

Goalkeeper passing behaviour in women's football: A graph-based analysis

DING, Junyuan <<http://orcid.org/0009-0002-6455-6844>>, STRAFFORD, Ben <<http://orcid.org/0000-0003-4506-9370>>, KUBAYI, Alliance <<http://orcid.org/0000-0002-8370-3056>> and STONE, Joseph <<http://orcid.org/0000-0002-9861-4443>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<https://shura.shu.ac.uk/37881/>

This document is the Accepted Version [AM]

Citation:

DING, Junyuan, STRAFFORD, Ben, KUBAYI, Alliance and STONE, Joseph (2026). Goalkeeper passing behaviour in women's football: A graph-based analysis. *Journal of Sports Sciences*, 44 (16), 2150-2166. [Article]

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

Goalkeeper Passing Behaviour in Women's Football: A Graph-Based Analysis

Ding, J.¹, Strafford, B.W.¹, Kubayi, A.² & Stone J.A.¹

¹School of Sport and Physical Activity, Sheffield Hallam University, United Kingdom

²Department of Human Movement and Therapeutic Sciences, Tshwane University of
Technology, South Africa

[Junyuan Ding: https://orcid.org/0009-0002-6455-6844](https://orcid.org/0009-0002-6455-6844)

Ben W. Strafford: <https://orcid.org/0000-0003-4506-9370>

Alliance Kubayi: <https://orcid.org/0000-0002-8370-3056>

Joseph A. Stone: <https://orcid.org/0000-0002-9861-4443>

-

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. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.

Correspondence concerning this article should be addressed to Junyuan Ding, at Junyuan.Ding@shu.ac.uk

23 **Abstract**

24 As elite women's football evolves tactically, understanding goalkeeper distribution
25 behaviour is increasingly important for performance analysis and coaching.
26 Traditional approaches rely on isolated passing metrics, providing limited insight into
27 how goalkeeper actions are organised within possession structures. This study applied
28 a graph-based clustering framework to classify goalkeeper passing behaviours in the
29 English Women's Super League across three seasons. The dataset comprised 29,911
30 goalkeeper passes from 27,895 possessions. Pass embeddings were represented within
31 a relational graph and clustered using K-means ($k = 14$), identifying recurrent
32 distribution patterns shaped by spatial context, defensive pressure, and possession
33 phase. Results indicated that behaviours were organised along continuous gradients
34 rather than discrete categories. Central clusters reflected adaptable distributions,
35 whereas peripheral clusters captured more specialised actions, including long passes
36 under defensive pressure. Lateral tendencies suggested interactions between
37 dominant-side preferences and spatial availability. Phase analysis showed goalkeeper
38 involvement was most prominent during transitional moments, particularly at
39 possession initiation and termination, with comparatively limited involvement in
40 sustained attacking sequences. Collectively, these findings provide a structured
41 account of goalkeeper distribution in elite women's football and demonstrate how
42 graph-based modelling can support behavioural profiling, tactical interpretation, and
43 evidence-based coaching practice.

44 **Keywords:** Goalkeeper, passing behaviour, women's football, ecological dynamics

1. Introduction

The rapid development of performance analysis technologies has significantly transformed football research over the past decade, particularly through the integration of big data and advanced computational modelling approaches (Rein & Memmert, 2016). Advances in automated tracking systems, Global Positioning Systems, and video motion analysis software have enabled the collection of large volumes of positional, technical, and physical data, providing more comprehensive insights into team and player performance (Cho et al., 2022; Herold et al., 2019). These technological developments have facilitated the use of machine and deep learning methods to analyse complex interactions and extract meaningful patterns from large datasets in sport (Ayala et al., 2019; Rico-González et al., 2023). The predictive and clustered outputs generated by machine learning have been applied to injury prevention, performance enhancement, and the prediction of athlete potential (Campbell et al., 2021; Oliver et al., 2020; Sheng et al., 2021; Klemp et al., 2026). Such approaches are particularly valuable in football due to the complex and dynamic nature of player behaviour and the interactive structure of the game (Herold et al., 2019).

Despite these methodological advances, current evidence still provides an incomplete picture of women's football performance, highlighting the need for systematic synthesis to consolidate what is known and identify key uncertainties (Cherappurath et al., 2024). The professionalisation of women's football has led to notable improvements in competition quality, increased media attention, and growing

67 participation worldwide (FIFA, 2019; 2023). However, the extent to which existing
68 knowledge provides an accurate and comprehensive understanding of the women's
69 game remains uncertain (Okholm Kryger et al., 2022). Differences in tactical
70 behaviour, physical characteristics, and match demands between men's and women's
71 football (Bradley et al, 2014; Datson et al, 2017) suggest that performance indicators
72 validated in men's competitions cannot be directly generalised to the women's
73 context (Okholm Kryger et al., 2022). Consequently, further research is required to
74 develop a contextualised understanding of performance characteristics in women's
75 football.

76 One understudied area concerns the evolving role of the goalkeeper as an active
77 participant in collective attacking organisation. Law changes, including restrictions on
78 handling back-passes and time limits for ball release, have increased the technical and
79 tactical demands placed on goalkeepers (Casal et al., 2023). In modern football,
80 goalkeepers are increasingly required to operate outside the penalty area and
81 contribute to build-up play as additional outfield players in defensive and offensive
82 phases (Lethole et al., 2024; Sarmiento, Anguera, et al., 2018). This shift reflects
83 broader tactical evolution in which goalkeepers initiate possession and support
84 attacking progression (Seaton & Campos, 2011; de Baranda et al., 2019). Evidence
85 from women's football further indicates that goalkeepers from teams progressing
86 beyond the group stage in international competitions show greater offensive
87 involvement and complete more successful passes across pitch areas (Sainz de
88 Baranda et al., 2019). Despite this growing importance, goalkeeper behaviour remains

insufficiently examined, particularly in women's football, where contextual and performance characteristics may differ from men's competitions.

Within this expanded tactical role of the goalkeeper, ball distribution represents a key component of offensive organisation. Passing in football is widely recognised as a core performance indicator reflecting team coordination and interaction (Clemente et al., 2015; Grund, 2012; Liu et al., 2015). Passing behaviour of teams can adapt to match objectives and contextual demands, including increasing forward passes to create scoring opportunities (Memmert et al., 2017), increasing passing frequency to maintain possession (Lago-Peñas & Dellal, 2010), and playing penetrative passes to advance into the final third (Tenga et al., 2010). The growing reliance on goalkeeper foot-based distribution highlights their role in shaping possession structures and initiating attacks (Seaton & Campos, 2011), and distribution is now considered a critical technical and tactical skill (Falcão et al., 2026). However, prior research has largely focused on isolated indicators such as pass frequency, success rates, and possession metrics (Castellano et al., 2012; Lago-Peñas et al., 2010; Liu et al., 2015), with limited attention to the structural organisation of goalkeeper passing behaviour and its interaction with contextual match factors.

