©Journal of Sports Science and Medicine (2025) **24**, 565-577
http://www.jssm.org DOI: https://doi.org/10.52082/jssm.2025.565

Research article

Beyond Playing Positions: Categorizing Soccer Players Based on Match-Specific Running Performance Using Machine Learning

Michel de Haan [1], Stephan van der Zwaard [1,2], Jurrit Sanders [3], Peter J. Beek [1] and Richard T. Jaspers [1]✉

[1] Department of Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, Netherlands; [2] Department of Cardiology, Amsterdam University Medical Center, location AMC, University of Amsterdam, Amsterdam, Netherlands; [3] PSV Eindhoven, Eindhoven, Netherlands

Abstract

Soccer players are frequently categorized by playing positions, both in the scientific literature and in practice. However, the utility of this approach in evaluating physical match performance and optimizing physical training programs remains unclear. This study compares the effectiveness of categorizing soccer players by their playing position versus using unsupervised machine learning based on match-specific running performance. Match-specific running data were collected from 40 young elite male soccer players over two seasons. Thirty-one of these players completed a 20-meter sprint test and a maximal incremental treadmill test to measure maximal oxygen uptake. Players were categorized both by playing position and by subgroups derived through *k*-means clustering based on match-specific running performance. Differences in sprint capacity, endurance capacity, and match-specific running performance were compared between and within playing positions, as well as between and within clusters. The two categorization methods were further compared for variance within subgroups and standardized differences between subgroups for total distance (TD), low-intensity running (LIR), moderate-intensity running (MIR), high-intensity running (HIR), and sprint distance during matches. Match-specific running performance differed between playing positions, despite notable inter-individual differences in running intensities within playing positions. Clustering based on match-specific running performance revealed less variance within groups (TD: P = 0.049, LIR: P = 0.032, HIR: P = 0.033) and larger standardized differences between groups (LIR: P = 0.037, MIR: P = 0.041, HIR: P = 0.035, Sprint: P = 0.018) compared to grouping by playing position. Moreover, 20-meter sprint speed differed between the sprint and high intensity endurance clusters (25.22 vs 23.75 km/h, P = 0.012), but not between playing positions. Using unsupervised machine learning to categorize soccer players improves the identification of player groups with similar match-specific running performance, thereby supporting performance evaluation and contributing to the optimization of physical training.

Key words: Clustering, football, artificial intelligence, physiology, sprint speed, $\dot{V}O_{2max}$.

Introduction

Soccer performance depends on multiple factors, including technical, tactical, and physical aspects. The technical and tactical factors are often viewed as the strongest determinants of soccer performance and hence receive the most attention in training (Fuhre et al., 2022; Soós et al., 2022; Wing et al., 2020). However, the physical factors underlying soccer performance are receiving more and more attention over the last decade, especially in view of the steep increase in match intensity, match load and the increasing number of matches played (Julian et al., 2021; Modric et al., 2021a).

Soccer is a dynamic sport, which requires high endurance capacity, since elite soccer players typically cover 10 to 12 kilometers during a 90-minute match (Stølen et al., 2005). In the context of this endurance performance, the players perform a wide array of explosive activities and actions involving changes in pace and direction, such as sprinting, jumping, kicking, tackling, and turning (Stølen et al., 2005). Studies have estimated that the average work rate during a match is approximately 70% of the maximal oxygen uptake ($\dot{V}O_{2max}$), with lactate concentrations averaging 3 - 9 mM, and values frequently exceeding 10 mM (Bangsbo, 1994).

Physical conditioning plays an important role in soccer performance, as evidenced by the fact that both endurance and sprint capacities are higher in elite soccer players compared to sub-elite or amateur players (Stølen et al., 2005). Additionally, physical conditioning appears to play a key role in improving distance covered, work intensity, number of sprints and number of ball involvements during a match (Helgerud et al., 2001). Furthermore, although injuries have many causes (Bittencourt et al., 2016), it has been shown that appropriate physical training can reduce the risk of injury and improve players' overall fitness, which in turn may help protect against injury, leading to greater physical output and resilience in competition (Gabbett, 2016, 2020; Towlson et al., 2021). This is critically important as the prevalence of injuries has increased markedly over the last 20 years (Ekstrand et al., 2016). These findings and data indicate that physical factors play an important role in soccer and should therefore not be ignored in training.

To efficiently train the physical aspects of soccer, it is essential to understand the load experienced during a match. Match load can be quantified through various methods, broadly categorized into internal and external load metrics. Internal load represents the psychophysiological stress experienced by the player and is often measured using heart rate or rate of perceived exertion, while external load represents the dose performed (Campos-Vazquez et al., 2015; Jaspers et al., 2018). External match load can be quantified through match-specific running performance,

Received: 26 June 2024 / Accepted: 27 June 2025 / Published (online): 01 September 2025

defined as the combination of distance covered and speeds achieved. Understanding the load experienced during a match is particularly important given the various playing positions in soccer, each of which is associated with different match tasks and varying external loads, as reflected in match-specific running performance (Bangsbo et al., 1991; Gil et al., 2007; Wisloeff et al., 1998). For example, Mohr et al. (2003) reported that the total distance covered and high intensity running distance were higher for midfield players, full-backs and attackers compared to central defenders. Additionally, attackers and full-backs covered a greater distance sprinting than midfield players and defenders. Given that different playing positions are associated with distinct external match loads; it may be advantageous to tailor physical training to the specific demands of each position. In line with this, several studies have emphasized the importance of tailoring specific physical abilities, such as $\dot{V}O_{2max}$, jump height, acceleration and sprint speed, to the playing positions of soccer players (Filter et al., 2023; Metaxas, 2021; Slimani et al., 2019). This practice of grouping players by playing position is common, not only in the scientific literature, but also in real-world training settings. There is, however, no standardized method for categorizing soccer players by playing position, as evidenced by the variety of methods used in previous studies (Gil et al., 2007). For instance, in some studies players are crudely classified into forwards, midfielders, and defenders (Soós et al., 2022), while in other studies defenders are differentiated further into central defenders and full-backs and midfielders into attacking and defending midfielders or wide/external and central midfielders (Bangsbo, 2014). In some studies, midfielders are even divided into three groups: central defensive, wide, and central attacking midfielders (Dellal et al., 2011). However, these categorizations are based on tactical positioning rather than on physical characteristics or match load factors. This means that despite these distinctions in playing positions, it is still unclear if this type of categorization is optimal for evaluating physical match performance and guiding physical training programs.

Although training should ideally be tailored to each individual player's specific physical capacities and match requirements, soccer is a multifaceted team sport that not only depends on physical performance but also requires high levels of proficiency in other domains. The substantial time needed to address other essential components, such as technical skills, tactical insight, and team cohesion, often limits the feasibility of implementing individualized physical training programs and seeing them through in full.

A potential alternative involves categorizing players into distinct subgroups based on their individual external load displayed during a match. In recent years, several studies have applied machine learning techniques to study internal and external load variables in soccer (Jaspers et al., 2018; Pillitteri et al., 2024; Rico-González et al., 2023). Machine learning techniques offer a promising approach for identifying clusters of players exhibiting similar external match load, which would enable the development and implementation of training strategies that specifically target players' match load. Such clustering analysis can be performed using unsupervised machine learning, such as by employing the *k*-means algorithm (Hartigan and Wong, 1979). This technique has previously been successful in identifying subgroups with similar physical characteristics in professional soccer players (Novack et al., 2013). Additionally, this technique has successfully identified distinct subgroups in other sports, such as elite cyclists based on their anthropometric characteristics (van der Zwaard et al., 2019). In this study we use *k*-means clustering to categorize soccer players according to their individual match-specific running performance to create subgroups of players with similar external match load.

This study compares the effectiveness of categorizing young elite soccer players based on playing position to categorizing them using clusters based on match-specific running performance. Sprint capacity, endurance capacity, and match-specific running performance are compared both between and within playing positions as well as between and within clusters. Additionally, the study explores the potential of unsupervised machine learning to enhance player categorization by assessing how well both categorization methods identify subgroups with similar match-specific running performance.

We hypothesize that distinct playing positions are associated with specific roles and tasks during matches. These differences in roles and demands are expected to be reflected in the match-specific running performances and sprint and endurance capacities exhibited by soccer players playing at different positions. However, we expect that employing clustering analysis offers a more nuanced approach for grouping soccer players, as positional roles can be executed in markedly different ways, depending on both the individual capabilities of the player and the strategic decisions made by the coaching staff.

Methods

Participants

Forty (40) young male elite soccer players at a professional football club of international caliber were included in this study (U18 and U21, age = 18.0 ± 1.1 years, height = 1.79 ± 0.06 m, weight = 71.4 ± 6.1 kg; mean ± standard deviation). U18 played in the highest league for their age group in the Netherlands (Eredivisie) and U21 played professionally in the second highest league in the Netherlands (KKD: Keuken Kampioen Divisie). They were grouped according to their most frequently played playing position: 10 forwards (F), 9 attacking midfielders (AM), 7 defending midfielders (DM), 6 full-backs (FB) and 8 central defenders (CD). Playing position was determined as assigned by the coach during matches (see Appendix A for a schematic overview of the playing positions). Goalkeepers were excluded from the study, because they have a distinctly different match-specific running performance than outfield players. Match-specific running data were collected from all 40 players over two full seasons, with 31 participants undergoing exercise testing at the start of the second season.

