Journal of Sports Sciences

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rjsp20

Association between match-related physical activity profiles and playing positions in different tasks: A data driven approach

Guglielmo Pillitteri, Alessio Rossi, Paolo Cintia, Athos Trecroci, Marco Petrucci & Giuseppe Battaglia

JOURNAL OF SPORTS SCIENCES
https://doi.org/10.1080/02640414.2024.2338026

Routledge
Taylor & Francis Group

SPORTS PERFORMANCE

Association between match-related physical activity profiles and playing positions in different tasks: A data driven approach

Guglielmo Pillitteri[a], Alessio Rossi[b,c,d], Paolo Cintia[c], Athos Trecroci[e], Marco Petrucci[f] and Giuseppe Battaglia[a]

[a]Sport and Exercise Research Unit, Department of Psychology, Educational Science and Human Movement, University of Palermo, Palermo, Italy; [b]Department of Research and Development, Feel Good Plus S.R.L. - MyPowerSet, Rome, Italy; [c]Department of Computer Science, University of Pisa, Pisa, Italy; [d]National Research Council (CNR), Institute of Information Science and Technologies (ISTI), Pisa, Italy; [e]Department of Biomedical Science for Health, University of Milan, Milan, Italy; [f]Department of Performance, Palermo Football Club, Palermo, Italy

ABSTRACT
Assessing the intensity characteristics of specific soccer drills (matches, small-side game, and match-based exercises) could help practitioners to plan training sessions by providing the optimal stimulus for every player. In this paper, we propose a data analytics framework to assess the neuromuscular or metabolic characteristics of a soccer-specific exercise in relation with the expected match intensity. GPS data describing the physical tasks' external intensity during an entire season of twenty-eight semi-professional soccer players competing at the fourth Italian division were used in this study. A supervised machine-learning approach was tested in order to detect difference in playing positions in different sport-specific drills. Moreover, a non-supervised machine-learning model was used to profile the match neuromuscular and metabolic characteristics. Players' playing positions during matches and match-based exercises are characterised by specific metabolic and neuromuscular characteristics related to tactical demands, while in the small-side game these differences are not detected. Additionally, our framework permits to evaluate if the match performance request is mirrored during training drills. Practitioners could evaluate the type of stimulus performed by a player in a specific training drill in order to assess if they reflect the matches characteristics of their specific playing position.

ARTICLE HISTORY
Received 29 September 2022
Accepted 25 March 2024

KEYWORDS
Small-side games; match-based exercises; neuromuscular and metabolic characteristics; GPS; training workload

Introduction

During a soccer match, players are required to address different both technical activities such as shot, passing, header, tackles and running at different speed intensity in the form of high-speed running, accelerations, and decelerations tasks (Bradley et al., 2010; Buchheit & Laursen, 2013) eliciting both metabolic and neuromuscular demands. In soccer, physical performance refers to intermittent cyclic and acyclic activities characterized by aerobic and anaerobic demands (Bloomfield, Polman, & O'Donoghue, 2007). As a matter of fact, 150–250 different actions and 1100 changes of direction were performed by players during a match (Krustrup et al., 2005). Professional players cover a total distance of between 10 and 13 km during a match (Stølen et al., 2005) including low intensities, while 22–24% is spent at intensities above 15 km/h, 8–9% above 20 km/h, and 2–3% above 25 km/h. Moreover, players can perform between 600 and 650 accelerations during a match highlighting their relevant contribution on physical demand. These demands were found to be related to playing positions (DiSalvo et al., 2007; Rampinini et al., 2007; Riboli et al., 2022). As a matter of fact, previous studies have shown differences in external load (EL, i.e., distance, speed and accelerations) (Abbott et al., 2018) and internal load (IL, i.e., through physiological factors assessment including heart rate and derived indices and/or rate of perceived effort scale) (Aslan et al., 2012) by playing position during both training and matches (Ademović, 2016; Orendurff et al., 2010). In particular, a previous study has shown that both central and wide midfielders spent most of the time in moderate- and high-intensity activities, also covering more total distance than other roles during a game. Moreover, they spent less time in low intensity showing shorter recovery bouts compared with other playing positions (Ademović, 2016; Brito et al., 2018; Orendurff et al., 2010). Differently, defenders spent significantly less time sprinting and running compared with other roles (Bloomfield, Polman, & O'Donoghue, 2007). Additionally, central backs showed the longest recoveries between consecutive high-intensity efforts (J. Ade et al., 2016) and the shortest time in low intensity activities (Ademović, 2016). It is worth noting that players in wide positions (i.e., wide defender and wide attacker) accelerated significantly more than central ones (Abbott et al., 2018). Notably, each position requires specific physical performance that could change in accordance with the tactics imposed by the coach along with other contextual factors such as match scoring and difficulty, i.e., level of opponents (Bradley et al., 2011).

In order to reflect the physical, technical, tactical, and cognitive demands of a match, Small-Sided Games (SSG) and Match-Based Exercises (MBE) may represent valid soccer-specific training solutions. If opportunely employed by practitioners, these particular drills are capable of driving both metabolic and neuromuscular physical demands differently

CONTACT Alessio Rossi alessio.rossi2@gmail.com Department of Research and Development, Feel Good Plus S.R.L. - MyPowerSet, Rome, Italy
Supplemental data for this article can be accessed online https://doi.org/10.1080/02640414.2024.2338026

(Gaudino et al., 2014). SSGs are among the most common drills, they are performed in a small area of the pitch with a reduced number of players and specific rules that allow modulating the training intensity (Clemente et al., 2014). MBE reflects soccer tasks carried out in larger pitches, with a greater number of players compared to SSG and games rules defined in accordance with Official Matches (OM). From a tactical perspective, MBE are more similar to match requests than SSG. It is worth noting that SSG does not completely reproduce the actual game-related physical demands (Gómez-Carmona et al., 2018). Rather, it can be considered a high-intensity task that stresses the players' neuromuscular components (Gómez-Carmona et al., 2018).