Understanding goalkeeper distribution requires theoretical approaches that account for interactions between players, tasks, and environmental constraints. Traditional performance analysis has often examined isolated variables without fully addressing contextual dynamics (McLean et al., 2017; Sarmiento, Clemente, et al., 2018). From an ecological dynamics perspective, behaviour emerges through

continuous performer – environment interaction, shaped by individual, task, and environmental constraints (Araújo et al., 2006; Araújo & Davids, 2016; Davids & Araújo, 2010). Players perceive affordances and adapt actions to evolving contextual information (Fajen et al., 2008; Travassos et al., 2013), with behaviour reflecting system reorganisation rather than isolated technique (Araújo et al., 2022).

Accordingly, passing should be conceptualised as a functional and adaptive process shaped by constraint interaction dynamics (Riley et al., 2011), with playing style representing recurrent behavioural patterns emerging in match contexts (Hewitt et al., 2016).

Recent developments in data-driven approaches have enabled more objective descriptions of passing behaviour and tactical patterns (Decroos et al., 2019; Peña & Touchette, 2012; Rein & Memmert, 2016). Previous studies have applied machine learning methods to quantify spatial invasion patterns (Bialkowski et al., 2014; Rein et al., 2017), identify player performance styles (Li et al., 2022), and analyse passing networks (Peña & Touchette, 2012; Yu et al., 2020). Other approaches have used zone-based representations and clustering techniques to examine passing styles, such as constructing heatmaps of passing distributions across spatial zones (Brooks et al., 2016). Additionally, representation learning methods, such as pass2vec, a contextual vector representation of football passes, have been developed to capture player-specific passing tendencies based on spatial and directional features (Cho et al., 2022), while quantitative frameworks have been proposed to evaluate pass effectiveness using multivariate modelling and dimensionality reduction techniques

(Goes et al., 2019). Although these approaches provide valuable insights into passing effectiveness and stylistic tendencies, they primarily focus on spatial distributions or player-level characteristics, with limited consideration of relational interactions and contextual constraints underlying passing decisions.

Graph Neural Networks (GNNs) provide a promising framework for modelling football as a system of interacting agents. By representing players or actions as nodes and interactions as edges, GNNs enable the modelling of relational structures and complex dependencies within game situations (Kipf & Welling, 2017; Veličković et al., 2018). These models have been applied to player value assessment (Decroos et al., 2019; Zhao et al., 2025), tactical evaluation (Stöckl et al., 2021), and pattern recognition in football (Anzer et al., 2022). The ability of GNNs to integrate contextual and relational information aligns with ecological dynamics perspectives by capturing behaviour as emerging from interaction systems. However, limited research has applied interaction-based representation learning to examine goalkeeper passing behaviour, particularly in women's football.

Therefore, the present study proposes a comprehensive framework for analysing goalkeeper passing behaviour in elite women's football by integrating contextual and situational information within a relational modelling approach. Specifically, this study aims to identify distinct goalkeeper passing behaviours using Graph Sample and Aggregation (GraphSAGE) to generate contextualised representations of passing actions, followed by K-means clustering to classify behavioural patterns.

Furthermore, the study examines the relationships between identified passing

behaviours and match situations, including game outcomes and possession success.

Hence, this research seeks to advance understanding of goalkeeper decision-making processes and provide new insights into the tactical role of goalkeepers in women's football.

2. Method

2.1. Analytical Framework

To examine goalkeeper passing behaviour within its interaction context, a graph-based representation learning framework was adopted. GraphSAGE is a spatial-based graph neural network that learns node representations by sampling and aggregating features from local neighbourhoods, enabling inductive learning on previously unseen nodes (Hamilton et al., 2018). Possession-level graphs were constructed from match event data, where nodes represented individual passes and edges captured temporal dependencies between consecutive actions. A graph neural network was then used to learn context-aware embeddings of passes. Finally, goalkeeper passes were extracted from the learned representations and clustered using k-means to identify distinct behavioural patterns. The whole construct process is illustrated in Figure 1.

****Insert Figure 1****

Figure 1. The Process of GraphSAGE Based Analytical Framework.

2.2. Data Sample

All 396 matches from the Football Association Women's Super League 2022/23, 2023/24 and 2024/25 season were selected with data sourced from Hudl: Statsbomb (https://www.hudl.com/en_gb/products/statsbomb). Each event

contains extensive contextual information (160+ variables per event, depending on event type), including event timing, player and team identifiers, action descriptors and outcomes, and spatial information such as pitch coordinates (e.g., x-y locations, and where available additional spatial components). The study was approved by the local university [Anonymise for peer-review] ethics committee (ID: ER70779229).

In total 27,895 possessions, containing 217,974 total passes, of which 29,911 were performed by the goalkeeper during the possessions were extracted from the data set. Pass-level features describing spatial, technical, and contextual characteristics were extracted for subsequent analysis (see Table 1). The starting and ending area variables of a pass are condensed from StatsBombs original x (0 to 120) and y (0 to 80) positions into 25 zones based on the field reference lines (see Figure S1). Each pass was uniquely identified by its match, possession, and sequential order within the possession. As all available league matches across the three seasons were included, the dataset represents a complete population sample rather than a sampled subset. Therefore, formal a priori sample size estimation was not required. Cluster stability was evaluated using internal validity indices (CH, DB, silhouette), providing additional support for model robustness.

****Insert Table 1****

Table 1. Variables and Definition.

2.3. Graph Modelling and Clustering

Possession-level graphs were constructed to represent passing sequences involving the goalkeeper. Within each possession, individual passes were defined as nodes, and directed edges were established between temporally adjacent passes to preserve the sequential flow of play. Each node was associated with a feature vector capturing spatial, technical, and contextual properties of the pass. GNN were employed to learn low-dimensional representations of passes by aggregating information from neighbouring actions. This process enabled each pass to be encoded not only by its intrinsic characteristics but also by its surrounding interaction context. The resulting embeddings were 64-dimensional and were subsequently used for clustering analysis.

The learned goalkeeper pass embeddings were standardised using z-score normalisation prior to clustering to ensure comparable feature scales. K-means clustering (Jain, 2010; Everitt et al., 2011) with Euclidean distance was then applied to group passes with similar behavioural characteristics in the embedding space. The number of clusters was determined using a multi-criteria approach incorporating gap statistics (Tibshirani et al., 2001), silhouette analysis (Rousseeuw, 1987; Shutaywi, 2021), and clustering stability (Hennig, 2007; Ullmann, 2022). Based on these evaluations, a solution with 14 clusters was selected. To enhance robustness, the algorithm was initialised 30 times with different centroid seeds ($n_{\text{init}} = 30$), retaining the configuration that minimised the within-cluster sum of squares. Each run used a maximum of 300 iterations with a fixed random state to ensure reproducibility. This

procedure enabled the identification of latent structures in goalkeeper passing behaviour not readily observable in the original feature space.