Ethical statement

The study was conducted in full compliance with the Declaration of Helsinki (2013) and was approved by The Scientific and Ethical Review Board (VCWE-2023-054) of the Faculty of Behavioural and Movement Sciences of the Vrije Universiteit Amsterdam. Participants provided written informed consent.

Match-specific running performance

Match-specific running performance data was collected over two seasons from the U18 and U21 teams using three positional tracking systems. The first system is a local position measurement system (Inmotio Local Position Measurement (LPM); Inmotio, Zeist, the Netherlands) with an overall sample frequency of 1,000 Hz divided by the number of active transponders on the field. The average measurement frequency per active transponder varied from 40 to 80 Hz over the measurement sessions, depending on the number of active transponders. The LPM system has been shown to be an accurate and valid measurement device for player tracking in football with a mean difference from the actual distance of at most -1.6% (Frencken et al., 2010; Ogris et al., 2012) and an accuracy of 10 cm according to the manufacturer. The second system is a GPS-based system (Inmotio GPS; Insiders, Lausanne, Switzerland), which measures with a frequency of 10 Hz and has an accuracy of 30 cm. The third system is an optical tracking system (SciSports Optical tracking; Panoris, Brno, Czech Republic), which has a measurement rate of 25 Hz. All collected data was processed using imoClient software (Inmotio, Zeist, the Netherlands). Data was available from all home and away games (n = 819, with an average of 20.5 games played per player). The same data filtering of positional data was applied to all tracking systems, using 100% weighted Gaussian average filter and a 500-ms speed frame interval. Data obtained during friendly games were not included in the dataset, as their physical performance in these games could differ from actual match condition (Modric et al., 2019). Furthermore, a minimum of 80 minutes of playing time was required for a match to be included. The position data was integrated over time and categorized into multiple speed ranges. Running distances were classified as follows: Low Intensity Running (LIR): <14 km/h, Moderate Intensity Running (MIR): ≥14 and <19 km/h, High Intensity Running (HIR): ≥19 and <24 km/h, and Sprinting: ≥24 km/h. These zones are based on threshold values established by the professional soccer team and closely correspond with those described in the literature (Gualtieri et al., 2023; Vieira et al., 2019). Additionally, a combined metric of all zones, referred to as total distance (TD), was included.

To investigate if playing positions exhibit distinctly different match-specific running performances, the five different positional groups (F, AM, DM, FB, CD) were compared regarding the distance covered at different intensities (LIR, MIR, HIR, and sprinting) as well as the total distance covered during the match (TD).

When categorizing players solely based on their most played position, players that distributed their playing time over two playing positions can unfairly influence the running performance of their most played position. As an example, one player played 15 games as full-back and 23 as central defenders. Therefore, to enable a fair comparison between the match-specific running performance belonging to a specific playing position, the match data was categorized based on the actual position played in each match, rather than according to the most frequently played position for each player. For the player in the example that would mean splitting the data into one data point reflecting 15 games played as full-back and one data point reflecting 23 games played as central defender. To address potential bias in match-specific running performance influenced by players occupying a position only once or twice for tactical reasons or out of necessity due to injury, an inclusion criterion was invoked, which required that a minimum of three games had to be played in a specific position. Of the 40 players assessed, 17 players had accumulated more than three matches at two different positions, while two players met this criterion at three distinct positions. One player did not meet the three-match threshold for any position, and the remaining 20 players had more than three matches played in only one position. This inclusion process ultimately resulted in 60 data points of average match-specific running performances (consisting of a total of 744 individual matches and an average of 12.4 matches per data point) at a playing position.

Physical capacity

Sprint capacity: The sprint capacity was measured during an all-out linear sprint test over 20 meters on an artificial grass surface. This sprint distance was used because it is representative for soccer sprints during competition and provides a relevant and practical indicator of the player's capacity to generate explosive power in match-specific contexts (Nikolaidis et al., 2016b). Before the sprint test, participants underwent a standard warm-up routine for soccer practice designed by the physical training staff of the team. This routine consisted of dynamic stretching, running exercises and footwork drills. Participants were instructed to cover the 20 meters as fast as possible from a static start. They performed this measurement twice and the fastest time was used for further analysis. Participants were familiar with the test protocol as this is a standard testing procedure performed in daily practice by the team. Positional data was obtained using LPM (Inmotio, Zeist, the Netherlands; Inmotio GPS; Insiders, Lausanne, Switzerland) and integrated over time to determine the average sprint speed over 20 meters.

Endurance capacity: $\dot{V}O_{2max}$ is considered the gold standard for measuring endurance capacity and represents a critical determinant of endurance performance (van der Zwaard et al., 2021). In this study the endurance capacity was defined as the $\dot{V}O_{2max}$ obtained during a maximal incremental treadmill test (Kemi et al., 2003). The speed of the motorized treadmill (H/P/COSMOS - Pulsar 3P, Samcon bvba, Melle, Belgium) was increased by 1.5 km/h every 2 minutes, starting at 8.5 km/h. The measurement ended either when the player was unable to run at the speed of the treadmill and voluntarily stepped off or when they fell and were suspended by the safety harness. Breath-by-breath gas exchange analysis (Vyntus CPX, Jaeger-Care-Fusion, UK) was used to measure $\dot{V}O_{2max}$. Calibration was performed according to the manufacturer's instructions. The gas analyzer was calibrated using automatic reference gas calibration (15% O2, 5% CO2, 80% N2) and volume transducer was calibrated using the automatic integrated blower. Breath-by-breath data was smoothed and $\dot{V}O_{2max}$ was calculated as the highest 30-s value. $\dot{V}O_{2max}$ was normalized to lean body mass$^{2/3}$ to eliminate the influence of body size (McCann and Adams, 2002; van der Zwaard et al., 2018a). Note that no normalization for body size was performed for sprint capacity, as the physical dimension of speed is already size-independent, dividing distance by time (LT^{-1}). Participants were instructed to avoid strenuous exercise for 30 hours leading up to the exercise testing.

Machine learning

Clustering was used to identify subgroups of players with similar match-specific running performance characteristics. The clustering was performed using the *k*-means algorithm, an unsupervised machine learning technique, which is described in detail in the literature (Hartigan and Wong, 1979; van der Zwaard et al., 2019). During multiple iterations, data points were assigned to the most nearby centroid based on their Euclidean distance. Initial centroid positions were obtained at random. During each iteration, the location of the centroid was recalculated as the average position of all assigned data points to that specific cluster. This process was repeated until the total within sum of square was minimized, and the location of the centroids no longer changed. For the present analysis, a maximum of 100 iterations was used and optimization was performed using 15 random starting partitions to enhance cluster stability. As input variables, we used the four speed categories, LIR, MIR, HIR, and sprint distance, which were normalized to Z-scores before being entered into the algorithm. The optimal number of clusters for the present data was determined to be 5, based on an evaluation of the elbow and silhouette plots. The stability of the cluster was evaluated by repeating the *k*-means algorithm 1,000 times and examining whether cluster assignment was consistent over these 1,000 runs. After this machine-learning analysis, differences in sprint, endurance and match-specific running performance were evaluated between the identified clusters. Additionally, we evaluated how playing positions were distributed among the clusters.

Statistical analysis

To ensure the validity of the statistical analyses, the Shapiro-Wilk test and visual inspection of data distribution plots were used to verify normality. One-way ANOVAs or the non-parametric Kruskal-Wallis were performed to assess differences between playing positions on average 20-meter sprint speed, $\dot{V}O_{2max}$ and match-specific running performance, specifically total distance, LIR, MIR, HIR, and sprint distance. Similarly, unique characteristics of the clusters identified by machine learning were compared between the clusters using a one-way ANOVA. Following the ANOVA tests, the Bonferroni post-hoc test was used to identify which specific groups exhibited statistically significant differences. Additionally, the two methods of grouping soccer players - i.e. that based on playing position and that based on clustering of match-specific running performance - were compared statistically. To test which of these methods resulted in the most uniform subgroups, for each method we obtained the Coefficient of Variation (*CV*) within the subgroups and the effect sizes (Cohen's *d*) of the differences between subgroups using pairwise comparesons. Effect sizes are evaluated according to Cohen (1988). For total distance, LIR, MIR, HIR, and sprint distance, independent *t*-tests were employed to test for significant differences in CV and effect size between the two grouping methods. *F*-tests were conducted to examine whether the data from the two methods exhibited equal variation, and in cases of unequal variation, the Welch's correction was applied. In cases of a non-normal distribution, the non-parametric Mann-Whitney U test was used instead. Results were considered statistically significant if the *P*-value was equal to or less than 0.05 (α).

Results

Physical capacity

The 20-meter sprint test was performed by 27 players and the maximal incremental treadmill test by 28 players, with 24 players successfully completing both assessments. The average sprint speed of these players on the 20-meter sprint test was 24.36 ± 0.66 km/h, with values ranging from 23.08 to 26.17 km/h. The average $\dot{V}O_{2max}$ relative to body weight was 57.93 ± 3.91 mL/kg/min. When normalized to $LBM^{2/3}$, the average $\dot{V}O_{2max}$ was 257.3 ± 14.51 $mL/kg_{LBM}^{2/3}/min$, with values ranging from 232.30 to 295.90.