It is known that both EL and IL can be affected by manipulating game and scoring rules, the number of players per team, and the pitch size (Clemente et al., 2019). Regarding pitch size, ~300 m^2 per player is the average match density experienced by players during regular matches (Castellano et al., 2015), while SSGs have been performed in more dense conditions (e.g., ~100 m^2 or ~200 m^2 per player) (J. D. Ade et al., 2014). It has also been previously reported that, in a smaller pitch (around 100 m^2, players' performance is characterised by multiple accelerations, decelerations, and changes of direction involving high neuromuscular demands). Differently, in a large field, such as in the Match Based Exercises, players address more high-speed running tasks (Castellano et al., 2015), eliciting more metabolic responses such as metabolic power and derived metabolic measures. This is probably due to the higher speed reached by players in large fields compared to smaller ones observed in Small-Sided Games (Gaudino et al., 2014). Hence, increasing the field dimension could be a proxy for increasing also acute physiological demands (i.e., heart rate, blood lactate, and RPE) (Hill-Haas et al., 2009). In this regard, coaches can consider "intense" tasks based on the type of stimulus (i.e., metabolic or neuromuscular) undergone by a player during a specific drill.

Although the literature has reported a lot of information on SSG and MBE, understanding the real impact of these specific drills on soccer performance could be a challenge for practitioners to schedule training weeks. Defining the intensity of specific soccer drills could help practitioners planning training sessions by providing the optimal stimulus for the players. The framework we propose in this study would aim to make it possible. In particular, it can assess the physical impact of different drills (matches, SSGs, and MBEs) for different players' playing positions. Specifically, through this approach it will be possible to define the intensity of a soccer exercise by indicating whether a stimulus is predominantly neuromuscular or metabolic also by playing position. Therefore, the first aim of this study is to discriminate roles in soccer through EL profiles considering different types of soccer drills using a supervised Machine Learning approach. Additionally, the second aim is to create physical activity profiles based on EL through a data driven approach: official matches' data become a comparison key to evaluate the intensity of physical tasks executed during a training session. Given the amount and the complexity of the available data, the use of Machine Learning (ML) (Rossi et al., 2021) approaches could help to outline the impact of these exercises on soccer's performance based on EL and tasks' density data (m^2 per player).

Methods

Subjects

Twenty-eight semi-professional soccer players (age: 25.21 ± 6.18 years, height: 183.42 ± 6.41 cm, weight: 75.24 ± 7.02 kg) competing at the fourth Italian division (Serie D) were recruited for this study during the 2019–2020 season. Due to the different physical demands, the goalkeeper was excluded from this study (DiSalvo et al., 2006; Pillitteri et al., 2023). In this paper, players were separated into six playing positions ($n = 7$ central back, $n = 7$ external strikers, $n = 2$ fullback, $n = 5$ midfielder, $n = 4$ striker, and $n = 3$ wide midfielder) in accordance with the zone of the pitch they used to play. A total of 996 training sessions (35.57 ± 21.79 per player) were recorded throughout the soccer season. To reduce inter-training variability between weeks induced by short microcycle, data from regular in-season training weeks (i.e., 6 days between matches) were used. A total of 33 weeks composed of 5 training days and 1 day off were used in this study. All the participants signed an informed consent before taking part in the experiment. The study, which complies with the principles of the Declaration of Helsinki, was approved by the Bioethics Committee of the University of Palermo (n. 68/2021).

Soccer-specific drills

Three different drills were evaluated in this study:

- Matches (official and friendly matches; $n = 384$): we considered a match density of about 300 m^2.
- Small-Side Game (SSG; $n = 296$): SSGs were performed in a small area of the pitch with a reduced number of players compared to regular matches and carried out with specific rules (e.g., number of touches of the ball, goal crossing the goal line, the use of three soccer goals). In this study, we included different types of SSG based on m^2/player (around 100 m^2/player), as shown in Table S3. SSGs were grouped by a data driven approach (supplementary material S1) as follows: Small Density with High Neuromuscular and Low Metabolic indicators (SD-HN-LM-ssg); Medium Density with Low Neuromuscular and Metabolic indicators (MD-LN-LM-ssg); Large Density, High Neuromuscular and High Metabolic indicators (LD-HN-HM-ssg); Very Large Density with Medium Neuromuscular and Metabolic indicators (VLD-MN-MM-ssg).
- Match-based exercises (MBE; $n = 316$): MBEs represent soccer specific tasks carried out in a larger area of the pitch and a relatively larger number of players compared to SSGs. Specific rules related to technical-tactical behaviours (similarly to matches' demands) characterise these tasks. The presence of both goalkeepers and the directionality of the game (i.e., the team in possession must score a goal in the opposing team's goal) makes these drills similar to friendly matches with the difference that they are played in smaller spaces and there are rules

specific tactics compliant with the tactical needs of the coach. Moreover, the playing positions are defined for each player.

- In this study, we included different types of MBE based on m^2/player (around 200 m^2/player) as shown in Table S4. MBEs were grouped by a data driven approach (supplementary material S1) as follows: Medium Density with Low Neuromuscular and Metabolic indicators (MD-LN-LI-mbe); Large Density with Medium Neuromuscular and Metabolic indicators (LD-MN-MI-mbe); Small Density with High Neuromuscular and Metabolic indicators (SD-HN-HI-mbe).

Data recording

Both matches and soccer-specific drills were played on a third-generation artificial pitch or natural grass. All players performed a 15-min warm-up before performing MBE and SSG and a typical 25-min pre-match warm-up before matches.

Data was collected through a Global Position System (Varley et al., 2012) (GPS unit, Qstarz BT-Q1000EX, 10 Hz) (Carlson et al., 2015) positioned on the upper back inserted in a special vest. The GPS device was started 15 min prior to the assessment to make available the acquisition of satellite signals. The average number of satellites in view was 12, the average horizontal dilution of precision (HDOP) was 0.8 (Duncan et al., 2013). All data was acquired through a dedicated software (LaGalaColli V: 8.6.4.3) during the entire 2019–2020 soccer season. EL indicators (normalized per minute) and recorded as global time (considering when the ball is both in and out of play and excluding recovery between set), obtained from GPS devices and considered in this study were both metabolic and neuromuscular, and they are detailed in Table 1.

Playing position prediction

Four machine-learning classifiers were trained to predict the players' playing positions based on EL parameters, respectively, in all the soccer-specific drills. The algorithms tested in this paper are Decision Tree Classifier (DTC), Random Forest Classifier (RFC), Extreme Gradient Boost Classifier (XGB), and Multinomial Logistic Regression (MLR). Moreover, a baseline classifier (Dummy) based on stratified approach (distribution of playing positions) was assessed in order to validate the prediction ability of the machine learning models. The models were trained in the 70% of the dataset, while the remaining 30% of the dataset was used as a test set. The metrics used to evaluate the models goodness are described in Supplementary material S2.