2.4.Goalkeeper passing behaviour profile

Following clustering, each cluster was characterised using multidimensional profiling to describe its behavioural properties. Distributions of spatial origin and destination, pass mechanics (e.g., height and body part), recipient positions, and key continuous variables were examined to capture the defining features of each passing pattern. Cluster profiling was conducted by examining spatial distributions through pass end-location heatmaps and by analysing cluster-level summaries of key passing features. To further characterise the identified clusters, principal component analysis (PCA) was conducted to examine the dominant dimensions underlying the embedding space and the variables driving cluster separation. Between-cluster differences were quantified using effect size measures and variance-based metrics (e.g., within- and between-cluster sum of squares) to assess cluster separation and distinctiveness. A two-dimensional PCA projection further illustrated the geometric relationships among clusters in the embedding space.

3. Result

The final k-means solution ($n = 29,911$; $p = 64$; $k = 14$) captured meaningful structure in goalkeeper passing behaviour, with 82.0% of total variance explained by between-cluster differences ($\text{BetweenSS}/\text{TotSS} = 0.820$). Internal validity indices indicated moderate cluster separation ($\text{Calinski-Harabasz} = 10,495$; $\text{Davies-Bouldin} = 1.34$).

The relatively low silhouette value (0.212) suggests that clusters reflect continuous behavioural tendencies rather than strictly discrete categories.

3.1.PCA analysis and effect size of cluster

The PCA projection (see Figure 2) demonstrated clear structural variation across clusters. Effect size analysis confirmed significant between-cluster differences across all variables (FDR-corrected $q < .001$). Spatial and distribution-related variables dominated cluster separation, with pass angle ($\eta^2 = .85$), end area ($\eta^2 = .78$), and pass length ($\eta^2 = .70$) showing the largest effects, followed by starting location ($\eta^2 = .39$). Moderate effects were observed for pass duration ($\eta^2 = .38$) and recipient position ($\eta^2 = .24$). Contextual indicators, including pass switching, defensive pressure, and through-ball actions, showed smaller but significant associations (Cramér's $V = .10-.17$), whereas possession length and passing speed demonstrated limited explanatory power ($\eta^2 < .10$). PC1 primarily reflected spatial-directional variation (pass angle and end location), while PC2 captured distribution length and tempo, driven by pass length and speed (see Figure S2).

****Insert Figure 2****

Figure 2. Two-dimensional principal component projection illustrating the distribution of K-means GK passing behaviours clusters.

3.2. Spatial profile of clusters

Figure 3 shows structured spatial organisation across the 14 clusters. Rather than random dispersion, clusters were arranged along lateral corridors and depth gradients. Six mirrored cluster pairs (4&9, 3&1, 14&10, 7&8, 11&12, 5&6) demonstrated left-

right lateral symmetry across pitch regions, occupying distinct vertical zones from defensive areas to advanced flanks. Starting locations were strongly concentrated in and around the penalty area, with most clusters originating in Zone 3 or adjacent boundary zones. A smaller number extended into deeper defensive regions. Overall, goalkeeper distributions exhibited a clear penalty-area-centred spatial gradient. Overall spatial distributions further supported this pattern (see Figure 4). Across all passes, starting locations were highly concentrated in the central defensive zone (Zone 3, 64%) and its neighbouring region (Zone 8, 18.4%), with minimal occurrences in advanced pitch areas. This confirms that goalkeeper passing actions were primarily initiated from a restricted defensive action space. In contrast, pass destinations were more spatially distributed, with the highest frequencies observed in central and wide build-up areas (Zones 10, 7, 6, and 8).

****Insert Figure 3 and 4****

Figure 3. Location profile of 14 clusters (Attacking Direction from left to right).

Figure 4. Location profile of all passes (Attacking Direction from left to right).

Pass angle distributions further revealed clear directional structure across clusters (Figure 5). Rather than exhibiting uniform angular distributions, passes within each cluster were concentrated within narrow directional ranges, indicating consistent orientation preferences. Several mirrored cluster pairs showed opposite lateral orientations, reflecting left–right symmetric distribution patterns. Directional profiles broadly corresponded to central progression and lateral outlet behaviours, supporting the interpretation of clusters as distinct goalkeeper distribution strategies.

****Insert Figure 5****

Figure 5. Pass angle of 14 clusters (Attacking Direction from left to right).

3.3. Passing Feature Profile

Recipient distributions revealed differentiated positional preferences (See Figure 6).

Most clusters primarily targeted centre-backs ($n = 8$), followed by forwards and wingers ($n = 4$), fullbacks and wingbacks ($n=2$), and with 1 cluster targeting defensive midfielders (30.2%), but still mainly targeting centre-backs (50.4%). Delivery type varied, with 7 clusters dominated by ground passes and 5 characterised by higher proportions of aerial distributions. Two clusters included relatively equal on ground and high passes distributions (Cluster 14 & 10). High pass clusters were typically associated with forward and winger, or fullback and wingback-oriented targets. Footed passes predominated overall; however, clusters 5 and 6 showed elevated drop-kick proportions ($>10\%$), whereas clusters 4, 9, 3, 1, 13 and 14 demonstrated higher rates of hand distributions ($>10\%$).

****Insert Figure 6****

Figure 6. Passing recipient, height and body part profile of GK pass clusters.

3.4. Passing phase profile and other features

Passing phase distributions differed systematically across clusters, revealing distinct temporal patterns of goalkeeper involvement within possessions. Clusters 4 and 9 were strongly dominated by start-phase actions, with Clusters 2 and 12 also showing relatively high initiation involvement. In contrast, Clusters 7 and 8 exhibited similar profiles characterised by lower start-phase involvement and greater representation in

middle and late phases, whereas Clusters 3, 1, and 13 were associated with higher proportions of early and middle-phase actions, reflecting progressive build-up behaviour. Clusters 5 and 6 displayed a distinct dual-phase structure, with substantial involvement in both start and end phases, suggesting more direct distribution tendencies, while Clusters 10, 11, and 14 showed greater participation in late or terminal possession stages. Overall, most clusters were primarily associated with actions occurring in the start and early phases of possession, highlighting the prominent role of goalkeepers in initiating attacking sequences.

****Insert Figure 7****

Figure 7. Percentage of GK Passing Occur in Different Phase.