Physical capacity and match-specific running performance across playing positions

Comparing sprint and endurance performance between playing positions revealed no significant differences in sprint speed ($P = 0.769$) or endurance capacity ($P = 0.241$) (Table 1, Figure 1a and Figure 1b). However, the playing positions showed marked differences in match-specific running performances. In competition, there were significant differences in total distance (F, AM, DM > CD; AM, DM > FB; DM > F), low intensity distance (DM, AM > FB; DM > CD = F), moderate intensity distance (DM = AM > F = FB = CD), high intensity distance (AM = F = FB = DM > CD) and sprint distance (F, FB > CD; F > DM) (Table 2, Figure 2). The positional analysis of sprint distance was performed using the non-parametric Kruskal-Wallis test with Bonferroni correction, because the data exhibited a non-normal distribution on a Shapiro Wilk test ($P = 0.001$).

The central defenders covered significantly less distance at high intensity (688 m vs 486 m, $P < 0.001$) or sprinting (249 m vs 150 m, $P = 0.012$) compared to the full-backs, with the full-backs nearly sprinting the most and the central defenders sprinting the least of all positional groups. Additionally, the forwards covered a significantly greater sprint distance (258 m vs 150 m, $P = 0.010$) and a significantly shorter total (10,726 m vs 11,388 m, $P = 0.028$), low (8,004 m vs 8,494 m, $P = 0.007$) and moderate intensity running distance (1,708 m vs 2,046 m, $P = 0.012$) compared to the defending midfielders, but not when compared to the attacking midfielders. Similarly, the central defenders covered a shorter low-intensity distance compared to the defending midfielders (7,960 m vs 8,494 m, $P = 0.007$), but not compared to the attacking midfielders. Even though there were significant differences between playing positions, there was also considerable interindividual variation within the different positional groups (average CV within positional groups: TD: 4.5%, LIR: 3.9%, MIR: 12.7%, HIR: 13.6% and sprint: 33.5%).

Categorizing players based on clusters with similar match-specific running performance

Players can also be grouped by similarities in match-specific running performance rather than by playing position. Using unsupervised machine learning, we identified five distinct clusters of players with similar match-specific running performances (Figure 3). Based on their characteristics, we refer to these clusters as Low Intensity Endurance group (LIE), High Intensity Endurance group (HIE), Sprint group (SPR), Low Sprint and Endurance group (LOW), and Balanced group (BLC). Cluster characteristics are summarized in Table 3 and visualized in Figure 4. Among the identified clusters, SPR (sprint) represented a sprint-focused subgroup, covering relatively few meters at speeds below 19 km/h. This group consisted of the most explosive full-backs and forwards. LOW (low sprint and endurance), on the other hand, contained players with the lowest physical load, as they covered a moderate distance at low intensity and short distances in all other intensity zones. This group consisted mostly of central defenders, but also contained one player from each other playing position. HIE (high intensity endurance) consisted of individuals who covered the greatest total distance and substantial distances at low, moderate, and high intensity levels. For the most part, this group consisted of attacking midfielders, apart from one defending midfielder and one forward. LIE (low intensity endurance) consisted of players who primarily covered a significant distance at low to moderate intensity and had the least sprint meters of all the cluster groups. This group consisted of attacking and defending midfielders and one forward. Lastly, BLC (balanced) represented a balanced group, incorporating a combination of different intensity levels. This group included players from all playing positions, but primarily full-backs and forwards. The physical capacities of the players in the clusters are displayed in Figure 5. Similar to the positional analysis, there were no significant differences between the five clusters for endurance capacity (P = 0.235). However, on the 20-meter sprint test, average speed of SPR was ~1.5 km/h faster than that of HIE (25.22 vs 23.75, P = 0.012).

Table 1. Average sprint and endurance capacity per position.

	Forward	Attacking midfielder	Defending midfielder	Central defender	Full-back
Average 20-m sprint speed	24.59 ± 0.7	24.13 ± 0.6	24.25 ± 0.9	24.35 ± 0.4	24.44 ± 0.8
Normalized $\dot{V}O_{2max}$	253.1 ± 14.5	255.4 ± 8.1	261.3 ± 23.7	250.3 ± 5.0	270.7 ± 10.2

Speed parameters are expressed in km/h. Normalized $\dot{V}O_{2max}$ is normalized to $LBM^{2/3}$ and is expressed as: $mL/kg_{LBM}^{2/3}/min$.

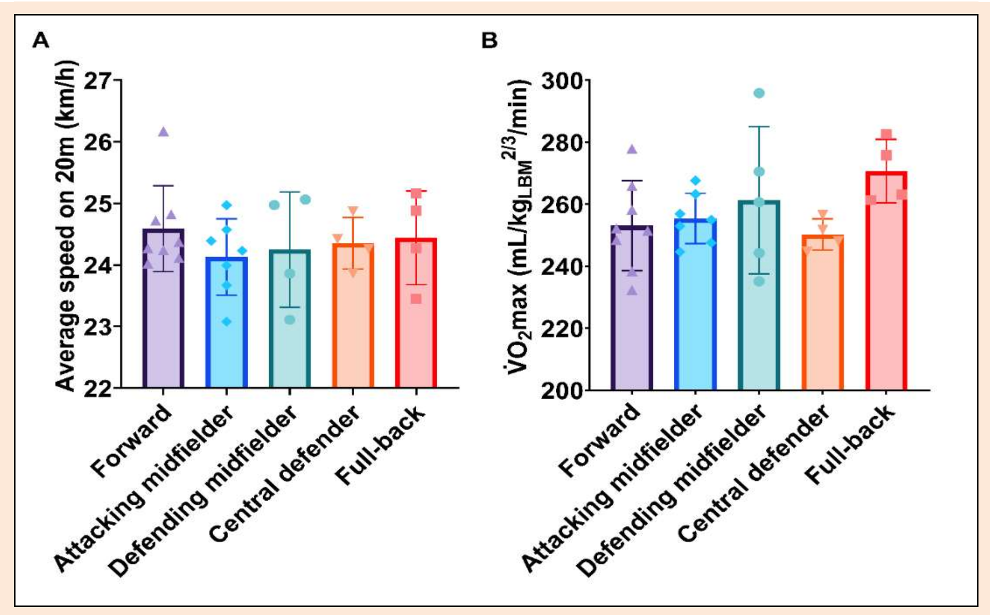

Figure 1. Average 20-meter sprint speed and $\dot{V}O_{2max}$, normalized to lean body mass, for each playing position. Each data point represents individually obtained values of male young elite soccer players. A) Means and standard deviations for average sprint speed are presented for forwards (n = 9), attacking midfielders (n = 8), defending midfielders (n = 4), central defenders (n = 4), and full-backs (n = 4). B) Means and standard deviations of normalized $\dot{V}O_{2max}$ are displayed for forwards (n = 9), attacking midfielders (n = 8), defending midfielders (n = 5), central defenders (n = 4), and full-backs (n = 4). There were no significant differences between the playing groups for average sprint speed (P = 0.769) and normalized $\dot{V}O_{2max}$ (P = 0.241).

Table 2. Average match-specific running performance per playing position.

	Forward	Attacking midfielder	Defending midfielder	Full-back	Central defender
TD	10726 ± 597	11295 ± 620	11388 ± 354	10539 ± 513	10115 ± 338
LIR	8004 ± 366	8309 ± 381	8494 ± 337	7899 ± 323	7960 ± 197
MIR	1708 ± 306	2011 ± 269	2046 ± 160	1685 ± 179	1501 ± 207
HIR	738 ± 114	762 ± 125	681 ± 83	688 ± 99	486 ± 46
Sprint	258 ± 77	195 ± 70	150 ± 65	249 ± 103	150 ± 26

Mean ± SD of the five different playing positions. TD = total running distance, LIR = low intensity running distance (<14 km/h), MIR = moderate intensity running distance (≥14 and <19 km/h), HIR = high intensity running distance (≥19 and <24 km/h), Sprint = sprinting distance (≥24 km/h) in meters.