Match' intensity profiles

A data driven approach was used to profile players in accordance with intensity demands during Matches. In particular, the data was initially normalized between 0 and 1, and then a feature dimension reduction was performed by using Principal Component Analysis (PCA). The elbow approach based on the explained variance was used to detect the best number of components for PCA.

A non-supervised machine learning model (k-means) based on the components derived from PCA analysis was fitted to partition n observations into k clusters in which each observation belongs to the cluster with the nearest cluster centroid. A silhouette analysis based on the k-means model was conducted in order to detect the best number of clusters into each drill. Finally, we assessed how the physical activity profiles detected in matches were distributed in SSGs and MBEs.

Results

Playing positions prediction

Table 2 shows the predictive performance in each drill. The players' positions in OM are more accurately discriminated by RFC (57%) compared with the other models. The very low accuracy of the baseline model (16%) indicates that RFC is a valid model able to distinguish between players' position in matches by assessing intensity parameters. In particular, %W > 35W, n° actions int/min, and metabolic power (Watt) are the three most important features used by RFC in the decision-making process showing an importance of 21.62%, 13.24%, and 13.08%, respectively.

In SSG, the playing positions of players are difficult to detect (RFC accuracy = 36%). The low ability to discriminate between players' playing positions during SSG and the low increment in predictive ability of RFC compared to baseline model (~12%) indicate that this kind of drill could not accurately represent the performance model of each role in the matches.

Table 1. EL Indicators descriptions.

Feature	Description
Watt	Metabolic Power (W · kg – 1) was calculated by multiplying EC (in J · kg – 1 · m – 1) by running speed (v; in m · s – 1) at any given moment (i.e., every 0.2 s): P met = EC · v. In order to assess metabolic power, considering the energy expenditure and derives, the equation developed by (DiPrampero et al., 2005). established on previous studies by (Minetti et al., 2002). and (Osgnach et al., 2010). was adopted.
%W > 35W	Percentage of metabolic power with intensity > 35 W.
n° actions int/min	Number of intense actions per minute. The threshold considered to establish the intense actions was 20 W · kg – 1 which represents the metabolic power when running at a constant speed of approximately 14.4 km · h – 1 (high speed) on grass (Osgnach et al., 2010).
dist/min	Metres covered in a minute.
PrT/min	Passive Recovery time per minute represents the sum of the seconds in a minute (for each minute) in which the player moves at a metabolic power between 0 and 5 W (i.e., 0–8 Km/h).
%int acc	This indicator represents the percentage of the number of accelerations >2 m/s2 out of the total number of accelerations.
%int dec	In an opposite way to the previous one, this parameter indicates the percentage of the number of decelerations <-2 m/s2 out of the total number of decelerations.
CdD/min >30	The number of changes of directions higher than 30° per minute.

Table 2. Predictive performance (weighted average of all the classes) of the machine learning algorithms and the baseline. The best model for each task was highlighted in green, while baseline (dummy) predictive performance was highlighted in grey.

Task	Model	Precision	Recall	F1-score	Accuracy
Matches	DTC	0.45	0.47	0.45	0.47
	RFC	0.56	0.57	0.56	0.57
	XGB	0.57	0.56	0.55	0.56
	MLR	0.47	0.51	0.47	0.51
	Dummy	0.16	0.16	0.16	0.16
SSG	DTC	0.29	0.29	0.29	0.29
	RFC	0.35	0.36	0.34	0.36
	XGB	0.32	0.35	0.33	0.35
	MLR	0.29	0.34	0.30	0.34
	Dummy	0.22	0.24	0.23	0.24
MBE	DTC	0.29	0.27	0.28	0.27
	RFC	0.35	0.36	0.35	0.36
	XGB	0.28	0.26	0.27	0.26
	MLR	0.38	0.40	0.38	0.40
	Dummy	0.17	0.17	0.17	0.17

During MBE, RFC is able to predict 40% of the total players' playing positions. It is a valid model to discriminate the roles of the players due to the higher accuracy compared with baseline (17%). Similar to the results obtained in Matches, the three most important features used by RFC to predict the players' playing positions are %W > 35W (21.30%), n° actions int/min (12.90%), and metabolic power (11.67%).

For more details about the prediction goodness of the models, see Figure 3.

Matches' intensity profiles

PCA analysis indicates that 3 components explain 97.63% of the data variance during Matches. By using these 3 components, it is possible to distinguish among 5 players' profiles (silhouette score = 0.34). We label them as very low intensity (VLI), low intensity (LI), moderate intensity (MI), high intensity (HI), and very high intensity (VHI).

Table 3 shows the centroids of the 5 profiles detected by the k-means analysis. Figure 1 reports the distribution of playing positions in each matches' intensity profile. During 62% and

Table 3. Matches' intensity profiles obtained by a data driven approach. The data were reported as mean (standard deviation). VLI, LI, MI, HI, and VHI correspond to very low intensity, low intensity, moderate intensity, high intensity, and very high intensity, respectively.

	VLI	LI	MI	HI	VHI
Watt	8.79 (0.49)	9.97 (0.40)	10.92 (0.35)	11.05 (0.37)	12.03 (0.44)
dist/min	92.64 (4.70)	104.31 (4.11)	114.88 (3.01)	114.86 (3.40)	126.07 (3.50)
%int acc	7.60 (1.30)	8.85 (1.32)	9.24 (1.28)	10.22 (1.22)	10.27 (1.61)
%int dec	8.27 (1.29)	9.52 (1.31)	10.16 (1.26)	10.02 (1.36)	10.63 (1.38)
CdD/min >30	14.39 (1.13)	15.04 (1.38)	15.58 (1.44)	14.91 (1.31)	15.26 (1.34)
%W > 35W	21.51 (4.14)	21.94 (5.72)	17.58 (3.75)	27.31 (2.82)	19.35 (2.95)
n° actions int/min	2.48 (0.45)	2.99 (0.29)	3.57 (0.35)	3.18 (0.39)	3.94 (0.30)
PrT/min	22.21 (1.91)	18.44 (2.73)	15.00 (2.41)	14.00 (2.12)	11.12 (2.19)

