Classifying Player Profiles in Elite Women's Football: A K-Means Clustering Analysis of Physical and Technical Data from the 2023 FIFA Women's World Cup

SHEN, Q, DING, Junyuan, SUMMERS., H.D., KUBAYI, A., STRAFFORD, Ben <http://orcid.org/0000-0003-4506-9370> and STONE, Joseph <http://orcid.org/0000-0002-9861-4443>

Citation:

SHEN, Q, DING, Junyuan, SUMMERS., H.D., KUBAYI, A., STRAFFORD, Ben and STONE, Joseph (2026). Classifying Player Profiles in Elite Women's Football: A K-Means Clustering Analysis of Physical and Technical Data from the 2023 FIFA Women's World Cup. Football Studies: 100026. [Article]

Classifying Player Profiles in Elite Women's Football: A K-Means Clustering Analysis of Physical and Technical Data from the 2023 FIFA Women's World Cup

No sources of funding from any funding agency in the public, commercial, or not for profit sectors were used to assist in the preparation of this article. We have no known conflict of interests.

Abstract

As elite women's football becomes increasingly physical and technically demanding, practitioners require objective, scalable methods to profile player behaviours and match-specific demands to inform coaching, recruitment, and training design. However, integrated clustering of physical tracking outputs and technical event indicators remains underused in this context. An exploratory, data-driven approach was used to classify player behaviours based on integrated physical tracking and offensive technical event data from the 2023 FIFA Women's World Cup. The full dataset comprised 1,885 player-match performances; after applying the study inclusion criteria, 1,599 player-match records from 539 players were retained for analysis. The K-means clustering analysis applied to the combined physical and technical dataset identified eighteen player clusters, reflecting distinct offensive behaviours operationalised as joint patterns of technical execution and physical output. The Principal Component Analysis (PCA) projection was used to provide a low-dimensional visual summary of cluster separability. Most clusters appeared well separated (77.2%), whereas a subset were positioned in close proximity with partial overlap of observations in the projected space. Independently of the PCA visualisation, longitudinal cluster membership indicated that 184 players showed stable single-cluster assignment across matches (specialised profiles), while the remaining players were assigned to two or more clusters across matches (hybrid profiles). These results suggest the relevance of integrating technical and physical data and the value of unsupervised learning approaches in capturing the diversity of player behaviours within women's football. The findings offer insights for coaching, recruitment, and training design by identifying nuanced player profiles, such as high-intensity forwards, aerial target strikers, and impact substitutes.

Introduction

In recent times, the availability of large-scale data sets in football combined with contemporary data analysis methods has led to more advanced player performance evaluations (Memmert & Raabe, 2023). Metrics such as physical outputs (e.g., sprints, high-speed running, distance covered) and technical actions (e.g., passes, dribbles, shots) are often collected to evaluate match performance (Jamil et al., 2021; Paul et al., 2015; Rodrigues & Pinto, 2022; Rodrigues et al., 2023). However, performance analysis methods often examine physical (e.g., Trewin et al., 2018; Vescovi & Falenchuk, 2019) and technical (e.g., Beare & Stone, 2019; Garcia- Unanue et al., 2020; Kubayi & Paul, 2020) indicators separately. This segmented approach does not reflect a holistic evaluation of football performance (Bradley & Ade, 2018).

In elite football match play, players are exposed to heterogeneous role demands that elicit distinct combinations of technical actions and physical output rather than uniform performance patterns (Forcher et al., 2022; Sarmento et al., 2024). For instance, some players are required to sustain high volumes of running and repeated high-intensity efforts over extended playing time, whereas others, such as late-match attacking substitutes, may accumulate relatively low total distance but are expected to produce a high density of explosive actions (e.g., sprints, accelerations) alongside direct attacking contributions such as shots and dribbles within a limited time window. Similarly, wide attacking players often combine frequent high-speed running and ball carrying to exploit space along the flanks, while central attacking players may exert influence primarily through passing and positional involvement with comparatively lower sprint demands. Empirical research has shown that both physical and technical match demands vary substantially according to playing role, position, and tactical context, supporting the notion that offensive performance is best represented by characteristic technical–physical configurations rather than isolated metrics

(Bradley et al., 2009; Di Salvo et al., 2007; Lago-Peñas et al., 2011; Forcher et al., 2022). Therefore, a combined physical–technical analytical approach is essential for uncovering meaningful patterns in player behaviour, because the interaction between physical output and technical execution often determines the effectiveness of match actions (Modric et al., 2022).

Currently, existing literature examining football performance focuses on domestic men's elite leagues (Kirkendall & Krustrup, 2021; Stone et al., 2025), with a relative lack of data-driven research targeting women's football (Harkness-Armstrong et al., 2022), particularly in international tournament settings (Bradley et al., 2010). Although research on women's football has increased in recent years, it still accounts for 20% of all football studies (Kirkendall & Krustrup, 2021), with a predominant emphasis on injury prevention, physical performance, psychosocial factors, motivation, and leadership (Ventaja-Cruz et al., 2024). Research on the profiles and playing styles of female footballers in terms of technique and physicality during matches remains limited. The FIFA Women's World Cup represents a unique high-performance environment characterised by distinct playing styles, condensed schedules, and international tactical variations (Garcia-Unanue et al., 2020; Oliva-Lozano et al., 2025). Therefore, a comprehensive analysis of World Cup performance data could yield valuable insights into the demands and success factors in international women's football games.

Clustering methods offer a data-driven means to uncover and classify complex features embedded within multivariate performance metrics (Cho et al., 2022; Herold et al., 2019). By applying statistical techniques and unsupervised algorithms, clustering can identify and categorise previously unlabelled patterns in team behaviours and performance characteristics without requiring pre-defined labels (Herold et al., 2019). This capability supports athlete performance optimisation and, when combined with visualisation and reporting, provides a structured route for generating and communicating insights relevant to football tactical research (Wang & You, 2021). Accordingly, clustering has become increasingly valuable in sports performance research, particularly for exploratory objectives where the goal is to characterise latent structures and heterogeneity within complex datasets rather than to evaluate a single pre-specified hypothesis (Teunissen et al., 2024). Practically, clustering enables the extraction of natural groupings that reflect the multidimensional nature of athletic performance and the joint contribution of multiple indicators (Carpita et al., 2023). Because performance indicators can combine and interact within match situations (Chang et al., 2024; Li et al., 2022), clustering provides a principled way to summarise these multivariate configurations into interpretable profiles. Interpretable features and decision rules can also be derived from match data, allowing behaviours to be represented in a visually accessible form that can inform strategy development and applied decision-making (Fujii, 2021). As such, clustering is well suited to exploratory profiling in sport, capturing emergent behavioural patterns without imposing rigid categorical assumptions. Beyond description, it can generate practitioner-relevant outputs by differentiating role-based profiles and informing training priorities, talent identification, and tactical preparation (Alberto et al., 2023; Serhat Emre Akhanl & Hennig, 2023).

Research has demonstrated the utility of clustering techniques in sport, providing a foundation for broader application in performance analysis (Herold et al., 2019). In football, clustering has been used to differentiate behavioural configurations, often incorporating positional and contextual information. For example, Sampaio et al. (2015) applied cluster analysis in small-sided games to distinguish behavioural patterns under different pitch dimensions, and Li et al. (2022) used K-means clustering in the Men’s Chinese Super League, highlighting how contextual information can shape cluster solutions. At the team level, Fernández-Navarro et al. (2016) categorised offensive and defensive patterns in elite Spanish and English teams and linked these patterns to match outcomes. Despite this growth in men's football, clustering applications remain comparatively underexplored in women's football (Harkness-Armstrong et al., 2022), particularly at the international tournament level.

