016/j.patrec.2019.09.024

Curriculum Vitae

Arian Skoki was born in 1995 in Rijeka, Croatia where he completed his elementary and high school education. In 2017, he earned a Bachelor's degree in Computer Engineering from the Faculty of Engineering at the University of Rijeka, followed by a Master's degree in 2019. His master's thesis, titled "Automatic Music Transcription for Traditional Woodwind Instruments Sopele," combined his passion for music with his technical expertise and was later published in an academic journal.

After completing his studies, he moved to Graz, Austria, to work as a Full Stack Developer. During his time in Graz, he gained valuable insights into the industry's challenges and the importance of community-driven initiatives like ML Graz, which foster both personal development and regional growth. In 2020 an opportunity came to combine computer engineering with his other passion - Sports. He became a Teaching Assistant at the Faculty of Engineering, University of Rijeka, where he also pursued a Ph.D. His research primarily focused on data analysis and machine learning in the field of soccer.

He was also teaching courses for Undergraduate studies: Embedded Systems, Computer Architecture, and Operating Systems. Apart from teaching roles, he was a leader and mentor of the Riteh Web Team (RWT) whose goal is to help students gain knowledge and experience to ease the transfer and bridge the gap between academia and industry. He was involved in mentoring students at the Lumen Development competition (2022) and Stem Games Technology arena (2022, 2023), the head of the organization for the "Ri-Hack" hackathon (2022, 2023), and conference "Ri-Comp" (2024), a member of the organization board for Summer School on Image Processing SSIP 2021 (Rijeka, Croatia) and SusTrainable Summer School 2022 (Rijeka, Croatia). In 2022, he joined the Erasmus+ project BLISS, which tackles the challenges of blended learning by integrating the best practices from both online and onsite teaching environments. This initiative aimed to enhance the quality and efficiency of education, using innovative approaches to combine the strengths of digital and traditional classrooms. Through this project, he had the opportunity to collaborate with experts from various countries, fostering valuable research connections and forming lasting friendships. That same year, he was awarded a scholarship from the SoBigData++ Consortium, allowing him to spend six weeks at the University of Pisa. This visit facilitated the exchange of ideas and collaboration with fellow researchers, further enriching his academic and professional development. He is also an active speaker at conferences such as DSC Adria, DSC Europe, ECML-PKDD, MLSA, and .debug, where he shares his passion for applying data analysis to sports.

Part II

Included Publications