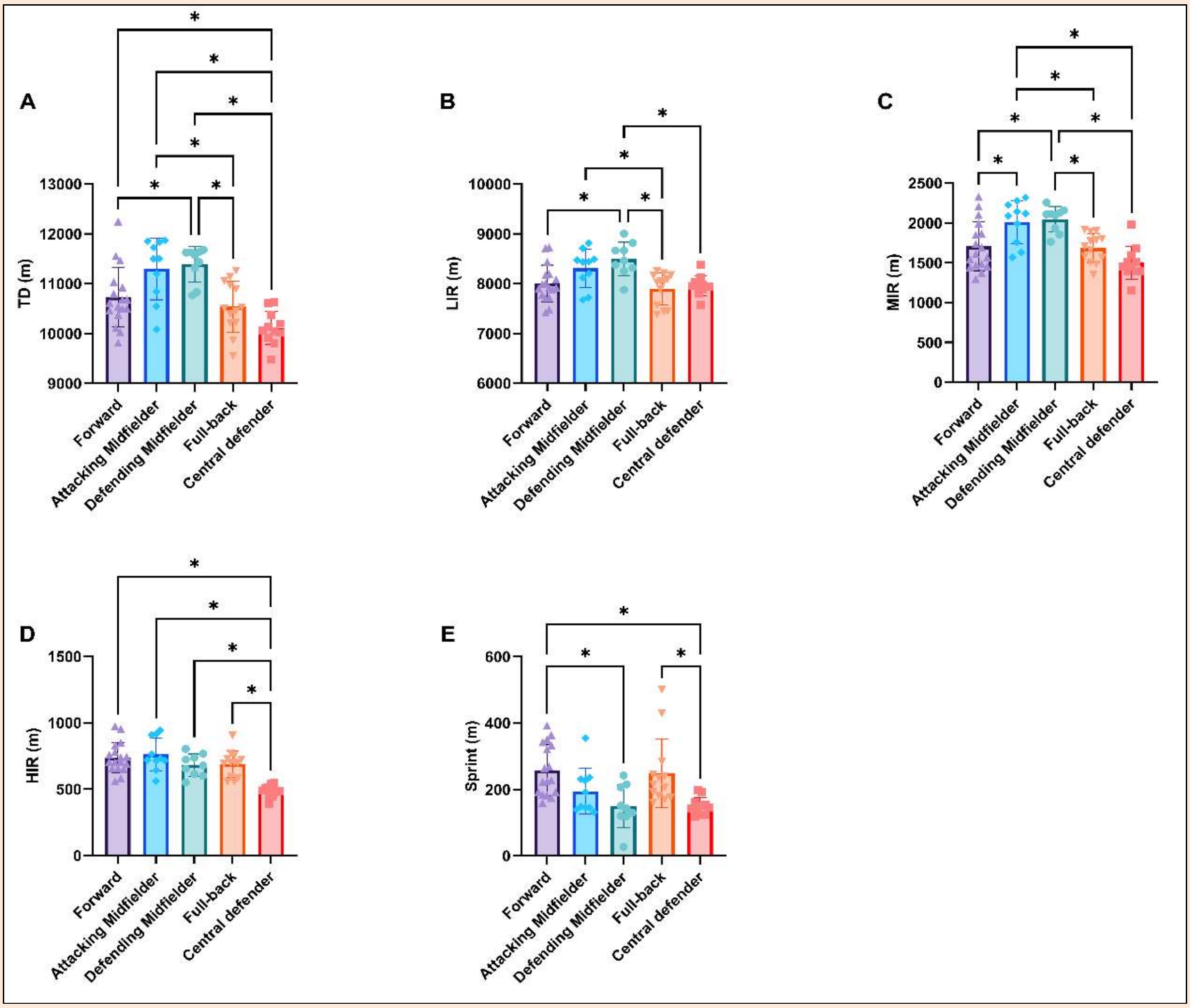

Figure 2. Average match-specific running performance for each playing position obtained over two full seasons. Match-specific running performance was determined based on A) Total Distance (TD), B) Low Intensity Running (LIR): <14 km/h, C) Moderate Intensity Running (MIR): ≥14 and <19 km/h, D) High Intensity Running (HIR): ≥19 and <24 km/h, and E) Sprinting: ≥24 km/h. Each data point represents the average match performance, in meters covered, of young elite male soccer players achieved when playing at a certain position (with a minimum of 3 games played at that position). Mean ± SD are shown for forwards (n = 17), attacking midfielders (n = 10), defending midfielders (n = 9), full-backs (n = 13), and central defenders (n = 11). Significance is indicated (* $P < 0.05$) and determined by one-way ANOVA with Bonferroni correction, except for sprint distance, where the non-parametric Kruskal-Wallis test with Bonferroni correction has been applied. There were significant differences between the different playing positions in total distance covered (F, AM, DM > CD; AM, DM > FB; DM > F), distance covered at low intensity (DM, AM > FB; DM > CD = F), distance covered at moderate intensity (DM = AM > F = FB = CD) , distance covered at high intensity (AM = F = FB = DM > CD) and distance covered while sprinting (F, FB > CD; F > DM).

Categorizing players by playing positions or clustering

We found notable interindividual variation in match-specific running performance when categorizing soccer players either by playing position or clustering. To test which of these methods resulted in the most uniform subgroups, we compared the within-group variation between categorizing based on playing position and clusters. We performed this analysis separately for the distances run at each of the four intensity zones and for the total running distance. There was significantly less variation in the cluster groups compared to the positional groups for total distance (3.1% vs 4.5%, $P = 0.049$), low (2.6% vs 3.9%, $P = 0.032$) and high intensity running (10.0% vs 13.6%, $P = 0.033$) distance. Additionally, there was a trend towards significance for moderate intensity running distance (8.0% vs 12.7%, $P = 0.077$). There was no significant difference in variance between categorization methods for sprint distance (25.0% vs 33.5%, $P = 0.237$). We also compared standardized differences between subgroups based on pairwise comparisons. After grouping based on playing positions, average effect sizes were large for all match-specific performance categories (TD: 1.42, LIR: 0.95, MIR: 1.29, HIR: 1.39, Sprint: 0.86), indicating distinct differences between positional subgroups. However, the effect sizes for the clustering method were even larger (TD: 2.54, LIR: 2.05, MIR: 2.92, HIR: 2.65, Sprint: 2.30). Comparison between the two methods showed that the effect size for the clustering method was significantly larger for low (2.05 vs 0.95, $P = 0.037$), moderate (2.92 vs 1.29, $P = 0.041$), high intensity running (2.65 vs 1.39, $P = 0.035$) and sprint (2.30 vs 0.86, $P = 0.018$) distance and showed a tendency for total distance (2.54 vs 1.42, $P = 0.080$).

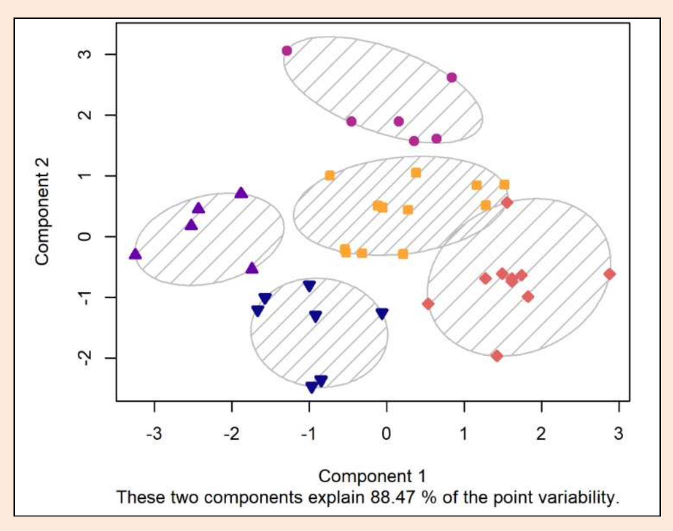

Figure 3. Cluster plot with a two-dimensional representation of the five physiological clusters. Clusters are displayed in the two most important dimensions (explaining 88.47% of point variability), which represent a combination of four different speed categories: LIR (<14 km/h), MIR (≥14 and <19 km/h), HIR (≥19 and <24 km/h) and Sprint ≥24 km/h). Individual values and spanning ellipses of clusters are presented for LIE (low intensity endurance cluster), HIE (high intensity endurance cluster), SPR (sprint cluster), LOW (low endurance and sprint cluster) and BLC (balanced cluster).

Table 3. Average match-specific running performance per cluster.

	LIE	HIE	SPR	LOW	BLC
TD	11463 ± 269	11793 ± 281	10641 ± 381	10152 ± 316	10554 ± 411
LIR	8646 ± 198	8420 ± 193	7832 ± 158	8049 ± 201	7890 ± 307
MIR	2016 ± 138	2252 ± 56	1645 ± 163	1440 ± 124	1748 ± 210
HIR	653 ± 72	887 ± 76	757 ± 77	500 ± 58	691 ± 60
Sprint	131 ± 55	217 ± 41	388 ± 67	146 ± 43	207 ± 36

Mean ± SD of the five different cluster groups. TD = total running distance, LIR = low intensity running distance (<14 km/h), MIR = moderate intensity running distance (≥14 and <19 km/h), HIR = high intensity running distance (≥19 and <24 km/h), Sprint = sprinting distance (≥24 km/h) in meters.

Discussion

This study compared the effectiveness of categorizing young elite soccer players based on playing position and categorizing them by using clusters derived from match-specific running performance. Sprint capacity, endurance capacity, and match-specific running performance were compared both between and within playing positions as well as between and within clusters. Although there were significant differences in match-specific running performance between the playing positions, these differences were not reflected in the maximal sprint and endurance capacity of the players. Despite the observed differences in match-specific running performance between the positions, substantial interindividual variation remained within playing positions. We introduced an alternative way of grouping players based on unsupervised machine learning to identify subgroups of players with similar match-specific running performance. This clustering method yielded more distinct subgroups in terms of external match load, as demonstrated by lower within-group variance and larger effect sizes between subgroups in match-specific running performance. As a result, this approach facilitates targeted training programs towards the specific load experienced in the match by the players. Both of our research hypotheses were thus confirmed.

Physical capacity is similar across playing positions

The average sprint speed of the present group of players on the 20-meter sprint test was 2 km/h faster (24.36 vs 22.42 km/h) than reported reference values for young elite soccer players of U18 and U21 available in literature (Nikolaidis et al., 2016a). The average relative $\dot{V}O_{2max}$ of the young elite soccer players that participated in this study was 57.93 ± 3.91 mL/kg/min, which is in line with average values reported in young elite Norwegian (58.1 mL/kg/min) and Tunisian (61.6 mL/kg/min) players (Chamari et al., 2004; Helgerud et al., 2001). However, exceptions exist, as evidenced by the U18 Hungarian team recording values as high as 73.9 mL/kg/min (Stølen et al., 2005). It was hypothesized that specific playing positions are associated with specific roles and tasks during matches (e.g., intercepting passes or building up play), resulting in varying match-specific running performances and physical capacities exhibited by the players. However, no significant difference in average 20-meter sprint speed or normalized $\dot{V}O_{2max}$ was found between the different playing positions of young elite soccer players. This finding is in line with that of Lago-Peñas et al. (2011), who found no significant difference in 30-meter sprint speed between the different positional groups in young male soccer players of Spanish regional representative teams, and that of Metaxas (2021) who found no significant difference between the $\dot{V}O_{2max}$ of positional groups in Greek elite male soccer players. However, Haugen et al. (2013) found that over a 20-meter sprint, forwards were faster than defenders who in turn were faster than midfielders, in male junior soccer players ranging from 5th division to national level. Bangsbo and Michalsik (2002) observed that for a group of elite Danish male soccer players, central defenders showed significantly lower $\dot{V}O_{2max}$ than full-backs, midfielders and forwards. These differences in findings could arise from different philosophies of teams or countries regarding physical training and application of the capabilities of the players in a match. A possible explanation for a lack of a significant difference in physical capacities between positional groups could be that trainers and coaches do not necessarily distinguish between playing positions in their training programs.