Cluster centroid feature profiles showed clear systematic differences in distribution behaviour. Early-start clusters (Cluster 4, 9, and 13) were characterised by short durations (1.34–1.46 s), low pressure exposure (0.06–0.13), short possession lengths (≈ 6.1 –7.8), and the lowest passing speeds (10.4–14.7). In contrast, later-stage clusters (Cluster 10 to 12 and Cluster 2) exhibited longer durations (2.21–2.67 s), substantially higher under-pressure rates (0.18–0.20), shorter possession lengths (≈ 4.7 –6.3), and higher speeds (17.0–18.8). Clusters 7 and 8 formed a distinct build-up profile, showing the longest possession lengths (8.83–9.09) with moderate durations (~ 1.60 –1.66 s) and mid-range speeds (~ 14.8 –15.6). Finally, Clusters 5 and 6 represented the most direct/terminal patterns, with the longest durations (3.22–3.41 s), high speeds (18.7–19.9), the shortest possession lengths (~ 4.37 –4.39), and the only non-zero through-ball rates (0.7–1.1%).

Insert Table 2

Table 2. Centroid characteristics of 14 GK passing clusters.

Interpretations of each passing cluster are provided in Table S1. Overall, these findings confirm that the clustering solution differentiated structurally distinct goalkeeper passing behaviours across spatial, temporal, and contextual dimensions.

4. Discussion

The aim of this study was to classify goalkeeper passing behaviours and examine their associations with game situations. The continuous gradient structure observed in the PCA projection suggests that goalkeeper passing behaviours are organised along continua rather than discrete categories, consistent with ecological dynamics perspectives in which coordinated actions emerge from interacting task and environmental constraints (Araújo & Davids, 2016; Newell, 1986). Observations nearer the origin reflected configurations aligned with dominant variance patterns, whereas peripheral clusters represented greater departures (Jolliffe, 2004). Overall, goalkeeper distributions typically originated from central defensive areas and were directed towards structured outlets across defensive and middle thirds, reflecting organised build-up behaviour.

Clusters 13 and 2 were located near the global centroid, representing passing configurations with limited directional specificity. Cluster 13 comprised central distributions from the penalty area to centre-backs or defensive midfielders, positions typically associated with high betweenness centrality in passing networks (Clemente, 2014), suggesting the transfer of build-up responsibility to outfield players. In

contrast, Cluster 2 involved centrally oriented distributions towards midfield and attacking zones under moderate-to-high pressure, promoting vertical progression to relieve defensive constraints (González-Rodenas et al., 2022; Sarmiento et al., 2014; Araújo & Davids, 2016). These patterns illustrate continuous perception–action regulation shaped by spatial and pressure constraints.

Totally 12 passing clusters demonstrated clear spatial mirror correspondences across pitch sides, although several pairs (Clusters 4&9, 3&1, and 7&8) showed functional similarity with weaker spatial symmetry. Such directional asymmetries are consistent with interactions between functional lateralisation and spatial affordances in goal-directed behaviour (Fajen et al., 2008; Vilar et al., 2012). In this dataset, 81.5% of goalkeeper passes were executed with the right foot, reflecting preferred-foot dominance in football (Carey et al., 2001; Stöckel & Carey, 2016). In low-position distributions (Clusters 4&9 and 3&1), goalkeepers under defensive pressure tended to distribute more frequently towards the dominant side, with right-oriented passes typically longer and faster than left-directed actions. From a practical perspective, these findings suggest that goalkeeper distribution should be evaluated in relation to foot preference and directional distribution preference rather than relying solely on overall passing statistics. Assessing dominant- and non-dominant-side behaviours separately may reveal important differences in decision-making and spatial use under defensive pressure. In practice, goalkeeper foot preference may also contribute to directional biases in team build-up, shaping preferred passing channels and attacking structures. Coaches and performance analysts should therefore consider

how consistent dominant-side distributions influence team organisation, predictability, and vulnerability to pressing, while encouraging greater adaptability in non-dominant-side distribution.

A related pattern emerged in higher-position distributions (Clusters 7 and 8, 14 and 10, 11 and 12, and 5 and 6), where left-directed passes were associated with greater switching behaviour and lower pressure exposure than right-oriented actions. This asymmetry aligns with the temporal demands of possession organisation under defensive pressure, whereby spatial constraints in one area promote circulation towards less constrained or advanced zones (Goes et al., 2022; González-Rodenas et al., 2022; Sarmiento et al., 2014). Overall, these differences reflect context-dependent behavioural regulation shaped by spatial and task constraints, with dominant-side distributions facilitating more direct progression and opposite-side actions supporting spatial reorganisation (Araújo & Davids, 2016; Goes et al., 2022). From a performance analysis perspective, these findings suggest that goalkeeper distribution should be evaluated in relation to directional context, pressure conditions, and spatial constraints rather than through aggregated passing metrics alone. Incorporating contextual indicators such as pressure exposure, switching behaviour, and spatial progression may provide a more accurate assessment of goalkeeper decision-making and tactical contribution during possession phases.

The spatial distribution of clusters suggest that goalkeepers are not solely responsible for unconstrained ball distribution but may also remain involved in possession sequences under pressure, contributing to early build-up phases. Most of

passes were made by the keepers' feet, which aligns with recent research on elite women's football goalkeepers, which reports that goalkeeper distribution plays an active role in offensive team performance and in transitional build-up situations (Casal et al., 2023). Furthermore, technical and tactical analyses of female international goalkeepers have demonstrated significant offensive involvement in open play, with successful teams' goalkeepers displaying higher passing participation across pitch zones (Sainz de Baranda et al., 2019). In line with broader analyses of modern goalkeeper roles, this supports the notion that contemporary goalkeepers are increasingly integrated into structured team possession rather than functioning solely as defensive specialists (Scholefield & Fitzpatrick, 2025).

Passing phase analysis showed that goalkeeper distributions were predominantly concentrated in the early stages of possession, typically representing the first action following restarts or the initial organisation of attacking play. In contrast, Clusters 5 and 6 occurred more frequently in the later phases of possession, highlighting goalkeeper involvement when teams reorganise defensively or respond to pressure through back passes or clearance-type actions. This pattern reinforces the goalkeeper's functional role in regulating transitions in possession, particularly at critical moments of initiation and resolution rather than during sustained attacking sequences (Anguera et al., 2022; Sarmiento et al., 2014).

Clusters 7 and 8, together with the mirrored pair Clusters 14 and 10, represented the primary passing configurations observed during mid-possession phases and were associated with more advanced spatial contexts. Clear differences emerged in the

situational constraints governing these actions: Clusters 7 and 8 occurred under relatively low defensive pressure, allowing greater freedom of ball control and spatial organisation, whereas Clusters 14 and 10 were characterised by higher pressure exposure and more constrained action spaces. These latter clusters appear to reflect context-specific tactical responses, such as retaining possession under pressure or manipulating defensive structure to create attacking space. The greater dispersion and increased distance from the cluster centroid observed in Clusters 14 and 10 further emphasise their specialised nature, indicating highly adaptive behavioural responses to spatial and defensive constraints during possession organisation (Goes et al., 2022).