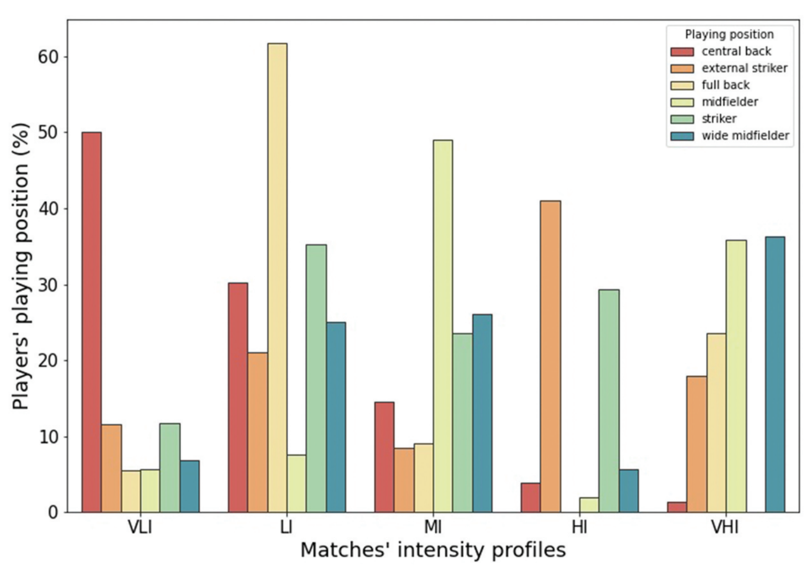

Figure 1. Percentage of playing position in each players' matches' intensity profiles. VLI, LI, MI, HI, and VHI refer to very low intensity, low intensity, moderate intensity, high intensity, and very high intensity, respectively.

50% of the time, full backs and central backs performed matches in VLI and LI profiles, respectively. Differently, external strikers usually show HI profiles (43%), while strikers execute different performances in each match showing 36%, 24%, and 29% of the time LI, MI, and HI profiles. Additionally, midfielders played with MI and VHI physical activity profiles for 50% and 37% of the time, respectively. Finally, 24%, 26% and 38% of the time, wide midfielders played with LI, MI, and VHI profiles, respectively.

Figure 2 and 3 shows the difference in occurrence distribution of matches' intensity profiles detected in OM for SSG and MBE in each player's playing position, respectively. In

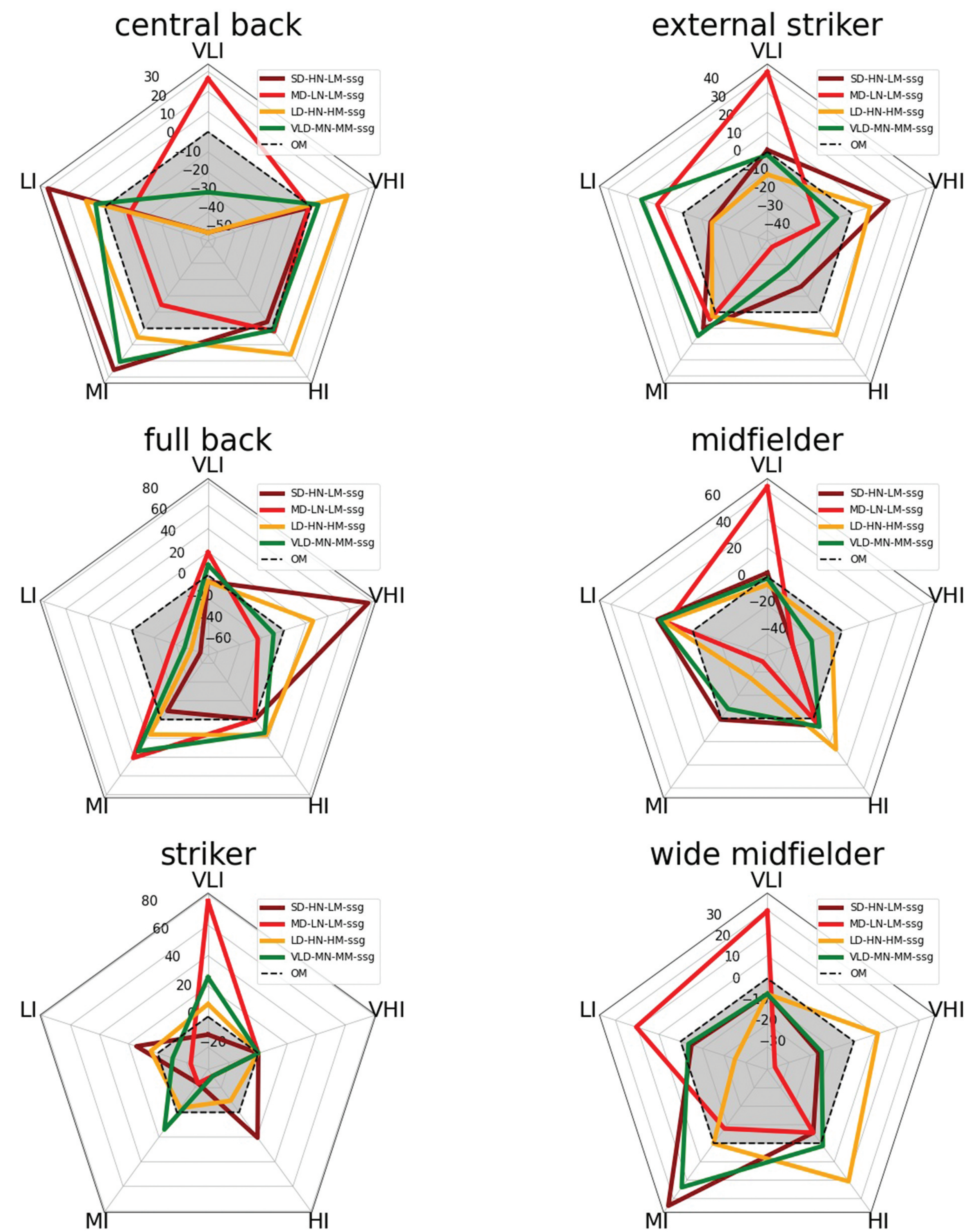

Figure 2. Percentage difference of matches' intensity profiles occurrence between each SSG group and OM for each playing position.