Alongside this methodological development, there is increasing interest in how attacking behaviours relate to physical output. Prior work suggests that different attacking behaviours (e.g., more direct play, possession-oriented circulation, or pressing-oriented approaches) are associated with distinct physical demands and functional contributions (Forcher et al., 2022; Izquierdo et al., 2023). Importantly, these behaviours are expressed through the joint contribution of technical actions and physical output, motivating integrated physical–technical analyses (González-Rodenas et al., 2020; Sarmento et al., 2018). While tactical "playing style" inference typically requires positional tracking and richer contextual variables, integrated physical and technical match metrics can still support the identification of interpretable, performance-based groupings relevant for applied decision-making. For example, Narayanan and Pifer (2023) developed a tailored model to characterise performance patterns in the United States Women's National Team to inform training and match preparation in women's football. Accordingly, understanding such performance-based profiles can support coaching decisions in talent identification, training focus, and match preparation (Fernandez-Navarro et al., 2016; Sarmento et al., 2017), and help align offensive contributions with appropriate physical capacities (Altmann et al., 2021).

In elite women's football, offensive contributions are expressed through the combined patterning of technical actions and physical output. Integrating these dimensions provides a practical basis for describing heterogeneity in how players contribute offensively, beyond what can be inferred from either dimension in isolation. Accordingly, this study adopts an exploratory, data-driven approach to derive offensive player profiles from integrated physical tracking and technical event metrics from the 2023 FIFA Women's World Cup using K-means clustering. Specifically, we aim to: (i) construct player–match observations combining physical outputs (e.g., high-speed running and sprints) and technical indicators (e.g., key passes and dribbles); (ii) apply K-means clustering to identify natural groupings in this combined feature space; and (iii) characterise the resulting profiles in terms of their distinguishing physical–technical features.

Methods

Match Sample

A purposive sample was used, comprising all eligible player–match observations from the 2023 FIFA Women's World Cup. This sampling strategy was selected a priori to align with the exploratory aim of deriving performance-based physical–technical profiles within an elite international tournament context, rather than to estimate population parameters through probabilistic sampling. The sample consisted of all 64 games, including 62, 90-minutes games and 2, 120-minutes games, which included 32 teams from the 2023 Women's World Cup. A total of 1,885 player–match record from 622 players were included (i.e., each observation corresponds to one player's aggregated physical-tracking and technical event metrics within a single match). Player-match records with ≥30 minutes of playing time were retained to avoid low-involvement bias by ensuring a minimum level of match exposure (Chang et al., 2024), and all variables were standardised to per-90-minute or per-120-minute (if extra time is applied) values to ensure comparability. Finally, 1,599 player-match records from 539 players were selected as the sample of study. The technical data was collected from the StatsBomb open-access FIFA Women's World Cup dataset. The physical data was collected from the Post Match Report from the FIFA Training Centre. The study was approved by the local [*Removed university name for anonymised for Peer Review*] University ethics committee [*ID number-Anonymised for Peer-Review*].

Data Collection and Integration

This study included both physical and technical performance data from the 2023 FIFA Women's World Cup to profile players' offensive behaviour in relation to their physical output. Physical performance data were manually extracted from individual Post-Match Summary Reports published by the FIFA Training Centre (FIFA, 2023), while technical event data were retrieved using the StatsBombR package using R-Studio from the publicly available StatsBomb Open Data platform. The FIFA reports were manually transcribed from portable document format (pdf) files into Excel (Microsoft, USA) format, to extract relevant metrics, including total distance (meters), distance across speed zones (meters), sprint counts, and maximum speed (km/h) were extracted for all matches. To ensure data accuracy and minimise transcription errors, a multi-step quality control procedure was applied. Specifically, two researchers independently extracted all physical performance variables from the original reports, and a third researcher subsequently cross-checked the extracted data against the source documents. Any discrepancies were discussed and resolved by consensus. This procedure was applied to all matches included in the study to ensure consistency between the extracted dataset and the original FIFA reports. In parallel, the StatsBomb JavaScript Object Notation (JSON) event data provided information on passes, shots, crosses, dribbles, and carries. Definitions are outlined in Table 1.

Insert Table 1 near here

The two datasets were merged by manually resolving discrepancies in player name formats (e.g., *LASTNAME Firstname* vs. *Firstname Middlename Lastname*) and appending missing match identifiers allowing comprehensive integration of all match records into a single dataset. The final dataset comprised 18 variables, including nine physical and nine offensive technical indicators (see Table 1). In terms of accuracy, all 64 games during the FIFA Women's World Cup 2023 were analysed using a multi-camera optical tracking system (TRACAB, ChyronHego, Sweden). All player movements were captured by high-definition cameras operating at 25Hz. The validity and accuracy of data collection by TRACAB Gen5 were provided by FIFA (FIFA, n.d.) and Linke et al (Linke et al., 2020). StatsBomb data were semi-automated and internally validated (see Hudl (2025) for data quality process); however, certain elements, such as key passes, may involve subjective judgement.

Statistical Analysis

A multi-step unsupervised learning workflow was applied to classify offensive playing profiles. First, numeric variables were scaled to [0,1] via min–max normalisation to ensure comparability across metrics of different scales and to prevent large-value features from dominating the clustering (Wongoutong, 2024). This normalisation also facilitated subsequent radar-chart visualisations by keeping all dimensions on a consistent scale. K-means clustering was then chosen for its effectiveness in discovering hidden patterns in unlabelled data (Jain, 2010; Everitt et al., 2011) and for minimising within-cluster variance with normalised inputs (Everitt, 1980). It assigns each data point to the nearest centroid by optimising the within-cluster sum-of-squares objective. This approach aligns with the study's goal of identifying cohesive behavioural patterns and has been demonstrated effective in sports science contexts (Sampaio et al., 2015).

The gap statistic (Tibshirani et al., 2001) is used, which is a method that compares the change in within-cluster dispersion with that expected under a null reference distribution, we selected the smallest number of clusters within one standard error of the maximum gap value, resulting in an 18-cluster solution. This choice provides a pragmatic balance between capturing meaningful structure in the data and maintaining interpretability. K-means clustering was run with 25 random initialisations, and cluster centres were rescaled to original units. A representative player closest to each centroid (by Euclidean distance) was selected to exemplify that cluster's characteristics, with both standardised and raw centroid values tabulated to provide a transparent numerical summary of each group (Fränti et al., 2014).

Collinearity among clustering features was assessed using pairwise Spearman correlations (Supplementary Figure S1). Expected correlations were observed among physical workload variables (e.g., total distance, speed-zone distances, high-speed runs, and sprints), reflecting shared underlying constructs. We retained these variables to preserve construct coverage across intensity bands and normalised all features prior to distance-based clustering to reduce scale effects and mitigate undue influence of any single variable. Highly correlated feature pairs ($|\rho| \geq 0.8$) are summarised in Appendix Table A1. To interpret and present the clustering results, three key visualisation methods were employed:

1. PCA-based cluster membership plot: Clusters were visualised in a 2D principal component space using the fviz_cluster function (Greenacre et al., 2022), illustrating how groups are distributed and separated in a reduced-dimensional view.
2. Quartile-based heatmap: Each cluster's centroid values were categorised as "Low," "Below Average," "Above Average," or "High" based on first, second, and third quartiles. These labels were plotted as a colour-gradient heatmap, which intuitively highlights the metrics in which each cluster excels providing an overview of player type distributions.
3. Radar charts: For each cluster's representative player, a radar chart was generated (using all 18 variables scaled 0–1) and annotated with the player's name, position, team, and match ID.