Overall, the passing phase analysis of each cluster reflects the specific tactical role of the goalkeeper. Given positional constraints and tactical demands, goalkeepers in this sample appear more involved in maintaining possession stability than in shaping attacking progression. This interpretation is consistent with findings in elite women's football showing that goalkeeper actions rarely lead directly to shots or goals (Casal et al., 2023). Although modern goalkeepers are increasingly expected to contribute to build-up play, their influence remains distinct from that of outfield attacking players (Falcão et al., 2026; Scholefield & Fitzpatrick, 2025).

This study offers several practical implications for coaching and performance analysis. Phase-specific patterns indicate that goalkeepers are most involved during possession initiation and pressure-related situations, suggesting that training tasks should replicate these transitional contexts to develop decision-making under defensive constraints. Observed lateral tendencies further suggest interactions

between dominant-side preference and spatial conditions, informing both self- and opposition-analysis as well as the development of bilateral technical competence. The finding that goalkeeper passing was more closely associated with possession stability than sustained attacking progression may also help clarify tactical expectations and align training objectives with positional demands in elite women's football.

Several limitations should be acknowledged. First, the use of event data restricts insight into the perceptual information available at the moment of decision-making. Integrating tracking data would allow greater consideration of opponent positioning, spatial relations, and team structure. Second, although contextual variables were included, clustering may still be influenced by spatial dominance, potentially emphasising territorial patterns over underlying decision processes. Future work could reduce reliance on spatial descriptors and link passing patterns to subsequent possession outcomes. Third, while the sample represents elite women's football, it reflects league-specific contexts and may not generalise across competitions. Finally, the study describes structural patterns rather than performance effectiveness; future research should examine how distribution types relate to progression, chance creation, or possession retention.

5. Conclusion

This study used a graph-based approach to characterise goalkeeper passing behaviour in elite women's football, revealing that distributions are organised along continuous spatial and contextual gradients rather than discrete categories. By representing pass embeddings as a structured graph, we identified distinct clusters influenced by

460 defensive pressure, lateral tendencies, and the phase of possession, highlighting how
461 goalkeepers adapt their distribution patterns in relation to game context.

462 The graph analysis showed that goalkeepers exert the greatest influence during
463 transitional phases, especially at possession initiation and termination, and appear
464 more involved in stabilising possession than in driving sustained attacking sequences.

465 Lateral biases and pressure-driven adaptations were also evident across cluster
466 relationships, underscoring the value of graph-based methods in capturing nuanced
467 behavioural patterns. Overall, this work demonstrates the utility of graph-based
468 representation for exploring complex passing behaviours and provides a foundation
469 for future efforts to link structural patterns with underlying decision-making processes
470 and performance outcomes within elite football.

471 **References**

- 472 Anguera, M. T., De Conti, G., Pic, M., & Santos, F. (2022). *T-pattern analysis of offensive*
473 *and defensive actions of youth football goalkeepers.*
- 474 Anzer, G., Bauer, P., Brefeld, U., Fassmeyer, D., & Berlin, H. B. (2022). *Detection of tactical*
475 *patterns using semi-supervised graph neural networks.* <https://bit.ly/3o2zIQ5>.
- 476 Araújo, D., Brito, H., & Carrilho, D. (2022). Team decision-making behavior: An ecological
477 dynamics approach. *Asian Journal of Sport and Exercise Psychology*, 3(1), 24–
478 29. <https://doi.org/10.1016/j.ajsep.2022.09.005>
- 479 Araújo, D., & Davids, K. (2016). Team synergies in sport: Theory and measures. *Frontiers in*
480 *Psychology*, 7(SEP), 1–13. <https://doi.org/10.3389/fpsyg.2016.01449>
- 481 Araújo, D., Davids, K., & Hristovski, R. (2006). The ecological dynamics of decision making
482 in sport. *Psychology of Sport and Exercise*, 7(6), 653–
483 676. <https://doi.org/10.1016/j.psychsport.2006.07.002>
- 484 Ayala, F., López-Valenciano, A., Gámez Martín, J. A., De Ste Croix, M., Vera-Garcia, F.,
485 García-Vaquero, M., Ruiz-Pérez, I., & Myer, G. (2019). A Preventive Model for
486 Hamstring Injuries in Professional Soccer: Learning Algorithms. *International Journal*
487 *of Sports Medicine*, 40(05), 344–353. <https://doi.org/10.1055/a-0826-1955>
- 488 Beamish, D., Bhatti, S. A., MacKenzie, I. S., & Wu, J. (2006). Fifty years later: a
489 neurodynamic explanation of Fitts' law. *Journal of The Royal Society Interface*, 3(10),
490 649–654. <https://doi.org/10.1098/rsif.2006.0123>

- 491 Bialkowski, A., Lucey, P., Carr, Peter, Yue, Y., Matthews, I., Carr, P., & Ram,
 492 F. (2014). “Win at Home and Draw Away”: Automatic Formation Analysis
 493 *Highlighting the Differences in Home and Away Team Behaviors*.
- 494 Bradley, Paul & A, Dellal & Mohr, Magni & Castellano, Julen & Wilkie, Anna. (2013).
 495 Gender differences in match performance characteristics of soccer players competing in
 496 the UEFA Champions League. *Human movement science*. 33.
 497 10.1016/j.humov.2013.07.024.
- 498 Brooks, J., Kerr, M., & Gutttag, J. (2016). Using machine learning to draw inferences from
 499 pass location data in soccer. *Statistical Analysis and Data Mining: The ASA Data*
 500 *Science Journal*, 9(5), 338–349. <https://doi.org/10.1002/sam.11318>
- 501 Buldu, J. M., Busquets, J., Echegoyen, I., & Seirul. lo, F. (2019). Defining a historic football
 502 team: Using Network Science to analyze Guardiola’s FC Barcelona. *Scientific*
 503 *Reports*, 9(1), 13602.
- 504 Campbell, P. G., Stewart, I. B., Sirotic, A. C., Drovandi, C., Foy, B. H., & Minett, G.
 505 M. (2021). Analysing the predictive capacity and dose-response of wellness in load
 506 monitoring. *Journal of Sports Sciences*, 39(12), 1339–
 507 1347. <https://doi.org/10.1080/02640414.2020.1870303>
- 508 Carey, D. P., Smith, G., Smith, D. T., Shepherd, J. W., Skriver, J., Ord, L., & Rutland,
 509 A. (2001). Footedness in world soccer: an analysis of France ’98. *Journal of Sports*
 510 *Sciences*, 19(11), 855–864. <https://doi.org/10.1080/026404101753113804>
- 511 Casal, C., Stone, J., Iván-Baragaño, I., & Losada, J. (2023). Effect of goalkeepers’ offensive
 512 participation on team performance in the women Spanish La Liga: a multinomial