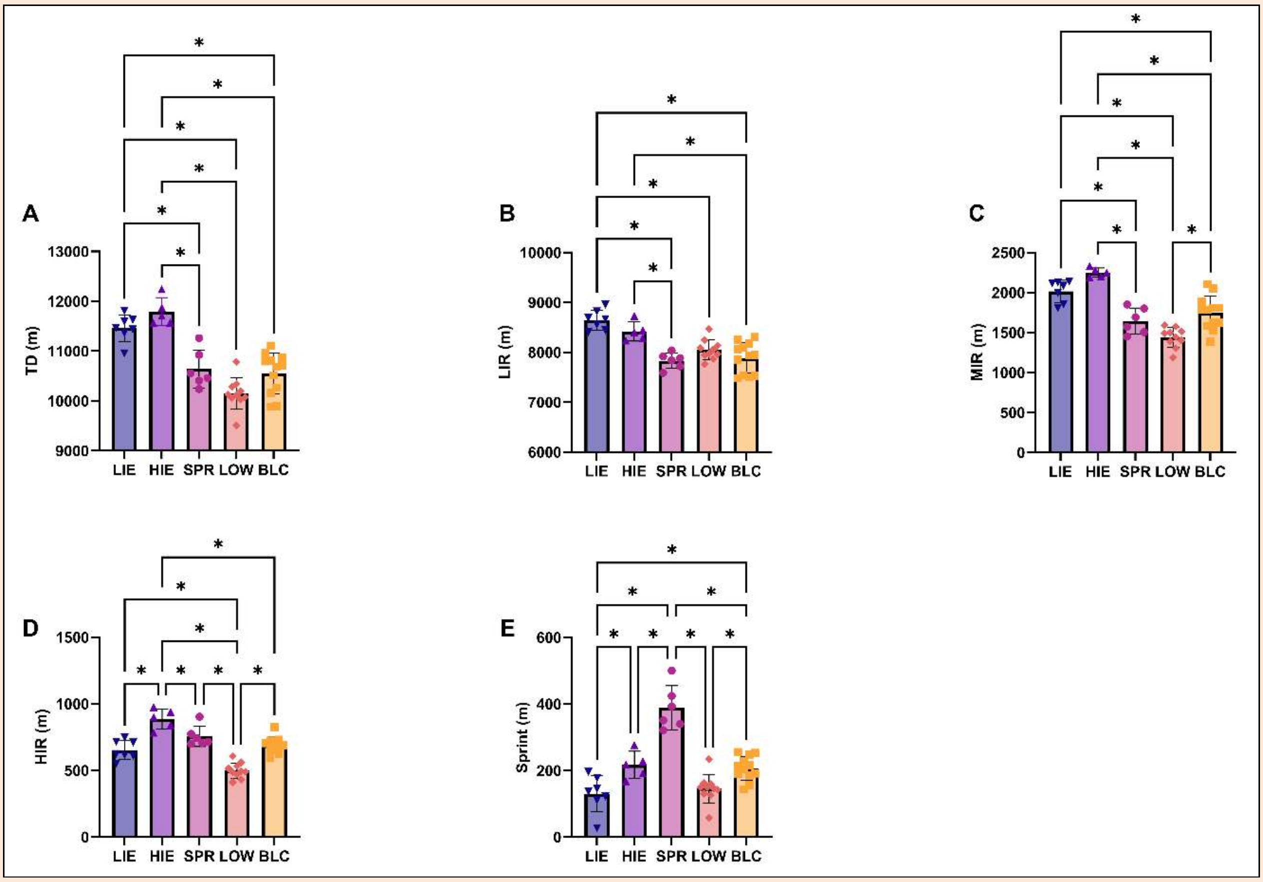

Figure 4. The average match-specific running performance across the five identified physiological clusters. Each data point corresponds to the average match performance, in meters covered, of a young elite soccer player over two full seasons, with mean and standard deviation presented. The clusters are denoted by their characteristics: LIE (low intensity endurance), HIE (high intensity endurance), SPR (sprint), LOW (low endurance and sprint), and BLC (balanced). The comparison of these groups (LIE: n = 7, HIE: n = 5, SPR: n = 6, LOW: n = 10, BLC: n = 12) is depicted for A) total distance covered, B) distance covered at low intensity (<14 km/h), C) distance covered at moderate intensity (≥14 and <19 km/h), D) distance covered at high intensity (≥19 and <24 km/h), and E) distance covered while sprinting (≥24 km/h). Significance is indicated (* $P < 0.05$) and determined by one-way ANOVA with Bonferroni correction.

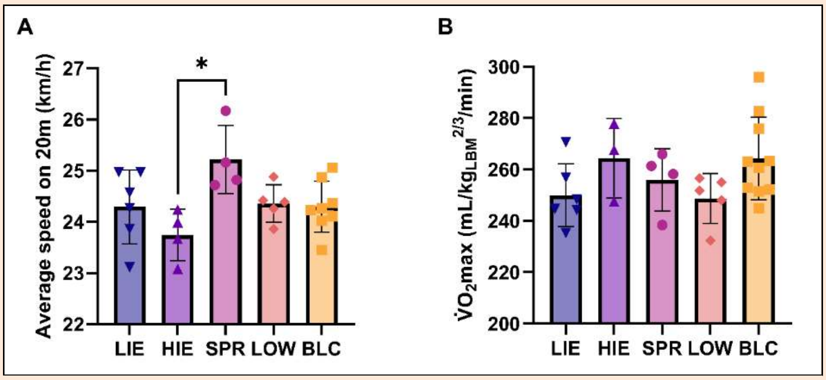

Figure 5. Average 20-meter sprint speed and $\dot{V}O_{2max}$, normalized to lean body mass$^{2/3}$, per playing position. Each data point represents individually obtained values of male young elite soccer players. A) The means and standard deviations for average sprint speed are shown for LIE (n = 6), HIE (n = 4), SPR (n = 4), LOW (n = 5), and BLC (n = 8). B) The means and standard deviations of normalized $\dot{V}O_{2max}$ are shown for LIE (n = 6), HIE (n = 3), SPR (n = 4), LOW (n = 5) and BLC (n = 10). Significance is indicated (* $P < 0.05$) and determined by one-way ANOVA with Bonferroni correction. There were no significant differences between the five different clusters for endurance capacity (P = 0.241). However, on the 20-meter sprint, the average speed of SPR was 1.5 km/h faster than HIE (25.22 vs 23.75, P = 0. 012).

Instead, they might apply their physical training to the entire group or tailor it to the individual players in the group. Individualized physical training can be performed in the gym using specific exercises tailored to each individual's match load or be applied in the form of sport-specific exercises during field training or in the form of small-sided games. It is essential to note that our study population received individualized training, particularly in the form of physical training in the gym. Another possible explanation is optimizing the physical aspects of the individual players is hampered by the busy match schedule, with a match day every 3 to 7 days. This might be easier in other sports with less frequent competitive demands. In line with this hypothesis, when comparing our results with studies in rowing and cycling, the observed range of $\dot{V}O_{2max}$ found in the present study was substantially smaller than in those cyclical sports (van der Zwaard et al., 2016; 2018b). Furthermore, it is possible that trainers tailor the physical training program to reflect the total weekly load experienced by the player, rather than focusing solely on match load. It could be argued that to optimally train all aspects of soccer throughout the week, players require a baseline of physical capacities necessary to perform training exercises effectively (Silva et al., 2016). In our case this would be a minimum average 20-meter sprint speed of 23.08 km/h and a minimum $\dot{V}O_{2max}$ of 49.88 mL/kg/min. This baseline might exceed the physical demands of a match (Buchheit et al., 2021). Such a scenario could lead to relatively uniform physical capacities across players, potentially reducing the observable differences between playing positions to a degree that makes detecting statistically significant differences challenging. Future research is necessary to determine whether there is a mismatch between the physical capacity and average match-specific running performance of soccer players and, if so, what are the causes of this phenomenon.