Interpretation of Cluster Membership Variability

Clustering was performed at the player–match observation level, whereby each individual match performance was treated as an independent analytical unit. Consequently, the same player could be assigned to different clusters across matches. Within this framework, overlap does not refer to insufficient separation between clusters, but rather to variability in a player's offensive performance profile across distinct match contexts.

Consistent cluster membership across matches was interpreted as indicative of a relatively stable expression of offensive behaviour, whereas dispersion of a player's match observations across multiple clusters was interpreted as reflecting adaptable or context-dependent offensive behaviour. Such variability was considered meaningful when cluster assignments corresponded to clearly differentiated combinations of physical output and technical actions, as evidenced by centroid profiles and representative player visualisations. This interpretation is consistent with an exploratory, behaviour-oriented perspective of football performance, which conceptualises player contributions as dynamic and situational rather than fixed individual traits.

Results

The K-means clustering analysis applied to physical and technical performance data produced eighteen player clusters, reflecting unique offensive profiles and physical demands. Eighteen representative clusters are highlighted to illustrate practical applications in talent identification and role profiling. For the selected k-means clustering solution (n = 1,599; p = 18), BetweenSS/TotSS was 0.721, indicating that a substantial proportion of the total variance was accounted for by between-cluster differences. The Calinski–Harabasz index was 240.21 and the Davies–Bouldin index was 1.78, suggesting moderate compactness and separation, with some overlap between clusters.

Cluster Membership Plot

Cluster distribution is visualised using a 2D principal component projection (see Figure 1). The plot reveals the spatial configuration of 18 clusters based on PCA reduction of the scaled performance data. Several clusters (for example, 1 and 17) form relatively compact groupings that are clearly separated from other clusters, suggesting low within-cluster variability and a coherent feature profile. By contrast, clusters 6 and 9 appear proximate with substantial intermingling in the PC1–PC2 plane, indicating limited separability along the dominant variance directions captured by these components. This pattern is consistent with clusters 6 and 9 representing closely related profiles. Clusters 6 and 9 show overlapping boundaries, suggesting more generalist or a transitional region along a continuum of playing styles, rather than two sharply distinct style categories. This structural layout aligns with the multifactorial nature of football performance, where physical and technical metrics are jointly expressed in match performance and can be summarised as co-occurring patterns at the player–match level. PC1 and PC2 explained 49.61% and 15.17% of the variance (64.78% cumulative), respectively. The highest absolute loadings on PC1 were total distance, high speed runs, Zone 2, playing time, Zone 3, whereas PC2 was primarily driven by playing time, sprints, Zone 1, Zone 4 and top speed.

Insert Figure 1 near here

Cluster Centroid Tables

The numerical characteristics of each cluster were summarised using both standardised and unstandardised centroids. The standardised centroid table (see Table 2) shows z-scored values across all 18-performance metrics.

Insert Table 2 near here

The unstandardised centroid table (see Table 3) restored the data to the original units, with Clusters 1 and 13 displaying high values in total distance (exceeding 9000 metres) and sprint counts, suggesting physically demanding, likely wide attackers or box-to-box midfielders. Whereas Cluster 4 shows low engagement across all metrics, potentially reflecting the minimal-impact players, for example, late substitutes, injured player after short time of play.

Insert Table 3 near here

Quartile-Based Heatmap

To facilitate interpretation, each cluster's values were categorised into quartiles (Q1 to Q4) and visualised as a heatmap (see Figure 2). Notably, Cluster 2 registers "Low" across nearly all variables, implying an outlier group with minimal physical or technical contribution. This visually signals a goalkeeper or non-involved role, which will be further reflected through representative profiles.

Insert Figure 2 near here

Representative Player Profiles

Each cluster's most prototypical player, defined as the observation with the minimum Euclidean distance to the centroid, was extracted and summarised in Table 4.

Insert Table 4 near here

These representative cases serve to humanise the data-driven groupings and allow for applied interpretation. This observation directly reinforces the interpretation drawn from Figure 2, where Cluster 2 was flagged as a goalkeeper. The alignment between these two forms of evidence suggests that the clustering structure effectively distinguishes not just stylistic profiles– but also functional archetypes within the team.

Radar Chart Profiles of Selected Clusters and Summary of Profiles

Radar charts visually represented the distinct performance profiles of selected clusters, enabling intuitive interpretation and clear differentiation of player profiles (Figure 3).

Insert Figure 3 near here

Discussion

This study identified and characterised distinct offensive performance profiles among elite women's football players by integrating technical and physical indicators from the 2023 FIFA Women's World Cup using K-means clustering. The resulting clusters should be interpreted as descriptive, tournament-specific profiles rather than generalisable player typologies beyond this context.

The gap statistic supported an 18-cluster solution, and the PCA-based membership plot indicated varying degrees of separation and overlap, consistent with both specialised and hybrid profiles. Although the statistical criterion favoured 18 clusters, several clusters were conceptually adjacent. Accordingly, the cluster solution is interpreted at a thematic level to provide a pragmatic balance between interpretability and model complexity while retaining sufficient granularity to distinguish meaningful performance profiles. The visual profiles (i.e., figure 3 radar charts) provide an accessible multivariate summary for non-technical stakeholders such as coaches and scouts (Yeung et al., 2024). Practically, these profiles can inform role-specific training design, tactical planning, and recruitment by linking observable match behaviours to physical and technical demands (Sampaio et al., 2015; Sarmento et al., 2014).

The quartile heatmap (figure 2) highlighted a clear "explosive winger" signature in Cluster 13, characterised by consistently high sprinting/high-speed running and dribbling-related metrics. This pattern aligns with evidence that wide players typically accumulate high volumes of high-intensity accelerations/decelerations and high-speed running in elite competition, reflecting repeated explosive efforts in wide channels (Bortnik et al., 2024; Bujnovsky et al., 2019; Izquierdo et al., 2023). Accordingly, conditioning for this archetype should prioritise repeated-sprint ability and speed-endurance, integrated with dribbling actions under fatigue to reflect match-specific constraints. These characteristics were also evident in the radar charts, reinforcing this player type. Collectively, the findings underline the importance of speed endurance and repeated sprint ability for this archetype in the women's game. Conditioning could be tailored to the high running demands of wide players by developing robust sprint capacity and the ability to repeat high-intensity efforts with limited fatigue. Hence, training that replicates match scenarios, such as repeated sprint bouts interspersed with short recovery periods, may be particularly relevant for preparing wingers for their role (Bujnovsky et al., 2019).

Cluster 1 exhibited consistently high performance across key offensive and physical indicators, most notably sprinting speed, shooting output, and high-intensity running, characterising a “complete attacking forward” profile. This integrated technical–physical signature is consistent with prior evidence that elite forwards demonstrate superior short-distance sprint performance and explosive power relative to players in other outfield positions (Boone et al., 2012), alongside strong finishing capability reflected in higher shooting accuracy observed even at youth levels (Kozina et al., 2023). In match-play, forwards also tend to incur substantial high-intensity demands during high-pressure phases, including greater sprint and acceleration outputs (Bortnik et al., 2024), and their somatotype characteristics may further support explosive actions (Kolena et al., 2024). Collectively, these attributes help explain why Cluster 1 aligns with central attacking roles that require the combination of decisive technical execution and intense physical output, consistent with the radar profile observed for Player 1. Conversely, Cluster 2 showed uniformly low technical and physical metrics, highlighting goalkeepers’ distinct positional demands (Casal et al., 2023). This profile reinforces the specificity required in goalkeeper training, conditioning, and selection processes, consistent with previous research emphasising role-specific preparations (Lethole et al., 2024).