- 513 logistic regression analysis. *Biology of Sport*, 3–
 514 13. <https://doi.org/10.5114/biolSport.2024.125592>
- 515 Castellano, J., Casamichana, D., & Lago, C. (2012). The Use of Match Statistics that
 516 Discriminate Between Successful and Unsuccessful Soccer Teams. *Journal of human*
 517 *kinetics*, 31, 139–147. <https://doi.org/10.2478/v10078-012-0015-7>
- 518 Cherappurath, N., Shamshadali, P., Elayaraja, M., & KI, D. A. K. (2024). Mapping the field:
 519 A bibliometric analysis of women's football research trends and future directions.
 520 *Apunts Sports Medicine*, 59(223), 100448.
- 521 Cho, H., Ryu, H., & Song, M. (2022). Pass2vec: Analyzing soccer players' passing style
 522 using deep learning. *International Journal of Sports Science & Coaching*, 17(2), 355–
 523 365. <https://doi.org/10.1177/17479541211033078>
- 524 Clemente, F. M., Martins, F. M. L., Kalamaras, D., Wong, D. P., & Mendes, R. S. (2015).
 525 General network analysis of national soccer teams in Fifa World Cup
 526 2014. *International Journal of Performance Analysis in Sport*, 15(1), 80–
 527 96. <https://doi.org/10.1080/24748668.2015.11868778>
- 528 Datson, N., Drust, B., Weston, M., Jarman, I. H., Lisboa, P. J., & Gregson, W. (2017). Match
 529 Physical Performance of Elite Female Soccer Players During International Competition.
 530 *Journal of strength and conditioning research*, 31(9), 2379–2387.
 531 <https://doi.org/10.1519/JSC.0000000000001575>
- 532 Davids, K., & Araújo, D. (2010). The concept of 'Organismic Asymmetry' in sport
 533 science. *Journal of Science and Medicine in Sport*, 13(6), 633–
 534 640. <https://doi.org/10.1016/j.jsams.2010.05.002>

- 535 de Baranda, P. S., Adán, L., García-Angulo, A., Gómez-López, M., Nikolic, B., & Ortega-
 536 Toro, E. (2019). Differences in the offensive and defensive actions of the goalkeepers at
 537 women's FIFA world cup 2011. *Frontiers in Psychology*, 10(FEB), 1–
 538 10. <https://doi.org/10.3389/fpsyg.2019.00223>
- 539 Decroos, T., Van Haaren, J., Bransen, L., & Davis, J. (2019). Actions speak louder than goals:
 540 Valuing player actions in soccer. *Proceedings of the ACM SIGKDD International*
 541 *Conference on Knowledge Discovery and Data Mining*, 1851–
 542 1861. <https://doi.org/10.1145/3292500.3330758>
- 543 Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster analysis (5th ed.). Wile
- 544 Fajen, B. R., Riley, M. A., & Turvey, M. T. (2008). Information, affordances, and the control
 545 of action in sport. *International Journal of Sport Psychology*, 40(1), 79–107.
- 546 Falcão, H., Pompeo, A., Afonso, J., Cirillo, E., da Costa, I. T., Machado, G., González-
 547 Villora, S., & Casanova, F. (2026). The tactical profile of goalkeepers in Football: A
 548 scoping review with evidence gap-map. *Football Studies*, 1,
 549 100006. <https://doi.org/10.1016/j.footst.2025.100006>
- 550 FIFA. (2019). *FIFA Women's World Cup 2019™ watched by more than 1 billion*. Fédération
 551 Internationale de Football
 552 Association. [https://www.fifa.com/tournaments/womens/womensworldcup/france2019/n](https://www.fifa.com/tournaments/womens/womensworldcup/france2019/news/fifa-women-s-world-cup-2019tm-watched-by-more-than-1-billion)
 553 [ews/fifa-women-s-world-cup-2019tm-watched-by-more-than-1-billion](https://www.fifa.com/tournaments/womens/womensworldcup/france2019/news/fifa-women-s-world-cup-2019tm-watched-by-more-than-1-billion)
- 554 FIFA. (2023). *Setting The Pace – FIFA Women's Football Benchmarking Report*. Inside
 555 FIFA.

- 556 Grund, T. U. (2012). Network structure and team performance: The case of English Premier
 557 League soccer teams. *Social Networks*, 34(4), 682–690.
 558 <https://doi.org/10.1016/j.socnet.2012.08.004>
- 559 Hamilton, W. L., Ying, R., & Leskovec, J. (2018). *Inductive Representation Learning on*
 560 *Large Graphs*. <http://arxiv.org/abs/1706.02216>
- 561 Hennig, C. (2007) Cluster-wise assessment of cluster stability. *Computational Statistics &*
 562 *Data Analysis*, 52(1), 258–271. <https://doi.org/10.1016/j.csda.2006.11.025>
- 563 Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C., & Meyer, T. (2019). Machine
 564 learning in men’s professional football: Current applications and future directions for
 565 improving attacking play. *International Journal of Sports Science & Coaching*, 14(6),
 566 798–817. <https://doi.org/10.1177/1747954119879350>
- 567 Hewitt, A., Greenham, G., & Norton, K. (2016). Game style in soccer: What is it and can we
 568 quantify it? *International Journal of Performance Analysis in Sport*, 16(1), 355–
 569 372. <https://doi.org/10.1080/24748668.2016.11868892>
- 570 Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition*
 571 *Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- 572 Jolliffe, I. T. . (2004). *Principal component analysis*. Springer.
- 573 Kasap, S., & Kasap, N. (2005). Development of a Database and Decision Support System for
 574 Performance Evaluation of Soccer Players. *35th International Conference on Computers*
 575 *and Industrial Engineering*, (April 2015), 1097–1102. www.fstats.net
- 576 Klemp, M., Bassek, M., Garnica Caparròs, M., Bakhtiar, L. A., & Memmert, D. (2026). The
 577 use of machine learning in performance analysis in invasion games: Umbrella review of