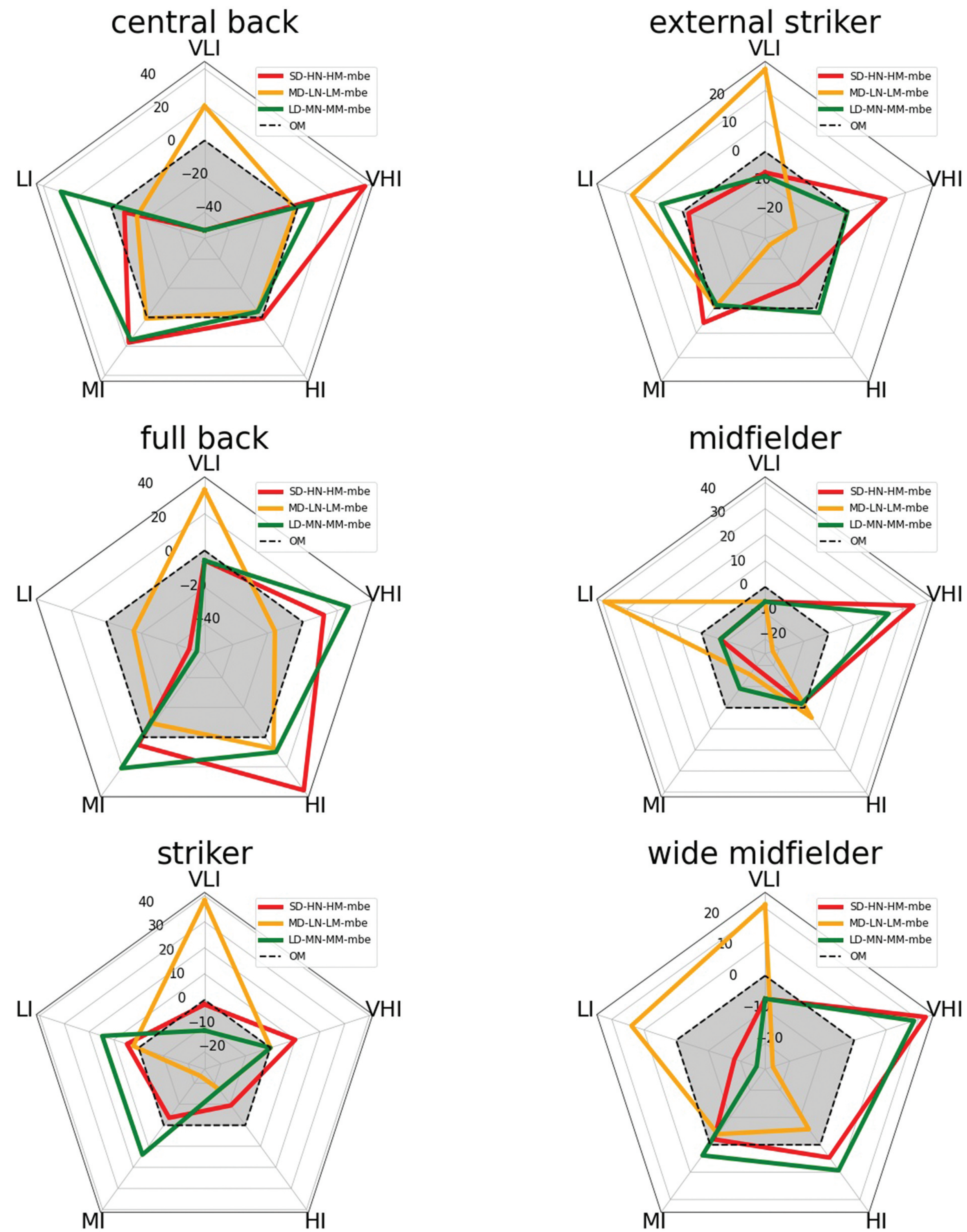

Figure 3. Percentage difference of matches' intensity profiles occurrence between each MBE group and OM for each playing position.

accordance with the different typologies of SSGs and MBEs and with the players' playing positions, the distribution of matches' intensity profiles changes. Moreover, different patterns of intensity profiles were detected between sport specific drills (i.e., SSG and MBE) and OM. Considering SSG exercises (Figure 2), central backs performed most of the matches at moderate-low intensity reflecting a similar trend in all the SSG typologies. LD-HN-HM-ssg induces the highest intensity in the players competing in this playing position. Moreover, external strikers often performed high intensity matches (HI and VHI) that could be simulated only by SD-HN-LM-ssg and LD-HN-HM-ssg. Additionally, full backs resulted in LI, but in all SSG types performed more MI. 43% and 38% of times midfielders performed matches in MI and VHI profiles. SD-HN-LM-ssg and VLD-MN-MM-ssg show a similar MI pattern, while LD-HN-HM-ssg produce more intense activities. Strikers and midfielders performed all of the physical activity profiles in similar percentages, but SSG typologies show specific patterns that do not reflect the OM ones.

Figure 3 reports the difference in occurrence distribution of matches' intensity profiles detected in OM for MBE in each player's playing position. During SD-HN-HM-mbe the players performed the highest percentage of VHI compared to other MBE types and OM. Differently, MD-LN-LM-mbe results in higher percentage of VLI and LI activities and lower HI and VHI ones compared to other MBE types and OM in external strikers. Similar to MD-LN-LM-mbe, full backs presented the highest percentage of LI in OM, while others MBE types showed highest intensity. Similar patterns were detected between OM and MBE types in strikers and wide midfielders.

Discussion

The first main result of this work is that the players' playing positions during matches and MBE are characterised by specific metabolic and neuromuscular stimuli, while during SSG players show similar physical stimuli even if they play in different playing positions (Table 2). Moreover, the framework of data analytics provided in this paper permits to profile the players' metabolic and neuromuscular performance during matches with the aim of understanding if the physical demands of a player requested during matches are reflected in the training drills.

In accordance with matches' tactical demands, the usual playing position of a player could vary during a match or between matches and consequently the role-related physical demands (Carling et al., 2016; Rampinini et al., 2007) change. Hence, for these reasons, the model is able to moderately discriminate the players' roles in all the matches and MBE examples. As a matter of fact, during matches and MBE, the wide midfielders were often confused with external strikers probably because of similar tactical demands among these two roles. Moreover, as full backs and wide midfielders played in an external pitch position with similar tactical and physical demands, our model showed some predictiction errors related to these specific player's playing positions. Finally, midfielders are almost always exchanged with wide midfielders. These results indicate that different physical demands are required by specific roles in Matches and MBE. Differently, the performance model for each role in SSG seems completely different compared to OM. The reason why the players' playing positions are not distinguishable during SSG is that the maintenance of a specific role in these kinds of exercises is not mainly required, despite being rarely assigned, but practitioners used these drills to induce technical and physical stimuli. Worth noting, during the training microcycle prescription process, different types of tasks across the training week are required to elicit the targeted physical stimuli. For example, during the MD-4 the use of SSG expose the players to "special" strength such as acceleration and deceleration, while the players' metabolic response could be limited. Conversely, during the MD-3 players are often exposed larger playing spaces (e.g., MBE) to reach high-speed running distance, sprinting, peak speed, and higher metabolic power (i.e., higher intensity and metabolic stimuli).