Cluster 18 reflected a distinct second striker profile, with high shooting involvement, particularly headed shots, alongside moderate physical contributions. Compared with other clusters, these players were more directly associated with end-product actions such as shots and headed shots, but showed lower involvement in ball progression and creation, including dribbles, carries, and headed passes. This pattern is consistent with operating primarily in the opponent’s penalty area, often converting receptions, especially from crosses, into immediate attempts on goal. Lower values for total distance covered and Zone 5 running further indicate an emphasis on scoring rather than sustained defensive or off-ball work such as pressing, tackling, and duels (Riezebos, 2021). This aligns with prior evidence that strikers generally cover shorter distances and experience lower running loads than other outfield roles (Bloomfield et al., 2007; Marcelli et al., 2024). Overall, the profile accords with the radar pattern observed for Player 17 and supports training emphases on penalty-area movement, balance, spatial positioning, and finishing under pressure.

An interesting finding emerged from the unstandardised centroid table, indicating that average running distances and moderate-speed activity in elite women's football are comparable to those reported in elite men's football (Bradley et al., 2013). This challenges some previous assumptions suggesting women’s football inherently involves lower overall physical match demands (Mohr et al., 2008). However, notable differences were observed in metrics involving sprints, high-speed running, and technical data such as dribbles, passes, and shot accuracy (Garnica-Caparrós & Memmert, 2021). Lower averages in high-speed metrics and technical execution emphasise areas for targeted improvement, reinforcing the need for training interventions focused on developing explosive capabilities and technical precision under fatigue conditions (Krustrup et al., 2005).

The occurrence of Player 13 appearing twice in the closest-player table for Clusters 13 and 17 suggests redundancy: both cluster centroids were defined by Player 13 performance vector, implying little difference between these clusters reveals significant insights about player versatility and tactical adaptability. Elite players capable of altering playing behaviours based on match scenarios demonstrate high adaptability, critical for tactical flexibility and competitive advantage (Carling et al., 2008). This finding emphasises the value of cultivating versatile players capable of fulfilling multiple tactical profiles, significantly impacting training programmes and talent scouting strategies.

These findings underscore the value of integrating technical and physical data and demonstrate how unsupervised learning can capture the diversity of behaviours in elite women's football. Rather than reproducing rigid position-based categories, the clustering approach identified performance-driven role profiles with practical relevance. This is particularly important given that most comparable clustering work has focused on male cohorts, and the present study therefore contributes to the development of sex-specific performance models (Li et al., 2022; Fernández-Navarro et al., 2016; Narayanan & Pifer, 2023). From an applied perspective, the profiles provide actionable insights for coaching, recruitment, and training design by supporting evidence-based decisions about role-specific demands, for example high-intensity forwards, aerial finishers, and impact substitutes. More broadly, the approach offers a replicable framework for profiling offensive roles in international women's competitions, with potential to be extended across age groups and competitive levels, and, where appropriate, to other contexts including the men's game.

Limitations and Future Studies

Despite offering valuable insights given the exploratory nature of this study and the large number of observed outcomes, the findings should be interpreted cautiously (Wasserstein & Lazar, 2016). Furthermore, the analysis was confined to a single international tournament (the 2023 FIFA Women's World Cup), which may limit the generalisability of the identified clusters across broader contexts or other competition levels. Player–match performance observations were treated as the unit of analysis. This approach captures match-to-match variability but does not explicitly model repeated observations within players; future work could adopt longitudinal or multilevel approaches to account for within-player dependence. Phase-specific physical outputs (e.g., in- and out-of-possession) were not available; therefore, physical variables reflect whole-match loads rather than possession- or phase-conditioned demands. Although goalkeepers were retained to reflect their contemporary involvement in build-up play, goalkeeper-specific profiling and position- or phase-conditioned analyses should be prioritised in future work to improve role specificity and interpretability. Lastly, the manual extraction and integration of physical and technical data from disparate sources could introduce minor inconsistencies or human errors, despite careful handling.

In addition, several contextual factors that may influence match-specific offensive behaviour were not explicitly modelled, including match status (e.g., winning vs. losing), quality of opposition, match location and travel demands, accumulated fatigue across the tournament, and short-term injury or return-to-play status. As a result, some cluster variability may reflect situational effects rather than stable player tendencies.

This work provides a tournament-level, performance-based profiling framework for the 2023 FIFA Women's World Cup that can inform applied practices such as player profiling, training design, and match preparation. For example, coaches could use the cluster labels and radar profiles as role templates when planning a micro-cycle. If match preparation indicates a need for wide high-intensity penetration, players aligned with the explosive winger profile can be prioritised, and training can emphasise repeated sprint bouts combined with ball-carrying and crossing under fatigue. Analysts can also benchmark individuals against cluster centroids to identify under- or over-indexed attributes and translate these into targeted development objectives. Given the absence of positional tactical tracking variables, the derived clusters are interpreted as physical–technical performance profiles rather than tactical-role or playing-style inferences and therefore provide a descriptive baseline for future validation and extension. Future studies could consider longitudinal designs to verify the stability and evolution of player profiles across multiple tournaments and seasons. Beyond temporal analysis, future work may also investigate the relationship between offensive playing behaviours and match outcomes, exploring whether specific player profiles correlate with team success or tactical efficiency. Furthermore, comparative analyses between different sexes (e.g., women's vs. men's football), age groups (youth vs. senior), and contextual factors such as playing position, competitive level, or geographical region (e.g., Europe vs. Asia) could uncover deeper insights into how biological, developmental, and cultural factors influence performance profiles. These directions would offer coaches and analysts more specific, context-sensitive tools to optimise training, recruitment, and tactical planning in elite football.

Conclusion

This study categorised offensive performance profiles in elite women's football by integrating physical and technical data through K-means clustering analysis, revealing eighteen distinct player profiles. The approach advances performance analysis by unifying multiple dimensions of in-match behaviour, offering a more realistic understanding of football performance beyond isolated metrics. Theoretically, this study bridges statistical modelling, athlete profiling, and applied sport science, demonstrating the value of unsupervised learning in revealing latent player archetypes. It supports the shift toward more holistic, ecologically valid frameworks in performance analysis, where context, role, and physical capacity are considered jointly.

References

Alberto, F., Drews, R., Denardi, R. A., Travassos, B., & Corrêa, U. C. (2023). Identification of football teams styles of play by cluster analysis. *International Journal of Sports Science & Coaching*, *19*(3), 1019-1034. https://doi.org/10.1177/17479541231186796

Akhanli, S. E., & Hennig, C. (2023). Clustering of football players based on performance data and aggregated clustering validity indexes. *Journal of Quantitative Analysis in Sports*, *19*(2), 103-123. https://doi.org/10.1515/jqas-2022-0037

Altmann, S., Forcher, L., Ruf, L., Beavan, A., Groß, T., Lussi, P., Woll, A., & Härtel, S. (2021). Match-related physical performance in professional soccer: Position or player specific? *PLOS ONE*, *16*(9), e0256695. https://doi.org/10.1371/journal.pone.0256695

Beare, H., & Stone, J. A. (2019). Analysis of attacking corner kick strategies in the FA women's super league 2017/2018. *International Journal of Performance Analysis in Sport*, *19*(6), 893-903

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), 63–70.