Match-specific running performance differs between playing positions

Our study demonstrates that central defenders exhibit the shortest high-intensity running distance (≥19 and <24 km/h) of 681 ± 83 m, while attacking and defending midfielders cover the greatest total, low (<14 km/h) and moderate-intensity (≥14 and <19 km/h) distances. Furthermore, forwards and full-backs were shown to perform significantly more sprint (≥24 km/h) meters compared to central defenders. These finding aligns with previous research. Di Salvo et al. (2009) observed significant differences in high-intensity running distance (>19.8 km/h) between wide midfielders (1,049 ± 106 m) and central defenders (681 ± 128 m). Similarly, Modric et al. (2021b) reported that for professional soccer players, low-intensity distance covered (<14.3 km/h) was highest among central midfielders (8,876 ± 545 m), while wide midfielders covered the greatest sprinting distance (>25.2 km/h) of 272 ± 101 meter and high-speed running distance (19.8–25.2 km/h) of 564 ± 115 m. Central midfielders achieved the greatest total distance covered (11,160 ± 644 m), and central defenders achieved the lowest total distance covered (9,257 ± 690 m). Moreover, Bradley et al. (2009) demonstrated that in the English Premier League, wide midfielders (3,138 ± 565 m) covered significantly greater distances during high-intensity running (>14.4 km/h) than central defenders (1,834 ± 256 m), full-backs (2,605 ± 387 m), central midfielders (2,825 ± 473 m), and attackers (2,341 ± 575 m), with central defenders performing significantly less high-intensity running than players in all other positions. Additionally, they showed that wide midfielders (346 ± 115 m) and full-backs (287 ± 98 m) covered greater sprint distance (>25.1 km/h) than central midfielders (204 ± 89 m), attackers (264 ± 87 m), and central defenders (152 ± 50 m). These findings highlight that significant differences in match-specific running performance exist between playing positions categorized as forwards, attacking midfielders, defending midfielders, full-backs, and central defenders.

Players should not be categorized in forwards, midfielders, and defenders when evaluating match-specific running performance

While early studies commonly categorized players into broad groups such as defenders, midfielders, and forwards (Bloomfield et al., 2007; Deprez et al., 2015; Gil et al., 2007; Metaxas, 2021; Wisloeff et al., 1998), more recent research increasingly adopts more refined positional classifications. Despite this shift, these traditional groupings remain in use in certain applied and research contexts due to their simplicity and practicality (Kavanagh et al., 2024; Smpokos et al.; Wei et al., 2024). However, in this study we found a significant difference in match-specific running performance between central defenders and full-backs. The central defenders covered significantly less distance at high intensity or while sprinting compared to full-backs, with full-backs nearly sprinting the most of all the positional groups and central defenders sprinting the least. This could be attributed to the fact that full-backs are positioned alongside the field where they can cover large distances at a high pace to contribute to both offense and defense (Rhini et al., 2024). This fact is overlooked when averaging over all defenders. Creating training programs for full-backs and central defenders together would create a situation in which the full-backs receive too little explosive training, and this could hamper their match performance. Moreover, we found significant differences between forwards and midfielders, but only for defending midfielders. Forwards covered a significantly greater sprint distance and a significantly shorter total, low and moderate intensity distance compared to defending midfielders, but not when compared to attacking midfielders. Similarly, central defenders covered a shorter low-intensity distance compared to defending midfielders, but not when compared to attacking midfielders. This indicates that grouping these attacking and defending midfielders together is suboptimal as the distinct differences in their match performance, compared to forwards and full-backs, would be averaged out. We therefore recommend refraining from grouping players as forwards, midfielders, and defenders when studying their physical capacities, designing their training programs or evaluating their match-specific running performance.

Grouping based on playing position is not optimal for training purposes

While significant differences in match-specific running performance exist among playing positions, our analysis additionally revealed a substantial degree of interindividual variation in match-specific running performance within each positional group. While the interindividual variation is not explicitly quantified using the coefficient of variation in literature about match-specific running performance, substantial interindividual differences in match-specific running performance within playing positions are evident. For example, using match-specific running performance data from Table 1 of Rhini et al. (2024), and Table 3 of Modric et al. (2021b), we calculated the average CV for distance run at high intensity (19.8 - 25.1 km/h) for all positional groups. We applied the same methodology as used for our data, specifically calculating the coefficient of variation for each playing position by dividing the standard deviation by the mean, multiplying by 100, and then averaging the resulting CVs across all playing positions. This results in average CVs of 16.2% and 21.9%, for Rhini et al. (2024) and Modric et al. (2021b) respectively. These values are in line with the substantial average CV of 13.6% for distance run at high intensity (≥19 and <24 km/h) observed across the positional groups in our study. Moreover, This finding is also in line with Bangsbo and Michalsik (2002), who noted considerable interindividual variation within physical capacity of the positional groups, possibly due to different playing styles within the same playing position. As such, grouping players based on position alone is not the most effective strategy to assess match-specific running performance of soccer players; a more personalized approach based on individual match-specific running performance seems more fruitful.

Clustering based on match-specific running performance as an alternative approach to group soccer players

Positional roles can be executed through diverse strategies as evidenced in our study by the considerable variation within positional groups. Using *k*-means clustering allowed us to distinguish five distinct subgroups based on match-specific running performance. These were labelled Low Intensity Endurance group (LIE), High Intensity Endurance group (HIE), Sprint group (SPR), Low Sprint and Endurance group (LOW), and Balanced group (BLC). *k*-means clustering has the advantage that it is not dependent on the a priori assumption that players within the same playing position have the same match-specific running performance. While pairwise comparison of positional groups displayed large average effect sizes for match-specific running performance across all speed categories, cluster groups exhibited average effect sizes approximately twice as large. Additionally, the variation within groups was also smaller for the cluster groups than for playing positions. Furthermore, the SPR and HIE clusters showed significantly different sprint capacity, while there were no differences in physical capacity between the playing positions. This indicates that clustering based on match-specific running performance yields more distinct groups than playing position. Implementation of *k*-means clustering into soccer research and practice could therefore provide coaches and trainers with a better and more nuanced understanding of the individual needs of players within a team, such as tailoring explosive power training for sprint-focused players (SPR group) and emphasizing endurance development for players who cover greater distances (END group), allowing for more targeted and effective training programs.

Limitations

Match-specific running performance is significantly influenced by strategic considerations. For instance, a short sprint distance during a match may not solely reflect a player's physical capacity but could instead be a deliberate strategic choice for tactical purposes. Although this study has deliberately minimized the impact of strategic variations arising from matchups by collecting data over two complete seasons, the overall coaching philosophy and the coach's specific instructions for a player might significantly influence match demand. Additionally, average match-specific running performance is not necessarily the same as match demand or optimal match performance. For example, it is possible that the running performance in a match is limited by the player's physical capacity. For instance, a match may demand a higher level of physical performance from a player than they can deliver, resulting in a discrepancy between match-specific running performance and match demand. Hence, this study does not establish definitive match requirements, but rather provides insights into the current performance levels and the associated clusters players belong to. It is the coach's responsibility to interpret these clustering results with respect to their strategic plan and specific performance requirements for their players. Additionally, this study focuses on youth players, who are still developing, which means the current match performance will likely differ from the match performance necessary to play at the highest level, for which they are being prepared. However, achieving good results at their present playing level is crucial for potential selection into elite competitions.

Results from the *k*-means clustering are specific to the included group of players. It should be noted that when applying this methodology to another team or club, a new cluster analysis will have to be performed, resulting into new clusters specific to that team or club. However, this would also be the case for grouping by playing position as another team or club could have other team compositions, playing styles and/or different requirements for the specific playing positions. Likewise, using a different clustering method could yield, and indeed is likely to yield, different clusters. The robustness of the present findings needs to be established in future studies comparing the results of different clustering methods. However, based on the present findings, it seems highly unlikely that the main conclusion of this study, i.e. that categorization of players based on their position is not optimal from a match evaluation and physical training standpoint, will be refuted in such a comparison.

A sample size of 40 young elite soccer players could be viewed as relatively small in the context of machine learning. However, contrary to traditional statistical analysis, the statistical power for cluster analysis primarily depends on cluster separation, and not on the sample size (Dalmaijer et al., 2022). Among our participants, match-specific running performances were rather heterogeneously distributed, which makes it more difficult to designate all players to distinct subgroups (as indicated by the average silhouette score of 0.34). Nonetheless, Figure 3 displays that we identified distinct and non-overlapping subgroups in the cluster plot. Furthermore, it has been demonstrated that robust and distinct clusters could be obtained with large cluster separation in a similar sample of 36 cyclists (van der Zwaard et al., 2019). Additionally, data points used for clustering reflected participant data collected over two full seasons, consisting of 819 underlying measurements, with an average of 20.5 matches per player.

We expected the cluster analysis to show markedly different sprint and endurance capacities between the groups. However, only on the 20-meter sprint test, the average speed of SPR was 1.5 km/h faster than HIE. A possible explanation for the lack of significant differences between any other clusters could be the small sample size of the cluster groups. The HIE group consisted of only four players, of which only three completed the maximal incremental exercise test. Another possibility is that the difference in physical capacities is only significant between the two most different groups, while the variations among the other groups might be too small to yield statistically significant results. Furthermore, the small sample size also hampered making meaningful comparisons between wingers (n = 7) and central forwards (n = 3), with an insufficient representation of central forwards in the study population.

Conclusion

The present findings show considerable interindividual variation in match-specific running performance within positional groups. This suggests that studying the physical capacities, designing training programs or evaluating match-specific running performance of soccer players based on their playing positions is suboptimal. Particularly grouping by forwards, midfielders and defenders should be avoided as this averages out the differences between central defenders and full-backs and attacking and defending midfielders. Identifying subgroups based on match-specific running performance using clustering analysis seems a promising alternative as it leads to more distinct subgroups of soccer players and may provide valuable indicators for evaluating players' match-specific running performance and optimizing training programs.