This finding corroborates the results of a previous paper in which Riboli et al. (Riboli et al., 2020) found that training tasks with goalkeepers better simulated the match intensity. This is because the use of the goalkeepers gives a direction to the game, resulting in a tactical strategy involvement compared to other drills. Moreover, SSG without goalkeepers usually shows higher intensity compared to goalkeeper-based tasks (Riboli et al., 2020). Additionally, based on the low predictive performance of RFC detected in this study, we could speculate that the metabolic and neuromuscular performances are poorly related to playing positions in SSG. The pitch size, the players' density on the field and the exercise rules induces different metabolic and neuromuscular responses in both SSG and MBE drills. Indeed, some authors reported that specific strength components (accelerations, decelerations, and changes of direction) are emphasised in small spaces, while high-speed running and metabolic stimuli are greater in larger pitch sizes (Casamichana et al., 2012; Castellano et al., 2015; Hill-Haas et al., 2009). As a matter of fact, Gaudino et al. reported an increase in acceleration/deceleration requirements and total number of changes in velocity when the number of players and pitch dimensions decreased in SSG (Gaudino et al., 2014).

Profiling players' performance in accordance with the match's metabolic and neuromuscular characteristics would allow coaching staff to fast and easily assess the stimuli sustained by players. In particular, evaluating the match physical performance could be useful to schedule the training program during the weekly microcycle evaluating if the training target and intensity were reached for each player's roles. Actually, as shown in Figure 1, each player's playing position results in a prevalence of specific matches' intensity characteristics showing that the players' roles are characterised by a specific prestative model. However, based on match characteristics, players could show a different matches' intensity profile compared to the common one detected for their specific playing position. For example, full backs usually show a prevalence of the LI profile (62% of the time), but 24% of the time they perform in the VHI profile. It is probably due to the different intensity and defensive demands of the match.

Thanks to match profiling, it is possible to assess whether the model detected during a match in a specific player or playing position is mirrored during training drills such as SSG and MBE. Moreover, to train specific metabolic and neuromuscular characteristics, it might be possible for practitioners to assess the type of workload sustained by a player in a specific playing position during such drills. Hence, the choice of a specific SSG and MBE could be made in accordance with the training purpose and intensity (McArdle et al., 2010). As shown in Figures 3 and 4, similar distribution of physical profiles was detected in MBEs compared with OMs, while SSGs showed a different distribution. The reason why the players' playing positions show different physical profiles distribution during SSGs compared to the matches one, while similar one during MBEs is related to the different nature of the two tasks. As a matter of fact, MBEs are match oriented due to the fact that it is characterised by similar players' playing positions and rules,

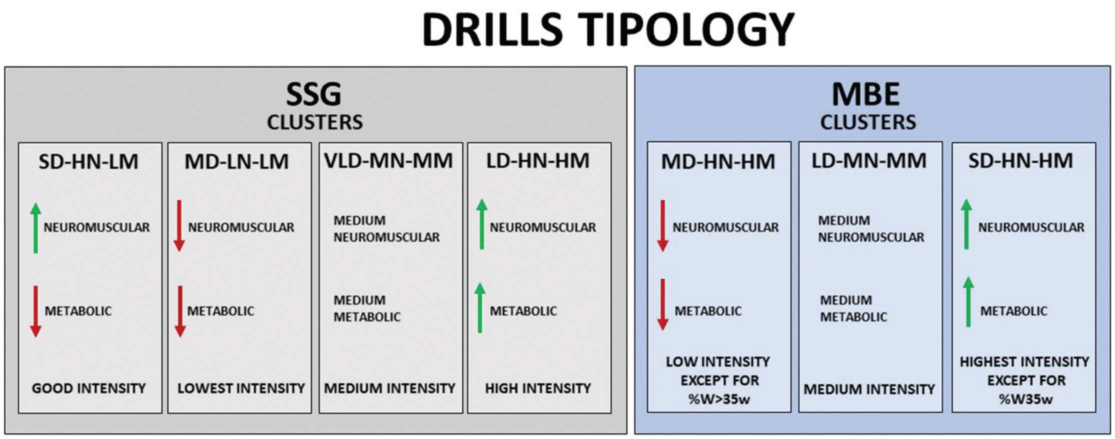

Figure 4. Drills cluster framework based on the ML analysis.

while SSGs are more focused on free play without fixing specific roles but with specific rules not often match-related (Riboli et al., 2020, 2022).

The study's primary limitation lies in its applicability solely to the specific soccer team under analysis. Variations in player characteristics, team playing styles, and training regimens necessitate a tailored adaptation of our framework for each team. However, this limitation also underscores the advantage of our work's personalized nature. Indeed, our paper introduces an innovative method for evaluating players' physical performance during both matches and training sessions, eschewing the provision of generalized insights into role behaviours. Thus, the principal strength of our approach lies in its utilization of machine learning, which facilitates the detection of multidimensional patterns within extensive datasets, enabling the assessment of training objectives and intensity characteristics.

Practical applications

The analysis conducted in this study provides a deeper understanding regarding the roles-specific model. As a matter of fact, in order to describe the players' performance during matches and comparing it with training exercises in different playing positions, specific matches players' profiles could be obtained from their EL data. Actually, the framework used in this work allows practitioners to better recognize the workloads performed by players during training tasks. In detail, this approach permits to detect relevant information from training exercises regarding both target task (i.e., neuromuscular or metabolic) and the task intensity based on EL indicators, the players' density on the pitch, and the type of tasks. Therefore, it is possible to understand if a specific task could be more intense compared to official matches, considering each player's playing position.

In detail, as shown in Figure 4, each cluster provides valuable information on the kind of stimuli elicited (i.e., metabolic or neuromuscular) and the physical intensity for each playing position. As regards SSG, clustered as SD-HN-LM showed the smallest density and consequently eliciting high neuromuscular, low metabolic stimuli and good intensity. Cluster MD-LN-LM presented medium density but the lowest intensity as shown for both neuromuscular and metabolic indicators. While cluster VLD-MN-MM revealed the highest density and medium intensity. The highest intensity was found in Cluster LD-HN-HM. Concerning MBEs clusters, MD-LN-LM showed medium density and lowest intensity (due to the highest Tr pass/min) considering all indicators except for %W > 35w that presented the highest value. Cluster LD-MN-MM presented the highest density and medium-high metabolic and neuromuscular indicators, while Cluster SD-HI-HN showed the smallest density, highest intensity considering all indicators except %W > 35w.