Boone, J., Vaeyens, R., Steyaert, A., Bossche, L. V., & Bourgois, J. (2012). Physical Fitness of Elite Belgian Soccer Players by Player Position. *Journal of Strength and Conditioning Research*, *26*(8), 2051–2057. https://doi.org/10.1519/jsc.0b013e318239f84f

Bortnik, L., Bruce-Low, S., Burger, J. A., Alexander, J., Harper, D., Morgans, R., Carling, C., McDaid, K., & Rhodes, D. (2024). Physical match demands across different playing positions during transitional play and high-pressure activities in elite soccer. *Biology of Sport*, *41*(2). https://doi.org/10.5114/biolsport.2024.131815

Bradley, P. S., & Ade, J. D. (2018). Are Current Physical Match Performance Metrics in Elite Soccer Fit for Purpose or Is the Adoption of an Integrated Approach Needed?

International Journal of Sports Physiology and Performance, *13*(5), 656–664. https://doi.org/10.1123/ijspp.2017-0433

Bradley, P. S., Dellal, A., Mohr, M., Castellano, J., & Wilkie, A. (2014). Gender differences in match performance characteristics of soccer players competing in the UEFA Champions League. *Human Movement Science*, *33*, 159–171. https://doi.org/10.1016/j.humov.2013.07.024

Bradley, P. S., Di Mascio, M., Peart, D., Olsen, P., & Sheldon, B. (2010). High-Intensity Activity Profiles of Elite Soccer Players at Different Performance Levels. *Journal of Strength and Conditioning Research*, *24*(9), 2343–2351. https://doi.org/10.1519/jsc.0b013e3181aeb1b3

Bujnovky, D., Maly, T., Ford, K., Sugimoto, D., Kunzmann, E., Hank, M., & Zahalka, F. (2019). Physical Fitness Characteristics of High-level Youth Football Players: Influence of Playing Position. *Sports*, *7*(2), 46. https://doi.org/10.3390/sports7020046

Casal, C., Stone, J., Iván-Baragaño, I., & Losada, J. (2023). Effect of goalkeepers' offensive participation on team performance in the women Spanish La Liga: a multinomial logistic regression analysis. *Biology of Sport*, *41*(1), 29-39.

Carling, C., Reilly, T., & Williams, A. M. (2008). *Performance assessment for field sports: Physiological, psychological and match notational assessment in practice*. Routledge.

Carpita, M., Pasca, P., Arima, S., & Ciavolino, E. (2023). Clustering of variables methods and measurement models for soccer players' performances. *Annals of Operations Research*, *325*(1), 37–56. https://doi.org/10.1007/s10479-023-05185-w

Chang, V., Sreeram Sajeev, Qianwen Ariel Xu, Tan, M., & Wang, H. (2024). Football Analytics: Assessing the Correlation between Workload, Injury and Performance of Football Players in the English Premier League. *Applied Sciences*, *14*(16), 7217–7217. https://doi.org/10.3390/app14167217

Cho, H., Ryu, H., & Song, M. (2022). Pass2vec: Analyzing soccer players' passing style using deep learning. *International Journal of Sports Science & Coaching*, *17*(2), 355-365. https://doi.org/10.1177/17479541211033078

Di Salvo, V., Baron, R., Tschan, H., Calderon Montero, F., 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

Everitt, B. (1980). Cluster analysis. *Quality and Quantity*, *14*(1), 75–100. https://doi.org/10.1007/bf00154794

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley

Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P. R., & McRobert, A. P. (2016). Attacking and defensive styles of play in soccer: analysis of Spanish and English elite teams. *Journal of Sports Sciences*, *34*(24), 2195–2204. https://doi.org/10.1080/02640414.2016.1169309

FIFA. (n.d.). *Electronic Performance and Tracking Systems (EPTS)*. FIFA Quality Programme: Global Standards. Retrieved January 4, 2026, from https://quality.fifa.com/innovation/standards/eptsFIFAepts

FIFA. (2023, July 14). Post Match Summary Reports – FIFA Women's World Cup Australia & New Zealand 2023™ (Match report hub). FIFA Training Centre. Retrieved December 30, 2025, from https://www.fifatrainingcentre.com/en/game/tournaments/fifa-womens-world-cup/2023/match-report-hub/post-match-summary-reports.php

Forcher, L., Forcher, L., Härtel, S., Jekauc, D., Wäsche, H., Woll, A., Gross, T., & Altmann, S. (2022). Does Technical Match Performance in Professional Soccer Depend on the Positional Role or the Individuality of the Player? *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.813206

Forcher, L., Forcher, L., Wäsche, H., Jekauc, D., Woll, A., & Altmann, S. (2022). The influence of tactical formation on physical and technical match performance in male soccer: A systematic review. *International Journal of Sports Science & Coaching*, *18*(5), 174795412211013. https://doi.org/10.1177/17479541221101363

Fränti, P., Rezaei, M., & Zhao, Q. (2014). Centroid index: Cluster level similarity measure. *Pattern Recognition*, *47*(9), 3034–3045. https://doi.org/10.1016/j.patcog.2014.03.017

Fujii, K. (2021). Data-driven analysis for understanding team sports behaviors. *Journal of Robotics and Mechatronics*, *33*(3), 505-514. https://doi.org/10.20965/jrm.2021.p0505

Garcia-Unanue, J., Fernandez-Luna, A., Burillo, P., Gallardo, L., Sanchez-Sanchez, J., Manzano-Carrasco, S., & Felipe, J. L. (2020). Key performance indicators at FIFA Women's World Cup in different playing surfaces. *PLOS ONE*, *15*(10), e0241385. https://doi.org/10.1371/journal.pone.0241385

Garnica-Caparrós, M., & Memmert, D. (2021). Understanding gender differences in professional European football through machine learning interpretability and match actions data. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-90264-w

González-Rodenas, J., Aranda-Malaves, R., Tudela-Desantes, A., Nieto, F., Usó, F., & Aranda, R. (2020). Playing tactics, contextual variables and offensive effectiveness in English Premier League soccer matches. A multilevel analysis. *PLOS ONE*, *15*(2), e0226978. https://doi.org/10.1371/journal.pone.0226978

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, *2*(1), 1–21. https://doi.org/10.1038/s43586-022-00184-w

Harkness-Armstrong, A., Till, K., Datson, N., & Emmonds, S. (2023). Influence of match status and possession status on the physical and technical characteristics of elite youth female soccer match-play. *Journal of Sports Sciences*, *41*(15), 1437-1449. https://doi.org/10.1080/02640414.2023.2273653

Harkness-Armstrong, A., Till, K., Datson, N., Myhill, N., & Emmonds, S. (2022). A systematic review of match-play characteristics in women's soccer. *PLOS ONE*, *17*(6), e0268334. https://doi.org/10.1371/journal.pone.0268334

Hegedüs, P., Csala, D., Tóth, J., & Tóth, J., Jr. (2025). Age-Related Effects of Physical Performance on Technical and Tactical Outcomes in Youth Soccer. Sports, 13(6), 162. https://doi.org/10.3390/sports13060162

Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C., & Meyer, T. (2019). Machine learning in men's professional football: Current applications and future directions for improving attacking play. *International Journal of Sports Science & Coaching*, *14*(6), 798–817. https://doi.org/10.1177/1747954119879350

Hudl. (2025). Data Quality Process. Retrieved Janaurty 06, 2026, from https://support.hudl.com/s/article/event-data-quality?language=en_US.

Izquierdo, J. M., Marqués-Jiménez, D., & Redondo, J. C. (2023). Running demands and tactical individual actions of wingers appear to depend on the playing formations within an amateur football team. *Scientific Reports*, *13*(1), 8927. https://doi.org/10.1038/s41598-023-36157-6

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jamil, M., Liu, H., Phatak, A., & Memmert, D. (2021). An Investigation Identifying Which Key Performance Indicators Influence the Chances of Promotion to the Elite Leagues in Professional European Football. *International Journal of Performance Analysis in Sport*, *21*(4), 641–650.