- 578 reviews. *Journal of Sports Sciences*, 1-28.
- 579 <https://doi.org/10.1080/02640414.2026.2636863>
- 580 Kipf, T. N., & Welling, M. (2017). *Semi-Supervised Classification with Graph Convolutional*
- 581 *Networks*. <http://arxiv.org/abs/1609.02907>
- 582 Lago-Peñas, C., & Dellal, A. (2010). Ball possession strategies in elite soccer according to the
- 583 evolution of the match-score: The influence of situational variables. *Journal of Human*
- 584 *Kinetics*, 25(1), 93–100. <https://doi.org/10.2478/v10078-010-0036-z>
- 585 Lago-Peñas, C., Lago-Ballesteros, J., Dellal, A., & Gómez, M. (2010). Game-Related
- 586 Statistics that Discriminated Winning, Drawing and Losing Teams from the Spanish
- 587 Soccer League. *Journal of Sports Science & Medicine*, 9(2), 288–293.
- 588 Lethole, L., Kubayi, A., Toriola, A., Larkin, P., & Stone, J. A. (2024). ‘Goalkeepers are players
- 589 too’: key attributes coaches’ look for in talented youth soccer goalkeepers. *BMC Sports*
- 590 *Science, Medicine and Rehabilitation*, 16(1), 210.
- 591 Li, Y., Zong, S., Shen, Y., Pu, Z., Gómez, M.-Á., & Cui, Y. (2022). Characterizing player’s
- 592 playing styles based on player vectors for each playing position in the Chinese Football
- 593 Super League. *Journal of Sports Sciences*, 40(14), 1629–
- 594 1640. <https://doi.org/10.1080/02640414.2022.2096771>
- 595 Liu, H., Gomez, M. Á., Lago-Peñas, C., & Sampaio, J. (2015). Match statistics related to
- 596 winning in the group stage of 2014 Brazil FIFA World Cup. *Journal of sports sciences*,
- 597 33(12), 1205–1213. <https://doi.org/10.1080/02640414.2015.1022578>

- 598 McLean, S., Salmon, P. M., Gorman, A. D., Read, G. J. M., & Solomon, C. (2017). What's in
 599 a game? A systems approach to enhancing performance analysis in football. *PLoS*
 600 *ONE*, 12(2), 1–15. <https://doi.org/10.1371/journal.pone.0172565>
- 601 Memmert, D., Lemmink, K. A. P. M., & Sampaio, J. (2017). Current Approaches to Tactical
 602 Performance Analyses in Soccer Using Position Data. *Sports Medicine*, 47(1), 1–
 603 10. <https://doi.org/10.1007/s40279-016-0562-5>
- 604 Newell, K. M. (1986). Constraints on the Development of Coordination. In *Motor*
 605 *Development in Children: Aspects of Coordination and Control* (pp. 341–360). Springer
 606 Netherlands. https://doi.org/10.1007/978-94-009-4460-2_19
- 607 Okholm Kryger, K., Wang, A., Mehta, R., Impellizzeri, F. M., Massey, A., & McCall,
 608 A. (2022). Research on women's football: a scoping review. *Science and Medicine in*
 609 *Football*, 6(5), 549–558. <https://doi.org/10.1080/24733938.2020.1868560>
- 610 Oliver, J. L., Ayala, F., De Ste Croix, M. B. A., Lloyd, R. S., Myer, G. D., & Read, P.
 611 J. (2020). Using machine learning to improve our understanding of injury risk and
 612 prediction in elite male youth football players. *Journal of Science and Medicine in*
 613 *Sport*, 23(11), 1044–1048. <https://doi.org/10.1016/j.jsams.2020.04.021>
- 614 Peña, J. L., & Touchette, H. (2012). *A network theory analysis of football*
 615 *strategies*. <http://arxiv.org/abs/1206.6904>
- 616 Pfister, G. (2015). Assessing the sociology of sport: On women and football. *International*
 617 *Review for the Sociology of Sport*, 50(4–5), 563–
 618 569. <https://doi.org/10.1177/1012690214566646>

- Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1410. <https://doi.org/10.1186/s40064-016-3108-2>
- Rein, R., Raabe, D., & Memmert, D. (2017). “Which pass is better?” Novel approaches to assess passing effectiveness in elite soccer. *Human Movement Science*, 55(August), 172–181. <https://doi.org/10.1016/j.humov.2017.07.010>
- Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F., & 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>
- Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology*, 2(MAR), 1–7. <https://doi.org/10.3389/fpsyg.2011.00038>
- Rousseeuw, P. J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Sarmiento, H., Anguera, M. T., Pereira, A., & Araújo, D. (2018). Talent Identification and Development in Male Football: A Systematic Review. *Sports Medicine*, 48(4), 907–931. <https://doi.org/10.1007/s40279-017-0851-7>
- Sarmiento, H., Clemente, F. M., Araújo, D., Davids, K., McRobert, A., & Figueiredo, A. (2018). 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>

- 641 Scholefield, P., & Fitzpatrick, A. (2025). The technical and tactical demands and capabilities
 642 of six English Category 1 academy goalkeepers: A case study. *International Journal of*
 643 *Sports Science & Coaching*, 20(6), 2670–
 644 2680. <https://doi.org/10.1177/17479541251367296>
- 645 Seaton, M., & Campos, J. (2011). Distribution competence of a football clubs
 646 goalkeepers. *International Journal of Performance Analysis in Sport*, 11(2), 314–
 647 324. <https://doi.org/10.1080/24748668.2011.11868551>
- 648 Sheng, B., Li, P., Zhang, Y., Mao, L., & Chen, C. L. P. (2021). GreenSea: Visual Soccer
 649 Analysis Using Broad Learning System. *IEEE Transactions on Cybernetics*, 51(3),
 650 1463–1477. <https://doi.org/10.1109/TCYB.2020.2988792>
- 651 Shutaywi, M., & Kachouie, N. N. (2021). Silhouette Analysis for Performance Evaluation in
 652 Machine Learning with Applications to Clustering. *Entropy*, 23(6), 759.
 653 <https://doi.org/10.3390/e23060759>
- 654 Stöckel, T., & Carey, D. P. (2016). Laterality Effects on Performance in Team Sports.
 655 In *Laterality in Sports* (pp. 309–328). Elsevier. [https://doi.org/10.1016/B978-0-12-](https://doi.org/10.1016/B978-0-12-801426-4.00014-6)
 656 [801426-4.00014-6](https://doi.org/10.1016/B978-0-12-801426-4.00014-6)
- 657 Stöckl, M., Seidl, T., Marley, D., & Power, P. (2021). *Making Offensive Play Predictable-*
 658 *Using a Graph Convolutional Network to Understand Defensive Performance in*
 659 *Soccer*.
- 660 Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010). Effect of playing tactics on
 661 achieving score-box possessions in a random series of team possessions from Norwegian