As regards the difference between playing position, Table 4 shows an overview about the drills clustered (i.e., SSG and MBE) providing the physical intensity (low, medium, high and very high) for each role based on the type of drills selected.

In this way, practitioners could select with more appropriateness the type of task during the scheduling microcycle process based on the suggestions provided by the proposed machine learning model.

Table 4. Overview of the difference in physical intensity between playing positions based on the different drills clustered (SSG and MBE). SSG: (SD-HN-LM) – (MD-LN-LM) - (LD-HM-HN) – (VLD-MN-MM); MBE: (SD-HM-HN) – (MD-LN-LM) – (LD-MN-MM). Moreover, VLI, LI, MI, HI and VHI refer to very-low-intensity, low-intensity, medium-intensity, high-intensity and very-high-intensity, respectively.

	DRILLS TIPOLOGY						
	SSG				MBE		
PLAYING POSITION	SD-HN-LM	MD-LN-LM	LD-HN-HM	VLD-MN-MM	SD-HN-HM	MD-LN-LM	LD-MN-MM
CB	LI	VLI	HI/VHI	MI	VHI	VLI	LI
ES	VHI	VLI	HI	LI	VHI	LI/VLI	LI
FB	VHI	MI	VHI	MI	HI	VLI	MI/VHI
MD	LI	VLI	HI	LI	VHI	LI/VLI	VHI
S	HI	VLI	VLI	VLI	VHI	VLI	LI/MI
WM	MI	LI/VLI	HI/VHI	MI	HI/VHI	LI/VLI	HI/VHI

Conclusion

Playing position shows specific intensity profiles based on training task targets and on both neuromuscular and metabolic components during both training drills and matches. The framework of data analytics based on the machine learning approach proposed in this paper permits the evaluation of the intensity of soccer tasks in order to easily and quickly evaluate the physical performance of a player. In this way, practitioners could evaluate the type of stimulus performed by a player in a specific training drill in order to assess if they reflect the matches characteristics of their specific playing position. Moreover, practitioners could use this approach to schedule the training program in order to maximise the training effect while reducing the risk of injury by monitoring both intensity and the neuromuscular and metabolic components of each training task.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the European Community's H2020 Program under the funding scheme H2020-INFRAIA-2019-1 Research Infrastructures grant agreement [871042], www.sobigdata.eu, SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ORCID

Alessio Rossi http://orcid.org/0000-0002-6400-5914

References

Abbott, W., Brickley, G., & Smeeton, N. J. (2018). Physical demands of playing position within English Premier League academy soccer. *Journal of Human Sport & Exercise*, *13*(2). https://doi.org/10.14198/jhse.2018.132.04

Ade, J., Fitzpatrick, J., & Bradley, P. S. J. J. O. S. S. (2016). High-intensity efforts in elite soccer matches and associated movement patterns, technical skills and tactical actions. Information for position-specific training drills. *Journal of Sports Sciences*, *34*(24), 2205–2214. https://doi.org/10.1080/02640414.2016.1217343

Ade, J. D., Harley, J. A., & Bradley, P. S. (2014). Physiological response, time-motion characteristics, and reproducibility of various speed-endurance drills in elite youth soccer players: Small-sided games versus generic running. *International Journal of Sports Physiology and Performance*, *9*(3), 471–479. https://doi.org/10.1123/ijspp.2013-0390

Ademović, A. J. H. S. (2016). Differences in the quantity and intensity of playing in elite soccer players of different position in the game. *18*(1), 26–31. UDC: 796.332.058.

Aslan, A., et al. (2012). Metabolic demands of match performance in young soccer players. *11*(1), 170.

Bloomfield, J., Polman, R., & O'Donoghue, P. (2007). Physical demands of different positions in FA Premier League soccer. *Journal of Sports Science and Medicine*, *6*(1), 63.

Bloomfield, J., Polman, R., & O'Donoghue, P. (2007). Physical demands of different positions in FA Premier League soccer. *Journal of Sports Science & Medicine*, *6*(1), 63.

Bradley, P. S., Carling, C., Archer, D., Roberts, J., Dodds, A., DiMascio, M., Paul, D., Gomez Diaz, A., Peart, D., & Krustrup, P. (2011). The effect of playing formation on high-intensity running and technical profiles in English FA Premier League soccer matches. *Journal of Sports Sciences*, *29*(8), 821–830. https://doi.org/10.1080/02640414.2011.561868

Bradley, P. S., DiMascio, M., Peart, D., Olsen, P., & Sheldon, B. (2010). High-intensity activity profiles of elite soccer players at different performance levels. *The Journal of Strength & Conditioning Research*, *24*(9), 2343–2351. https://doi.org/10.1519/JSC.0b013e3181aeb1b3

Brito, Â., Roriz, P., Duarte, R., & Garganta, J. (2018). Match-running performance of young soccer players in different game formats. *International Journal of Performance Analysis in Sport*, *18*(3), 410–422. https://doi.org/10.1080/24748668.2018.1479924

Buchheit, M., & Laursen, P. B. (2013). High-intensity interval training, solutions to the programming puzzle. Part II: Anaerobic energy, neuromuscular load and practical applications. *Sports Medicine*, *43*(10), 927–954. https://doi.org/10.1007/s40279-013-0066-5

Carling, C., Bradley, P., McCall, A., & Dupont, G. (2016). Match-to-match variability in high-speed running activity in a professional soccer team. *Journal of Sports Sciences*, *34*(24), 2215–2223. https://doi.org/10.1080/02640414.2016.1176228

Carlson, J. A., Jankowska, M. M., Meseck, K., Godbole, S., Natarajan, L., Raab, F., Demchak, B., Patrick, K., & Kerr, J. (2015). Validity of PALMS GPS scoring of active and passive travel compared to SenseCam. *Medicine & Science in Sports and Exercise*, *47*(3), 662. https://doi.org/10.1249/MSS.0000000000000446

Casamichana, D., Castellano, J., & Castagna, C. (2012). Comparing the physical demands of friendly matches and small-sided games in semi-professional soccer players. *The Journal of Strength & Conditioning Research*, *26*(3), 837–843. https://doi.org/10.1519/JSC.0b013e31822a61cf