Kirkendall, D. T., & Krustrup, P. (2021). Studying professional and recreational female footballers: A bibliometric exercise. *Scandinavian Journal of Medicine & Science in Sports*, *32*, 12-26. https://doi.org/10.1111/sms.14019

Kolena, B., Šviríková, B., & Vondráková, M. (2024). From Strikers to Keepers: Somatotype of Football Players from Slovakia. *Sports*, *12*(10), 271. https://doi.org/10.3390/sports12100271

Kozina, Z., Protas, M., Siryi, O., Hresko, O., Zavada, V., Ovdiienko, P., & Semenov, O. (2023). Comparative characteristics of the young football players of different game roles technical and physical fitness at the specialized basic training stage. *Health Technologies*, *1*(3), 6–18. https://doi.org/10.58962/ht.2023.1.3.6-18

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

Kubayi, A., & Larkin, P. (2020). Technical performance of soccer teams according to match outcome at the 2019 FIFA women's World Cup. *International Journal of Performance Analysis in Sport*, *20*(5), 908–916. https://doi.org/10.1080/24748668.2020.1809320

Lago-Peñas, C., Lago-Ballesteros, J., Dellal, A., & Gómez, M. (2010). Game-Related Statistics that Discriminated Winning, Drawing and Losing Teams from the Spanish Soccer League. *Journal of sports science & medicine*, *9*(2), 288–293.

Lethole, L., Kubayi, A., Toriola, A., Larkin, P., & Stone, J. A. (2024). 'Goalkeepers are players too': key attributes coaches' look for in talented youth soccer goalkeepers. *BMC Sports Science, Medicine and Rehabilitation*, *16*(1), 210.

Li, Y., Zong, S., Shen, Y., Pu, Z., Gómez, M.-Á., & Cui, Y. (2022). Characterizing player's playing styles based on player vectors for each playing position in the Chinese Football

Super League. *Journal of Sports Sciences*, *40*(14), 1629–1640. https://doi.org/10.1080/02640414.2022.2096771

Linke, D., Link, D., & Lames, M. (2020). Football-specific validity of TRACAB's optical video tracking systems. *PLOS ONE*, *15*(3), e0230179. https://doi.org/10.1371/journal.pone.0230179Marcelli0230179

Marcelli, L., Silvestri, F., Di Pinto, G., Gallotta, M. C., & Curzi, D. (2024). How Match-Related Variables Influence the Physical Demands of Professional Female Soccer Players during the Regular Season. *Journal of Functional Morphology and Kinesiology*, *9*(3), 149. https://doi.org/10.3390/jfmk9030149

Mayhew, S. R., & Wenger, H. A. (1985). Time motion analysis of professional soccer. *Journal of Human Movement Studies*, *11*(1), 49–52.

Memmert, D., & Raabe, D. (2023). Data analytics in football: Positional data collection, modelling and analysis. Routledge.

Modric, T., Malone, J. J., Versic, S., Andrzejewski, M., Chmura, P., Konefał, M., ... & Sekulic, D. (2022). The influence of physical performance on technical and tactical outcomes in the UEFA Champions League. *BMC Sports Science, Medicine and Rehabilitation*, *14*(1), 179.

Mohr, M., Krustrup, P., Andersson, H., Kirkendal, D., & Bangsbo, J. (2008). Match Activities of Elite Women Soccer Players at Different Performance Levels. *Journal of Strength and Conditioning Research*, *22*(2), 341–349. https://doi.org/10.1519/jsc.0b013e318165fef6

Narayanan, S., & Pifer, N. D. (2023). A data-driven framing of player and team performance in U.S. Women's soccer. *Frontiers in Sports and Active Living*, *5*, 1125528. https://doi.org/10.3389/fspor.2023.1125528

Oliva-Lozano, J.M., Yousefian, F., Chmura, P., Gabbett, T. J., & Cost, R. (2025). Analysis of FIFA 2023 Women's World Cup match performance according to match outcome and phase of the tournament. *Biology of Sport*. https://doi.org/10.5114/biolsport.2025.142643

Paul, D. J., Bradley, P. S., & Nassis, G. P. (2015). Factors Affecting Match Running Performance of Elite Soccer Players: Shedding Some Light on the Complexity. *International Journal of Sports Physiology and Performance*, *10*(4), 516–519. https://doi.org/10.1123/ijspp.2015-0029

Riezebos, M. (2021). *Identifying play styles of football players based on match event data* (Master's thesis, Erasmus University Rotterdam). Erasmus University Thesis Repository.

Rodrigues, F., & Pinto, Â. (2022). Prediction of football match results with Machine Learning. *Procedia Computer Science*, *204*, 463–470. https://doi.org/10.1016/j.procs.2022.08.057

Rodrigues, V., Carlgren, D., Holash, R. J., & Katz, L. (2023). Technological Breakthroughs in Sport: Current Practice and Future Potential of Artificial Intelligence, Virtual Reality, Augmented Reality, and Modern Data Visualization in Performance Analysis. *Applied Sciences*, *13*(23), 12965–12965. https://doi.org/10.3390/app132312965

Sampaio, J., McGarry, T., Calleja-González, J., Jiménez Sáiz, S., Schelling i del Alcázar, X., & Balciunas, M. (2015). Exploring Game Performance in the National Basketball Association Using Player Tracking Data. *PLOS ONE*, *10*(7), e0132894. https://doi.org/10.1371/journal.pone.0132894

Sarmento, H., Clemente, F. M., Araújo, D., Davids, K., McRobert, A., & Figueiredo, A. (2017). What Performance Analysts Need to Know About Research Trends in Association Football (2012–2016): A Systematic Review. *Sports Medicine*, *48*(4), 799–836. https://doi.org/10.1007/s40279-017-0836-6

Sarmento, H., Figueiredo, A., Lago-Peñas, C., Milanovic, Z., Barbosa, A., Tadeu, P., & Bradley, P. S. (2018). Influence of Tactical and Situational Variables on Offensive Sequences During Elite Football Matches. *Journal of Strength and Conditioning Research*, *32*(8), 2331–2339. https://doi.org/10.1519/jsc.0000000000002147

Sarmento, H., Marcelino, R., Anguera, M. T., CampaniÇo, J., Matos, N., & LeitÃo, J. C. (2014). Match analysis in football: a systematic review. *Journal of Sports Sciences*, *32*(20), 1831–1843. https://doi.org/10.1080/02640414.2014.898852

Sarmento, H., Martinho, D. V., Gouveia, É. R., Afonso, J., Chmura, P., Field, A., Nestor Ordoñez Savedra, Oliveira, R., Gibson Praça, Silva, R., Barrera-Díaz, J., & Filipe Manuel Clemente. (2024). The Influence of Playing Position on Physical, Physiological, and Technical Demands in Adult Male Soccer Matches: A Systematic Scoping Review with Evidence Gap Map. *Sports Medicine, 54*. https://doi.org/10.1007/s40279-024-02088-z

Stone, J. A., Kubayi, A., Smith, A., & North, J. S. (2025). Analysis of throw-ins strategy on performance metrics in five men's European football leagues. *International Journal of Performance Analysis in Sport*, 1-15.