- 662 professional soccer matches. *Journal of Sports Sciences*, 28(3), 245–
 663 255. <https://doi.org/10.1080/02640410903502766>
- 664 Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data
 665 set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical*
 666 *Methodology)*, 63(2), 411–423. <https://doi.org/10.1111/1467-9868.00293>
- 667 Travassos, B., Davids, K., Araújo, D., & Esteves, P. T. (2013). Performance analysis in team
 668 sports: Advances from an ecological dynamics approach. *International Journal of*
 669 *Performance Analysis in Sport*, 13(1), 83–
 670 95. <https://doi.org/10.1080/24748668.2013.11868633>
- 671 Ullmann, T., Hennig, C., & Boulesteix, A.-L. (2022). Validation of cluster analysis results on
 672 validation data: A systematic framework. *WIREs Data Mining and Knowledge*
 673 *Discovery*, 12(3), e1444. <https://doi.org/10.1002/widm.1444>
- 674 Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lì, P., & Bengio,
 675 Y. (2018). *GRAPH ATTENTION NETWORKS*.
- 676 Vilar, L., Araújo, D., Davids, K., & Button, C. (2012). The role of ecological dynamics in
 677 analysing performance in team sports. *Sports medicine (Auckland, N.Z.)*, 42(1), 1–10.
 678 <https://doi.org/10.2165/11596520-000000000-00000>
- 679 Wang, B., Chen, J., Li, C., Zhou, S., Shi, Q., Gao, Y., Feng, Y., Chen, C., & Wang, C. (2024).
 680 Distributionally Robust Graph-based Recommendation System. *WWW 2024 -*
 681 *Proceedings of the ACM Web Conference*, 3777–
 682 3788. <https://doi.org/10.1145/3589334.3645598>

- 683 Yu, Q., Gai, Y., Gong, B., Gómez, M.-Á., & Cui, Y. (2020). Using passing network measures
684 to determine the performance difference between foreign and domestic outfielder players
685 in Chinese Football Super League. *International Journal of Sports Science &*
686 *Coaching*, 15(3), 398–404. <https://doi.org/10.1177/1747954120905726>
- 687 Zhao, S., Jiang, Y., Zhang, Z., Yu, Y., & Chen, H. (2025). *GRAIN: Multi-Granular and*
688 *Implicit Information Aggregation Graph Neural Network for Heterophilous*
689 *Graphs*. <http://arxiv.org/abs/2504.06649>

690 **Table 1.** Variables and Definition

Metric	Definition	Range of data
Start area	A calculated area by XY value of where the player executed the pass.	[1, 15]
End area	A calculated area by XY value of where the passing event ended.	[1, 15]
Pass angle	The angle in radian of pass.	$[-\pi, \pi]$
Pass height	Classification labels for pass height.	1. Ground pass: balls on the ground 2. Medium pass: balls above ground but not higher than head 3. High pass: pass over head.
Pass body part	Classification labels for the body part that the player used to execute the pass.	0. No touch 1. Drop kick 2. Keeper Arm 3. Foot 4. Other
Pass recipient position	The position of player who receive the pass	0. No recipients. 1. Keeper 2. Centre Back 3. Side Back 4. Winger Back 5. Defensive Midfield 6. Midfield 7. Side Midfield 8. Attacking Midfield 9. Winger 10. Forward
Under Pressure	Boolean data to present if the player is under pressure or not, when the player was executing	[0, 1] True or False

	passes.	
Pass Switch	Boolean data to present if the player is executing switching pass.	[0, 1] True or False
Pass Through Ball	Boolean data to present if the player is executing through ball pass.	[0, 1] True or False
Pass Phase	The scaled order number of the pass in a pass sequence.	[0, 1]
Speed	The speed of pass.	[0, 30]
Pass Length	Length of pass	[0, 120]

692 **Table 2.** Passing features profile of GK pass clusters' centroid.

Cluster	Duration	Pass Length	Pass Switch	Pass Through Ball	Under Pressure	Speed
4	1.417	12.324	0.000	0.000	0.059	10.416
9	1.462	13.984	0.000	0.000	0.093	11.453
3	1.681	18.710	0.000	0.000	0.102	13.294
1	1.688	18.988	0.003	0.000	0.117	13.131
13	1.337	13.128	0.000	0.000	0.135	14.748
7	1.660	21.285	0.005	0.000	0.084	14.803
8	1.602	21.074	0.002	0.000	0.103	15.620
14	2.214	30.743	0.060	0.000	0.195	15.910
10	2.230	33.073	0.037	0.000	0.200	17.017
11	2.533	40.527	0.043	0.000	0.161	18.200
12	2.669	43.910	0.013	0.000	0.182	18.828
2	2.487	40.726	0.001	0.000	0.181	18.834
5	3.219	53.068	0.024	0.007	0.110	19.908
6	3.406	53.707	0.009	0.011	0.174	18.675

694 **Table S1.** Archtypes of GK pass clustering.

Archtypes	Cluster	Key features
Low-position build up	4	Short ground passes from the goal area towards centre-backs on either side of the pitch. Actions occur predominantly in the start phase, with the shortest durations, lowest
	9	passing speeds, minimal pressure exposure, and short possession sequences.
Low-position wide build up	3	Ground passes directed diagonally towards full-backs and midfield areas. Actions occur mainly in early and middle phases, with moderate duration, speed, and possession
	1	length.
Short central build-up	13	Forward-oriented ground passes towards central defensive and midfield zones. Actions show moderate duration, moderate pressure exposure, and central receiving
		locations.
High-position organisation	7	Passes directed into midfield areas with the longest possession lengths. Actions show moderate duration and speed, with lower start-phase involvement and greater
	8	participation in middle and late phases.
Advanced wide build up	14	Longer actions with higher pressure exposure, increased use of high passes, and distribution towards advanced areas. Actions occur more frequently in later possession
	10	phases.
Advanced progression pass	11	Passes directed towards wide and forward areas with high passing speeds and relatively high pressure exposure. Actions occur primarily in later possession stages.

	12	
Direct long pass central	2	Forward high passes from the defensive third into advanced central areas, primarily received by forwards. Typically distributed at the start of possession, although relatively evenly represented across phases, using the foot (occasionally via drop-kick) under moderate-to-high pressure.
Direct long pass wide	5	Longest-duration actions with the highest passing speeds. Passes include a higher proportion of high balls and occasional through passes towards advanced areas. Actions occur in both start and end phases.

- 696 **Figure 1.** The Process of GraphSAGE Based Analytical Framework
- 697 **Figure 2.** Two-dimensional principal component projection illustrating the distribution
698 of K-means GK passing behaviours clusters.
- 699 **Figure 3.** Location profile of 14 clusters (Attacking Direction from left to right)
- 700 **Figure 4.** Location profile of all passes (Attacking Direction from left to right)
- 701 **Figure 5.** Pass angle of 14 clusters (Attacking Direction from left to right)
- 702 **Figure 6.** Passing recipient, height and body part profile of GK pass clusters.
- 703 **Figure 7.** Percentage of GK Passing Occur in Different Phase
- 704 **Figure S1.** Zones division of football pitch (Attacking Direction from left to right)
- 705 **Figure S2.** PCA1 and PCA 2 driver of clusters.