Acknowledgements

The authors thank Bart Vromans and Nicole Koopman-Verbaarschot for their great assistance with the data collection. We are thankful to Anita Loomans and Anna TopSupport for allowing us to utilize their exercise testing room, treadmill ergometer, and heart rate monitor during the testing week. Finally, we thank all the physicians that supervised these measurements. We express our gratitude to the soccer club where the present study was conducted for allowing our testing procedure to be a part of the yearly medical examination and their willingness to share the match data of their players. The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy rules that are in place at the professional soccer club in question. However, anonymous data are available upon request from the corresponding author: Richard Jaspers, r.t.jaspers@vu.nl. This work was supported by an NWO-ZonMW National Sportinnovator Grant (grant number: 538001045). Authors report no conflict of interest. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

References

Bangsbo, J. (1994) Energy demands in competitive soccer. *Journal of Sports Sciences* **12(sup1)**, 5-12. https://doi.org/10.1080/02640414.1994.12059272

Bangsbo, J. (2014) Physiological demands of football. *Scandinavian Sports Exercise* **27(125)**, 1-6.

Bangsbo, J. and Michalsik, L. (2002) *Assessment of the physiological capacity of elite soccer players* (1st ed.).

Bangsbo, J., Nørregaard, L. and Thorsø, F. (1991) Activity profile of competition soccer. *Journal of Sports Sciences* **16(2)**, 110-116.

Bittencourt, N. F., Meeuwisse, W., Mendonça, L., Nettel-Aguirre, A., Ocarino, J. and Fonseca, S. (2016) Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *British Journal of Sports Medicine* **50(21)**, 1309-1314. https://doi.org/10.1136/bjsports-2015-095850

Bloomfield, J., Polman, R. and O'Donoghue, P. (2007) Physical demands of different positions in FA Premier League soccer. *Journal of Sports Science and Medicine* **6(1)**, 63-70. https://pubmed.ncbi.nlm.nih.gov/24149226

Bradley, P. S., Sheldon, W., Wooster, B., Olsen, P., Boanas, P. and Krustrup, P. (2009) High-intensity running in English FA Premier League soccer matches. *Journal of Sports Sciences* **27(2)**, 159-168. https://doi.org/10.1080/02640410802512775

Buchheit, M., Simpson, B. M., Hader, K. and Lacome, M. (2021) Occurrences of near-to-maximal speed-running bouts in elite soccer: insights for training prescription and injury mitigation. *Science and Medicine in Football* **5(2)**, 105-110. https://doi.org/10.1080/24733938.2020.1802058

Campos-Vazquez, M. A., Mendez-Villanueva, A., Gonzalez-Jurado, J. A., León-Prados, J. A., Santalla, A. and Suarez-Arrones, L. (2015) Relationships between rating-of-perceived-exertion-and heart-rate-derived internal training load in professional soccer players: a comparison of on-field integrated training sessions. *International Journal of Sports Physiology and Performance* **10(5)**, 587-592. https://doi.org/10.1123/ijspp.2014-0294

Chamari, K., Hachana, Y., Ahmed, Y., Galy, O., Sghaier, F., Chatard, J., Hue, O. and Wisloff, U. (2004) Field and laboratory testing in young elite soccer players. *British Journal of Sports Medicine* **38(2)**, 191. https://doi.org/10.1136/bjsm.2002.004374

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.). Erbaum Press, Hillsdale, NJ, USA.

Dalmaijer, E. S., Nord, C. L. and Astle, D. E. (2022) Statistical power for cluster analysis. *BMC Bioinformatics* **23(1)**, 205. https://doi.org/10.1186/s12859-022-04675-1

Dellal, A., Chamari, K., Wong, D. P., Ahmaidi, S., Keller, D., Barros, R., Bisciotti, G. N. and Carling, C. (2011) Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European Journal of Sport Science* **11(1)**, 51-59. https://doi.org/10.1080/17461391.2010.481334

Deprez, D., Fransen, J., Boone, J., Lenoir, M., Philippaerts, R. and Vaeyens, R. (2015) Characteristics of high-level youth soccer players: variation by playing position. *Journal of Sports Sciences* **33(3)**, 243-254. https://doi.org/10.1080/02640414.2014.934707

Di Salvo, V., Gregson, W., Atkinson, G., Tordoff, P. and Drust, B. (2009) Analysis of high intensity activity in Premier League soccer. *International Journal of Sports Medicine*, 205-212. https://doi.org/10.1055/s-0028-1105950

Ekstrand, J., Waldén, M. and Hägglund, M. (2016) Hamstring injuries have increased by 4% annually in men's professional football, since 2001: a 13-year longitudinal analysis of the UEFA Elite Club injury study. *British Journal of Sports Medicine* **50(12)**, 731-737. https://doi.org/10.1136/bjsports-2015-095359

Filter, A., Olivares-Jabalera, J., Dos'Santos, T., Madruga, M., Lozano, J., Molina, A., Santalla, A., Requena, B. and Loturco, I. (2023) High-intensity actions in elite soccer: current status and future perspectives. *International Journal of Sports Medicine* **44(8)**, 535-544. https://doi.org/10.1055/a-2013-1661

Frencken, W. G., Lemmink, K. A. and Delleman, N. J. (2010) Soccer-specific accuracy and validity of the local position measurement (LPM) system. *Journal of Science and Medicine in Sport* **13(6)**, 641-645. https://doi.org/10.1016/j.jsams.2010.04.003

Fuhre, J., Øygard, A. and Sæther, S. A. (2022) Coaches' criteria for talent identification of youth male soccer players. *Sports* **10(2)**, 14. https://doi.org/10.3390/sports10020014

Gabbett, T. J. (2016) The training - injury prevention paradox: should athletes be training smarter and harder? *British Journal of Sports Medicine* **50(5)**, 273-280. https://doi.org/10.1136/bjsports-2015-095788

Gabbett, T. J. (2020) Debunking the myths about training load, injury and performance: Empirical evidence, hot topics and recommendations for practitioners. *British Journal of Sports Medicine* **54(1)**, 58-66. https://doi.org/10.1136/bjsports-2018-099784

Gil, S. M., Gil, J., Ruiz, F., Irazusta, A. and Irazusta, J. (2007) Physiological and anthropometric characteristics of young soccer players according to their playing position: Relevance for the selection process. *Journal of Strength and Conditioning Research* **21(2)**, 438-445. https://doi.org/10.1519/R-19995.1

Gualtieri, A., Rampinini, E., Dello Iacono, A. and Beato, M. (2023) High-speed running and sprinting in professional adult soccer: Current thresholds definition, match demands and training strategies. A systematic review. *Frontiers in Sports and Active Living* **5**, 1116293. https://doi.org/10.3389/fspor.2023.1116293

Hartigan, J. A. and Wong, M. A. (1979) Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society* **28(1)**, 100-108. https://doi.org/10.2307/2346830

Haugen, T. A., Tønnessen, E. and Seiler, S. (2013) Anaerobic performance testing of professional soccer players 1995-2010. *International Journal of Sports Physiology and Performance* **8(2)**, 148-156. https://doi.org/10.1123/ijspp.8.2.148

Helgerud, J., Engen, L. C., Wisløff, U. and Hoff, J. (2001) Aerobic endurance training improves soccer performance. *Medicine and Science in Sports and Exercise* **33(11)**, 1925-1931. https://doi.org/10.1097/00005768-200111000-00019

Jaspers, A., De Beéck, T. O., Brink, M. S., Frencken, W. G., Staes, F., Davis, J. J. and Helsen, W. F. (2018) Relationships between the external and internal training load in professional soccer: What can we learn from machine learning? *International Journal of Sports Physiology and Performance* **13(5)**, 625-630. https://doi.org/10.1123/ijspp.2017-0299

Julian, R., Page, R. M. and Harper, L. D. (2021) The effect of fixture congestion on performance during professional male soccer match-play: A systematic critical review with meta-analysis. *Sports Medicine* **51**, 255-273. https://doi.org/10.1007/s40279-020-01359-9

Kavanagh, R., Di Michele, R., Oliveira, R., McDaid, K., Rhodes, D. and Morgans, R. (2024) The relationships between distances covered above generic and relative speed thresholds by male soccer players in English Premier League matches across two competitive seasons. The effects of positional demands and possession. *Biology of Sport* **41(4)**, 77-86. https://doi.org/10.5114/biolsport.2024.135416

Kemi, O., Hoff, J., Engen, L., Helgerud, J. and Wisloff, U. (2003) Soccer specific testing of maximal oxygen uptake. *Journal of Sports Medicine and Physical Fitness* **43(2)**, 139.