Serhat Emre Akhanl, & Hennig, C. (2023). Clustering of football players based on performance data and aggregated clustering validity indexes. *Journal of Quantitative Analysis in Sports*, *0*(0). https://doi.org/10.1515/jqas-2022-0037

Teunissen, J. W., De Bock, J., Schasfoort, D., Slembrouck, M., Verstockt, S., Lenoir, M., & Pion, J. (2024). "There is more that unites us than divides us". Optimizing talent transfer processes by clustering 34 sports by their task, individual and environmental similarities. *Frontiers in Sports and Active Living*, *6*. https://doi.org/10.3389/fspor.2024.1445510

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423. https://doi.org/10.1111/1467-9868.00293

Trewin, J., Meylan, C., Varley, M. C., & Cronin J. (2018). Effect of match factors on the running performance of elite female soccer players. *Journal of Strength & Conditioning Research*, *32*(7), 2002–2009. https://doi.org/10. 1519/JSC.0000000000002584

Ventaja-Cruz, J., Cuevas, M., Tejada-Medina, V., & Martín-Moya, R. (2024). A Bibliometric Study on the Evolution of Women's Football and Determinants Behind Its Growth over the Last 30 Years. *Sports*, *12*(12), 333–333. https://doi.org/10.3390/sports12120333

Vescovi, J. D., & Falenchuk, O. (2019). Contextual factors on physical demands in professional women's soccer: Female athletes in motion study. *European Journal of Sport Science*, *19*(2), 141–146. https://doi. org/10.1080/17461391.2018.1491628

Vilar, L., Araújo, D., Davids, K., & Button, C. (2012). The Role of Ecological Dynamics in Analysing Performance in Team Sports. *Sports Medicine*, *42*(1), 1–10. https://doi.org/10.2165/11596520-000000000-00000

Wang Z., & You. S. (2021) Soccer Tactical Analysis Methods and Development Tendency Based on Positional Data Under the Background of Big Data. *Journal of Shanghai University of Sport*, *45*(9): 60-69. doi:10.16099/j.sus.2021.09.006

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wongoutong, C. (2024). The impact of neglecting feature scaling in k-means clustering. *PLOS ONE*, *19*(12), e0310839. https://doi.org/10.1371/journal.pone.0310839

Yeung, C., Bunker, R., & Fujii, K. (2024). Unveiling Multi-Agent Strategies: A Data-Driven Approach for Extracting and Evaluating Team Tactics from Football Event and Freeze-Frame Data. *Journal of Robotics and Mechatronics*, *36*(3), 603–617. https://doi.org/10.20965/jrm.2024.p0603

Zhang, W., Gong, B., Tao, R., Zhou, F., Ángel, M., & Zhou, C. (2024). The influence of tactical formation on physical and technical performance across playing positions in the Chinese super league. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-53113-0

Table 1. The definition of physical (From FIFA) and technical indicators (From Statsbomb)

Variables	Definition	Category of data
Playing Time	Total playing time on the pitch of each players	Time
Total Distance	The total distance (meters) that the player covers during playing time	Physical
Zone 1	Meters cover by very low-speed activity, with the player moving within **[0,7] km/h**.	Physical
Zone 2	Meters cover by low-to-moderate speed running, with the player moving within (**7,13] km/h**.	Physical
Zone 3	Meters cover by moderate-to-high speed running, with the player moving within (**13,19] km/h**.	Physical
Zone 4	Meters cover by high-speed running, with the player moving within (**19,23] km/h**.	Physical
Zone 5	Meters covered by sprinting speeds, with the players moving at **23+ km/h**.	Physical
High Speed Runs	Number of efforts performed within (**19–23] km/h** (Zone 4).	Physical
Sprints	Number of efforts performed at **23+ km/h** (Zone 5).	Physical
Top Speed	The highest speed (km/h) a player reaches in the session or match.	Physical
Total Passes	Number of passes **attempted** by the player.	Performance
Key Passes	Number of passes that **directly lead to a shot by the receiving player.**	Performance
Crosses	Number of crosses (a ball passing from flank area, defined by Statsbomb) attempted by the player.	Performance
Headed Passes	Number of passes made by a player using their head.	Performance
Total Shots	Number of shots made by the player.	Performance
Headed Shots	Number of shots made by the player using their head.	Performance
Dribbles	Number of single attempts by a player to beat 1 opponent.	Performance
Carries	Number of controlling the ball at their feet while moving or standing still by the player.	Performance

Table 2. Standardised centroids for player clusters across physical and offensive performance metrics.

Cluster	Playing Time	Total Distance	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	High speed runs	Sprints	Top Speed	Total Passes	Key Passes	Crosses	Headed Passes	Total Shots	Headed Shots	Dribbles	Carries
1	0.71	0.73	0.56	0.70	0.67	0.46	0.11	0.70	0.46	0.65	0.27	0.08	0.06	0.15	0.08	0.02	0.12	0.27
2	0.06	0.18	0.18	0.19	0.19	0.17	0.09	0.20	0.18	0.63	0.07	0.03	0.03	0.06	0.04	0.01	0.04	0.08
3	0.68	0.66	0.60	0.57	0.53	0.52	0.26	0.60	0.55	0.74	0.48	0.18	0.31	0.17	0.10	0.02	0.17	0.43
4	0.31	0.38	0.37	0.33	0.32	0.36	0.24	0.36	0.42	0.77	0.13	0.10	0.16	0.06	0.13	0.05	0.19	0.16
5	0.68	0.62	0.59	0.57	0.49	0.37	0.15	0.54	0.40	0.68	0.24	0.05	0.05	0.11	0.05	0.02	0.07	0.21
6	0.63	0.56	0.61	0.44	0.39	0.46	0.40	0.47	0.57	0.81	0.16	0.07	0.14	0.06	0.12	0.03	0.29	0.23
7	0.68	0.63	0.61	0.55	0.50	0.45	0.20	0.54	0.49	0.73	0.22	0.08	0.15	0.07	0.12	0.04	0.53	0.29
8	0.97	0.89	0.84	0.74	0.68	0.63	0.28	0.76	0.64	0.75	0.38	0.11	0.15	0.23	0.14	0.06	0.14	0.32
9	0.19	0.30	0.28	0.29	0.28	0.26	0.11	0.30	0.28	0.68	0.11	0.04	0.05	0.09	0.05	0.02	0.05	0.12
10	0.73	0.25	0.59	0.12	0.05	0.03	0.01	0.08	0.03	0.38	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.09
11	0.70	0.66	0.59	0.57	0.54	0.55	0.27	0.61	0.60	0.75	0.22	0.05	0.05	0.16	0.05	0.01	0.08	0.19
12	0.71	0.57	0.61	0.56	0.35	0.26	0.12	0.42	0.30	0.69	0.18	0.01	0.02	0.15	0.03	0.03	0.02	0.14
13	0.70	0.63	0.60	0.61	0.44	0.29	0.12	0.50	0.31	0.66	0.57	0.05	0.04	0.20	0.05	0.03	0.05	0.53
14	0.74	0.72	0.64	0.57	0.58	0.69	0.46	0.66	0.77	0.79	0.22	0.13	0.19	0.12	0.19	0.03	0.20	0.25
15	0.67	0.62	0.60	0.53	0.47	0.45	0.23	0.55	0.51	0.73	0.24	0.09	0.08	0.14	0.36	0.50	0.12	0.24
16	0.42	0.49	0.43	0.45	0.42	0.41	0.16	0.47	0.44	0.71	0.14	0.04	0.05	0.10	0.08	0.03	0.08	0.15
17	0.73	0.64	0.63	0.59	0.48	0.37	0.14	0.54	0.38	0.68	0.30	0.06	0.03	0.49	0.05	0.02	0.04	0.22
18	0.44	0.49	0.42	0.50	0.41	0.22	0.05	0.43	0.22	0.57	0.22	0.03	0.02	0.11	0.06	0.04	0.08	0.21