Castellano, J., Puente, A., Echeazarra, I., & Casamichana, D. (2015). Influence of the number of players and the relative pitch area per player on heart rate and physical demands in youth soccer. *The Journal of Strength & Conditioning Research*, *29*(6), 1683–1691. https://doi.org/10.1519/JSC.0000000000000788

Clemente, F. M., Lourenço Martins, F. M., & Mendes, R. S. (2014). Developing aerobic and anaerobic fitness using small-sided soccer games: Methodological proposals. *Strength & Conditioning Journal*, *36*(3), 76–87. https://doi.org/10.1519/SSC.0000000000000063

Clemente, F. M., Praça, G. M., Bredt, S. D. G. T., Linden, C. M. I. V. D., & Serra-Olivares, J. (2019). External load variations between medium- and large-sided soccer games: Ball possession games Vs Regular games with small goals. *Journal of Human Kinetics*, *70*(1), 191–198. https://doi.org/10.2478/hukin-2019-0031

DiPrampero, P., Fusi, S. J. S. R. A. N. E. A. J. E. B., Sepulcri, L., Morin, J. B., Belli, A., & Antonutto, G. (2005). Sepulcri L Morin JB Belli A Antonutto G. *The Journal of Experimental Biology*, *208*(14), 2809–2816. https://doi.org/10.1242/jeb.01700

DiSalvo, V., Baron, R., Tschan, H., Montero, F.Calderon, Bachl, N., & Pigozzi, F. (2006). Performance characteristics according to playing position in elite soccer. *International Journal of Sports Medicine*, *28*(3), 222–227. https://doi.org/10.1055/s-2006-924294

DiSalvo, V., Baron, R., Tschan, H., Montero, F.Calderon, Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *International Journal of Sports Medicine*, *28*(3), 222–227. https://doi.org/10.1055/s-2006-924294

Duncan, S., Stewart, T. I., Oliver, M., Mavoa, S., MacRae, D., Badland, H. M., & Duncan, M. J. (2013). Portable global positioning system receivers: Static validity and environmental conditions. *American Journal of Preventive Medicine*, *44*(2), e19–e29. https://doi.org/10.1016/j.amepre.2012.10.013

Gaudino, P., Alberti, G., & Iaia, F. M. J. H. M. S. (2014). Estimated metabolic and mechanical demands during different small-sided games in elite soccer players. *Human Movement Science*, *36*, 123–133. https://doi.org/10.1016/j.humov.2014.05.006

Gómez-Carmona, C. D., Gamonales, J., Pino-Ortega, J., & Ibáñez, S. (2018). Comparative analysis of load profile between small-sided games and official matches in youth soccer players. *Sports*, *6*(4), 173. https://doi.org/10.3390/sports6040173

Hill-Haas, S. V., Dawson, B. T., Coutts, A. J., & Rowsell, G. J. (2009). Physiological responses and time–motion characteristics of various small-sided soccer games in youth players. *Journal of Sports Sciences*, *27*(1), 1–8. https://doi.org/10.1080/02640410802206857

Krustrup, P., Mohr, M., Ellingsgaard, H., & Bangsbo, J. (2005). Physical demands during an elite female soccer game: Importance of training status. *Medicine & Science in Sports & Exercise*, *37*(7), 1242–1248. https://doi.org/10.1249/01.mss.0000170062.73981.94

McArdle, W. D., Katch, F. I., & Katch, V. L. (2010). *Exercise physiology: Nutrition, energy, and human performance*. Lippincott Williams & Wilkins.

Minetti, A. E., Moia, C., Roi, G. S., Susta, D., & Ferretti, G. (2002). Energy cost of walking and running at extreme uphill and downhill slopes. *Journal of Applied Physiology*, *93*(3), 1039–1046. https://doi.org/10.1152/japplphysiol.01177.2001

Orendurff, M. S., Walker, J. D., Jovanovic, M., Tulchin, K. L., Levy, M., & Hoffmann, D. K. (2010). Intensity and duration of intermittent exercise and recovery during a soccer match. *The Journal of Strength & Conditioning Research*, *24*(10), 2683–2692. https://doi.org/10.1519/JSC.0b013e3181bac463

Osgnach, C., Poser, S., Bernardini, R., Rinaldo, R., & Diprampero, P. E. (2010). Energy cost and metabolic power in elite soccer: A new match analysis approach. *Medicine & Science in Sports and Exercise*, *42*(1), 170–178. https://doi.org/10.1249/MSS.0b013e3181ae5cfd

Pillitteri, G., Giustino, V., Petrucci, M., Rossi, A., Leale, I., Bellafiore, M., Thomas, E., Iovane, A., Palma, A., & Battaglia, G. (2023). Match load physical demands in U-19 professional soccer players assessed by a wearable inertial sensor. *Journal of Functional Morphology and Kinesiology*, *8*(1), 22. https://doi.org/10.3390/jfmk8010022

Rampinini, E., Coutts, A., Castagna, C., Sassi, R., Impellizzeri, F. (2007). Variation in top level soccer match performance. *International Journal of Sports Medicine*, *28*(12), 1018–1024. https://doi.org/10.1055/s-2007-965158

Riboli, A., Coratella, G., Rampichini, S., Cé, E., & Esposito, F. (2020). Area per player in small-sided games to replicate the external load and estimated physiological match demands in elite soccer players. *Public Library of Science ONE*, *15*(9), e0229194. https://doi.org/10.1371/journal.pone.0229194

Riboli, A., Esposito, F., & Coratella, G. (2022). Small-sided games in elite football: Practical solutions to replicate the 4-min match-derived maximal intensities. *The Journal of Strength & Conditioning Research*, *37*(2), 366–374. https://doi.org/10.1519/JSC.0000000000004249

Rossi, A., Pappalardo, L., & Cintia, P. (2021). A narrative review for a machine learning application in sports: An example based on injury forecasting in soccer. *Sports*, *10*(1), 5. https://doi.org/10.3390/sports10010005

Stølen, T., et al. (2005). *Physiology of soccer: An update*. Sports medicine.

Varley, M. C., Fairweather, I. H., & Aughey1, R. J. (2012). Validity and reliability of GPS for measuring instantaneous velocity during acceleration, deceleration, and constant motion. *Journal of Sports Sciences*, *30*(2), 121–127. https://doi.org/10.1080/02640414.2011.627941