Lago-Peñas, C., Casais, L., Dellal, A., Rey, E. and Domínguez, E. (2011) Anthropometric and physiological characteristics of young soccer players according to their playing positions: Relevance for competition success. *Journal of Strength and Conditioning Research* **25(12)**, 3358-3367. https://doi.org/10.1519/JSC.0b013e318216305d

McCann, D. J. and Adams, W. C. (2002) A theory for normalizing resting VO_2 for differences in body size. *Medicine and Science in Sports and Exercise* **34(8)**, 1382-1390. https://doi.org/10.1097/00005768-200208000-00022

Metaxas, T. I. (2021) Match running performance of elite soccer players: $\dot{V}O_2$max and players position influences. *Journal of Strength and Conditioning Research* **35(1)**, 162-168. https://doi.org/10.1519/JSC.0000000000002646

Modric, T., Jelicic, M. and Sekulic, D. (2021a) Relative training load and match outcome: Are professional soccer players actually undertrained during the in-season? *Sports* **9(10)**, 139. https://doi.org/10.3390/sports9100139

Modric, T., Versic, S. and Sekulic, D. (2021b) Does aerobic performance define match running performance among professional soccer players? A position-specific analysis. *Research in Sports Medicine* **29(4)**, 336-348. https://doi.org/10.1080/15438627.2021.1888107

Modric, T., Versic, S., Sekulic, D. and Liposek, S. (2019) Analysis of the association between running performance and game performance indicators in professional soccer players. *International Journal of Environmental Research and Public Health* **16(20)**, 4032. https://doi.org/10.3390/ijerph16204032

Mohr, M., Krustrup, P. and Bangsbo, J. (2003) Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sports Sciences* **21(7)**, 519-528. https://doi.org/10.1080/0264041031000071182

Nikolaidis, P. T., Knechtle, B., Clemente, F. and Torres-Luque, G. (2016a) Reference values for the sprint performance in male football players aged from 9-35 years. *Biomedical Human Kinetics* **8(1)**, 103-112. https://doi.org/10.1515/bhk-2016-0015

Nikolaidis, P. T., Ruano, M. A. G., de Oliveira, N. C., Portes, L. A., Freiwald, J., Leprêtre, P. M. and Knechtle, B. (2016b) Who runs the fastest? Anthropometric and physiological correlates of 20 m sprint performance in male soccer players. *Research in Sports Medicine* **24(4)**, 341-351. https://doi.org/10.1080/15438627.2016.1222281

Novack, L. F., Nascimento, V. B., Salgueirosa, F. M., Carignano, L. F., Fornaziero, A., Gomes, E. B. and Osiecki, R. (2013) Subgroup distribution based on physiological responses in professional soccer players by K-means cluster technique. *Revista Brasileira de Medicina do Esporte* **19**, 130-133. https://doi.org/10.1590/S1517-86922013000200012

Ogris, G., Leser, R., Horsak, B., Kornfeind, P., Heller, M. and Baca, A. (2012) Accuracy of the LPM tracking system considering dynamic position changes. *Journal of Sports Sciences* **30(14)**, 1503-1511. https://doi.org/10.1080/02640414.2012.712712

Pillitteri, G., Petrigna, L., Ficarra, S., Giustino, V., Thomas, E., Rossi, A., Clemente, F. M., Paoli, A., Petrucci, M. and Bellafiore, M. (2024) Relationship between external and internal load indicators and injury using machine learning in professional soccer: A systematic review and meta-analysis. *Research in Sports Medicine* **1-37**. https://doi.org/10.1080/15438627.2023.2297190

Rhini, M., Hickner, R., Naidoo, R. and Sookan, T. (2024) The physical demands of the match according to playing positions in a South African Premier Soccer League team. *South African Journal of Sports Medicine* **36(1)**. https://doi.org/10.17159/2078-516X/2024/v36i1a16752

Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F. and Baca, A. (2023) Machine learning application in soccer: A systematic review. *Biology of Sport* **40(1)**, 249-263. https://doi.org/10.5114/biolsport.2023.112970

Silva, J. R., Brito, J., Akenhead, R. and Nassis, G. P. (2016) The transition period in soccer: A window of opportunity. *Journal of Sports Medicine* **46**, 305-313. https://doi.org/10.1007/s40279-015-0419-3

Slimani, M., Znazen, H., Miarka, B. and Bragazzi, N. L. (2019) Maximum oxygen uptake of male soccer players according to their competitive level, playing position and age group: Implication from a network meta-analysis. *Journal of Human Kinetics* **66**, 233. https://doi.org/10.2478/hukin-2018-0060

Smpokos, E., Rikos, N., Manolarakis, G., Mourikis, C. and Linardakis, M. (2025) Physical match running performance indicators in professional Greek footballers competing in a national league: A two-consecutive-period survey (2021/22 and 2022/23). *International Journal of Novel Research in Interdisciplinary Studies* **12(3)**, 7-17.

Soós, I., Borysławski, K., Boraczyński, M., Ihasz, F. and Podstawski, R. (2022) Anthropometric and physiological profiles of Hungarian youth male soccer players of varying ages and playing positions: A multidimensional assessment with a critical approach. *International Journal of Environmental Research and Public Health* **19(17)**. https://doi.org/10.3390/ijerph191711041

Stølen, T., Chamari, K., Castagna, C. and Wisløff, U. (2005) Physiology of soccer: An update. *Sports Medicine* **35**, 501-536. https://doi.org/10.2165/00007256-200535060-00004

Towlson, C., Salter, J., Ade, J. D., Enright, K., Harper, L. D., Page, R. M. and Malone, J. J. (2021) Maturity-associated considerations for training load, injury risk, and physical performance in youth soccer: One size does not fit all. *Journal of Sport and Health Science* **10(4)**, 403-412. https://doi.org/10.1016/j.jshs.2020.09.003

van der Zwaard, S., Brocherie, F. and Jaspers, R. T. (2021) Under the hood: Skeletal muscle determinants of endurance performance. *Frontiers in Sports and Active Living* **3**, 719434. https://doi.org/10.3389/fspor.2021.719434

van der Zwaard, S., de Ruiter, C. J., Jaspers, R. T. and de Koning, J. J. (2019) Anthropometric clusters of competitive cyclists and their sprint and endurance performance. *Frontiers in Physiology* **10**, 1276. https://doi.org/10.3389/fphys.2019.01276

van der Zwaard, S., de Ruiter, C. J., Noordhof, D. A., Sterrenburg, R., Bloemers, F. W., de Koning, J. J., Jaspers, R. T. and van der Laarse, W. J. (2016) Maximal oxygen uptake is proportional to muscle fiber oxidative capacity, from chronic heart failure patients to professional cyclists. *Journal of Applied Physiology* **121(3)**, 636-645. https://doi.org/10.1152/japplphysiol.00355.2016

van der Zwaard, S., van der Laarse, W. J., Weide, G., Bloemers, F. W., Hofmijster, M. J., Levels, K., Noordhof, D. A., de Koning, J. J., de Ruiter, C. J. and Jaspers, R. T. (2018a) Critical determinants of combined sprint and endurance performance: An integrative analysis from muscle fiber to the human body. *FASEB Journal* **32(4)**, 2110-2123. https://doi.org/10.1096/fj.201700827R

van der Zwaard, S., Weide, G., Levels, K., Eikelboom, M. R. I., Noordhof, D. A., Hofmijster, M. J., van der Laarse, W. J., de Koning, J. J., de Ruiter, C. J. and Jaspers, R. T. (2018b) Muscle morphology of the vastus lateralis is strongly related to ergometer performance, sprint capacity and endurance capacity in Olympic rowers. *Journal of Sports Sciences* **36(18)**, 2111-2120. https://doi.org/10.1080/02640414.2018.1439434

Vieira, L. H. P., Christopher, C., Barbieri, F. A., Aquino, R. and Santiago, P. R. P. (2019) Match running performance in young soccer players: A systematic review. *Journal of Sports Medicine* **49**, 289-318. https://doi.org/10.1007/s40279-018-01048-8

Wei, X., Zhang, R., Maitiniyazi, A., Chmura, P., Randers, M. B. and Krustrup, P. (2024) Running performance of substitutes in the FIFA World Cup after the change in substitution rule: Influence of match outcome, playing position and sex. *International Journal of Performance Analysis in Sport* **24(6)**, 666-680. https://doi.org/10.1080/24748668.2024.2358268

Wing, C. E., Turner, A. N. and Bishop, C. J. (2020) Importance of strength and power on key performance indicators in elite youth soccer. *Journal of Strength and Conditioning Research* **34(7)**, 2006-2014. https://doi.org/10.1519/JSC.0000000000002446

Wisloeff, U., Helgerud, J. and Hoff, J. (1998) Strength and endurance of elite soccer players. *Medicine and Science in Sports and Exercise* **30(3)**, 462-467. https://doi.org/10.1097/00005768-199803000-00019

Key points

- There is considerable interindividual variation in match-specific running performance within positional groups.
- Studying the physical capacities, designing training programs or evaluating match-specific running performance of soccer players based on their playing positions is suboptimal.
- Particularly grouping by forwards, midfielders and defenders should be avoided when evaluating match-specific running performance
- Identifying subgroups based on match-specific running performance using clustering analysis seems a promising alternative for categorizing soccer players.

AUTHOR BIOGRAPHY

Michel de HAAN
Employment
Department of Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, Netherlands
Degree
MSc, PhD student
Research interests
Exercise physiology, Muscle physiology, Sports science, Data science
E-mail: m.h.de.haan@vu.nl

Stephan VAN DER ZWAARD
Employment
Department of Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, Netherlands
Department of Cardiology, Amsterdam University Medical Center, location AMC, University of Amsterdam, Amsterdam, Netherlands
Degree
Assistant professor
Research interests
Sports science, Exercise physiology, Data science
E-mail: s.vander.zwaard@vu.nl

Jurrit SANDERS
Employment
PSV Eindhoven, Eindhoven, Netherlands
Degree
MSc
Research interests
Soccer, Applied physiology, Data science, Talent development
E-mail: j.sanders@psv.nl

Peter BEEK
Employment
Department of Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, Netherlands
Degree
Full professor
Research interests
Coordination dynamics, Perceptual-motor skills, Sports science, Rehabilitation
E-mail: p.j.beek@vu.nl

Richard JASPERS
Employment
Department of Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, Netherlands
Degree
Full professor
Research interests
Exercise physiology, Muscle physiology, Molecular physiology, Sports science
E-mail: r.t.jaspers@vu.nl

✉ **Richard Jaspers**
Vrije Universiteit Amsterdam, De Boelelaan 1108, 1081 HZ Amsterdam, Netherlands