Table 3. Unstandardised centroids for player clusters across physical and offensive performance metrics

Cluster	Playing Time	Total Distance	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	High Speed	Sprints	Top Speed	Total Passes	Key Passes	Crosses	Headed Passes	Total Shots	Headed Shots	Dribbles	Carries
1	98.7	11131.8	3686.7	3772.3	2689.1	687.2	296.8	218.8	59.7	29.7	46.5	2.7	4.4	1.3	2.7	0.2	2.4	41.1
2	99.2	4544.7	3420.0	854.9	236.6	28.8	4.5	27.6	2.8	23.0	27.7	0.0	0.0	0.0	0.0	0.0	0.0	12.6
3	96.9	9221.1	3494.3	3637.2	1677.4	304.2	108.1	145.6	27.1	28.1	26.9	0.2	0.3	1.5	0.4	0.1	0.3	16.1
4	68.2	7574.9	2554.4	2669.7	1730.8	431.9	188.1	144.2	39.4	29.3	21.6	0.6	1.0	0.9	1.3	0.2	1.4	20.1
5	98.9	9933.5	3642.7	3694.9	2074.2	393.7	127.9	172.4	32.8	28.3	48.1	0.5	0.4	5.9	0.5	0.1	0.5	30.2
6	94.4	9658.3	3433.4	3429.6	2112.3	467.0	216.1	178.0	43.7	29.6	31.9	0.3	0.7	1.4	0.6	0.1	1.0	23.0
7	96.8	9955.6	3500.7	4010.2	2030.2	311.1	103.5	170.8	27.1	27.9	97.0	0.6	0.6	2.7	0.6	0.1	0.4	78.5
8	95.4	10239.1	3482.2	3680.9	2333.6	528.8	213.5	191.0	44.6	29.2	76.5	1.6	3.1	2.0	1.1	0.1	2.0	58.1
9	97.8	10835.6	3481.7	3752.0	2645.2	688.4	268.1	217.8	59.2	29.3	33.7	0.5	0.8	2.0	1.0	0.1	1.3	27.4
10	92.6	9582.2	3503.1	3372.8	2076.7	444.1	185.5	167.4	39.9	29.1	34.2	0.8	1.6	0.9	1.5	0.2	6.4	38.1
11	95.2	9819.1	3650.3	3037.5	2075.0	610.9	445.4	178.9	61.6	30.7	28.8	1.1	2.0	0.7	2.4	0.1	4.5	34.0
12	97.0	9573.3	3554.5	3737.2	1888.4	295.8	97.5	156.4	25.9	27.8	58.4	0.4	0.4	1.6	0.8	0.3	0.9	45.0
13	122.7	13493.7	4817.9	4769.4	2998.3	664.7	243.7	247.2	55.1	29.3	59.8	0.8	1.6	2.9	1.9	0.3	1.8	42.2
14	48.7	5320.6	1768.9	1917.7	1235.2	288.2	110.6	101.7	25.0	28.5	20.2	0.5	0.8	1.0	0.8	0.1	0.8	17.2
15	96.3	10977.3	3325.3	4256.1	2826.6	472.7	96.6	216.0	37.7	27.6	41.6	0.8	0.7	2.0	0.9	0.1	1.4	33.1
16	36.1	3606.7	1229.1	1264.4	857.4	181.9	73.9	69.2	15.6	27.3	12.2	0.3	0.4	0.7	0.5	0.1	0.6	10.1
17	69.8	7829.5	2494.3	3141.6	1867.5	270.3	55.6	145.9	21.9	26.8	29.9	0.3	0.3	1.4	0.7	0.1	1.0	24.4
18	93.2	9668.7	3517.3	3373.6	2085.6	489.2	202.9	178.9	44.2	29.1	36.2	1.0	0.9	1.7	4.9	2.5	1.7	32.4

Table 4. Representative players closest to each cluster centroid with position, team, and match context.

Player Name	Position	Team Name	Playing Time	Total Distance	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	High Speed	Sprints	Top Speed	TotalPasse	Key Passes	Crosses	Headed	Total	Headed	Dribbles	Carries	Cluster
Player 1	Forward	South Africa Women's	95	10817.5	3317	3497	3009	707	285	231	62	28.8	32	3	4	1	3	0	2	26	1
Player 2	Goalkeeper	Zambia Women's	100	4495	3261	951.1	260.7	22	0	34	3	23.1	31	0	0	0	0	0	0	18	2
Player 3	Defender	Denmark Women's	94	8987.6	3274	3722	1621	271	98	128	24	28.7	26	0	0	2	0	0	0	15	3
Player 4	Forward	France Women's	66	7054.3	2308	2638	1520	393	193	117	33	28.9	31	1	1	1	1	0	2	21	4
Player 5	Defender/Midfielder	Netherlands Women's	98	9604.7	3385	3562	2211	350	94	190	32	30.5	53	0	0	5	1	0	0	37	5
Player 6	Defender	New Zealand Women's	100	10038.5	3511	3625	2172	507	220	165	41	28.8	41	0	1	2	0	0	0	22	6
Player 7	Midfielder	Netherlands Women's	95	10078.2	4089	3590	1920	345	131	159	24	27.7	90	1	1	2	0	0	1	72	7
Player 8	Defender	France Women's	103	10730.3	3365	3988	2582	549	244	212	51	29.4	60	1	4	1	1	0	2	56	8
Player 9	Forward/Midfielder	Republic of Ireland Women's	97	10496.9	3350	3823	2339	694	288	189	55	29.7	39	0	1	3	1	0	3	27	9
Player 10	Forward	Switzerland Women's	85	8627.2	3451	2671	1879	425	198	154	36	29.2	43	0	1	0	1	0	6	38	10
Player 11	Midfielder	Sweden Women's	89	10186.6	3667	3196	2135	678	509	181	64	30.1	34	1	1	1	0	0	4	33	11
Player 12	Midfielder/Defender	China PR Women's	103	10112.4	3734	4161	1752	325	138	149	24	27.3	51	0	0	1	1	0	1	39	12
Player 13	Winger	Sweden Women's	121	13951.1	5094	4792	3143	679	241	264	53	29.1	46	1	1	3	2	1	0	45	13
Player 14	Forward	England Women's	46	5358.5	1785	1755	1439	267	109	104	26	29	12	0	2	1	0	0	1	17	14

Player 15	Winger	Switzerland Women's	94	11152.6	3718	4088	2836	421	88	206	35	27.7	30	1	1	2	1	0	2	23	15
Player 16	Midfielder	Nigeria Women's	32	3773	1324	1352	878	166	50	66	15	26.8	11	1	0	1	0	0	0	10	16
Player 13	Winger	Sweden Women's	67	7753.2	2591	2819	1946	303	93	161	26	27	30	0	1	1	1	0	0	23	17
Player 17	Forward	France Women's	96	10343.5	3445	3840	2448	474	134	190	41	27.6	38	1	1	1	3	2	2	42	18

Table S1. High-correlation feature pairs (| ρ |⩾0.8/0.9)

var1	var2	rho
Zone 4	sprints	0.9417122
total distance	high.speed.runs	0.9098041
total distance	Zone 2	0.8871384
total distance	Zone 3	0.8855692
total passes	carries	0.845553

Figure 1. Two-dimensional principal component projection illustrating the distribution of K-means player clusters.

Figure 2. Quartile-based heat map illustrating relative performance metrics for each player cluster.

Figure 3. Radar charts illustrating representative players' normalised performance profiles for Clusters 1, 2, 4, 8, 10, and 13.

Figure A1. Spearman correlation heatmap