Nusantara Journal of Behavioral and Social Science. 2025; 4(1): 51-64
ISSN 2828-1039 (print), ISSN 2828-0733 (online)
https://doi.org/10.47679/njbss.202573

RESEARCH ARTICLE

A Two-Stage Spatio-Geometrical clustering of football team shape for post-match review

Ali Zare Zardiny [1*], Zahra Bahramian[2)]

Published online: 29 March 2025

Abstract: The increasing availability of spatio-temporal data in football has enabled advanced Post-Match Review techniques that analyze team performance from both individual and collective perspectives. While traditional approaches focus on individual player metrics, this study introduces a novel spatio-geometrical method for analyzing team shape dynamics. The proposed approach defines the team shape as a convex hull at each time frame, capturing its overall spatial structure. A comprehensive set of spatial, geometric, zone-based, and event-based descriptors is extracted to quantify the team′s shape and movement patterns. A two-stage clustering framework is employed to categorize team behavior. First, spatial clustering identifies broad positioning trends based on pitch location and zone overlap. Second, geometric clustering refines these clusters by analyzing shape variations, enabling the detection of distinct tactical patterns in both in-possession and out-of-possession scenarios. This process facilitates a data-driven interpretation of tactical strategies, helping analysts understand team behavior leading to goal-scoring opportunities, passing efficiency, and spatial control. The main contributions of this study include the development of a fully vector-based clustering approach that eliminates the need for computationally expensive image-processing techniques, the introduction of novel geometric descriptors tailored for team shape analysis, and the implementation of a two-stage clustering strategy that enhances the interpretability of tactical adjustments. The findings provide actionable insights for coaches and analysts, offering a quantitative framework for evaluating and optimizing team strategies.

Keywords: Spatio-temporal data, Post-Match Review, Team Shape, Clustering, Shape Descriptors, Football Analytics

INTRODUCTION

The advancement of technology in recent years has led to the collection of vast amounts of data in every football match. These data include player positions at each frame, ball trajectories, and event-based actions, all of which have a strong spatio-temporal nature. The availability of such large-scale data has significantly enhanced the potential for performing technical analyses that provide valuable insights into team dynamics and individual player performance (Gudmundsson & Horton, 2017; Liu et al., 2021). Football analysts and coaches can utilize these insights to optimize strategies, enhance player development, and improve overall team performance (Sarmento et al., 2014).

Various methodologies have been developed to analyze football team performance. These analyses can generally be categorized into two levels: micro-level and macro-level (Araújo et al., 2015). Micro-level analysis focuses on individual players' contributions, including running distances, speed, engagement in match events, and off-ball movement efficiency (Carling, Williams, & Reilly, 2005). Conversely, macro-level analysis examines the team as a collective entity, emphasizing aspects such as formation, pitch coverage, and team-wide movement (Zhang, 2022). While conventional macro-level studies predominantly rely on numerical metrics such as area, perimeter, and length-width ratios of team shape, fewer studies have explored the geometrical structure of the team beyond basic spatial descriptors. Most existing research also overlooks the impact of scale and orientation, despite their critical influence on spatial and tactical analysis.

This paper aims to conduct a Post-Match Review by evaluating the overall team behavior from a spatio-geometrical perspective. Unlike conventional approaches that analyze individual players' movements separately, this study defines the entire team shape as a single geometry. The analysis proceeds by first constructing the general shape of the team in each time frame using a convex hull representation. Following this, various spatial, geometric,

[1*, 2] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran

*) corresponding author

Ali Zare Zardiny
School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran

Email: zare_zardiny@ut.ac.ir

Submitted: 19 November 2024 Revised: 19 February 2025 Accepted: 27 February 2025

zone-based, and event-based descriptors are computed to quantify the team's formation and movement. The next step involves segmenting all frames into in-possession and out-of-possession states based on event descriptors. Subsequently, a two-stage clustering process is applied: in the first stage, team shapes are clustered according to pitch location and spatial overlap, while in the second stage, each main cluster is further divided based on geometric descriptors to identify sub-clusters representing distinct tactical patterns. Finally, the effectiveness of the identified clusters is examined in relation to match outcomes, including team strategy in goal-scoring opportunities, passing efficiency, challenge success rates, and time-space management.

The integration of spatio-temporal data in football has significantly advanced match analysis techniques (Bialkowski et al., 2014; Goncalves et al., 2019). However, existing studies typically rely on either image-processing techniques or simplified geometric models that fail to capture higher-order spatial relationships (Bueno et al., 2021). This study offers several key contributions. First, it introduces a novel clustering process based exclusively on vector data, eliminating the need for computationally expensive image processing techniques. Second, it proposes meaningful geometrical descriptors that better capture team shape variations. Third, the study employs a two-stage clustering framework, enhancing the interpretability of formation shifts and tactical adjustments. Lastly, the research develops performance indicators linking team shape behavior with match outcomes, providing actionable insights for coaches and analysts.

This paper is structured as follows: Section 2 presents a review of previous studies on team shape modeling and analysis. Section 3 introduces the proposed method for team shape description and spatio-geometrical clustering. Section 4 details the implementation of the proposed method and provides an interpretation of the results. Finally, Section 5 discusses conclusions and future research directions.

RESEARCH BACKGROUND

Post-match reviews analyzing the performance of the entire team or individual players provide coaches with valuable insights into game dynamics (Carling, Williams, & Reilly, 2005). These analyses support various aspects such as performance evaluation, tactical adjustments, player development, injury prevention, and team cohesion (Sarmento et al., 2014; Teixeira et al., 2021; Querido et al., 2022). With the increasing availability of spatio-temporal data, researchers have explored numerous methodologies for understanding team behavior from a geometrical and spatial perspective.

Zhang (2022) investigated team tactical behavior using player positional data, particularly in transitions between attack, defense, and possession states. This study utilized the convex hull method to define team shape and analyzed parameters such as geometric center, length, width, area, and displacement. Statistical analysis of these parameters over time revealed that teams exhibit significantly different spatial behaviors when in possession compared to out-of-possession states. During transitions from out-of-possession to in-possession, teams expand width more than length, whereas defensive transitions show an increased width compared to moments when possession is lost.

Shaw and Glickman (2020) proposed a hierarchical clustering approach for analyzing defensive and offensive team formations. Their study employed cumulative clustering of player positions to determine strategic patterns over time. Tactical summaries were generated using Bayesian model selection, enabling probability estimation for different tactical formations. An example of their clustering results demonstrated that team structures could be dynamically categorized into probable shapes during possession and defensive situations. An example of the created clusters is shown in Figure 1.

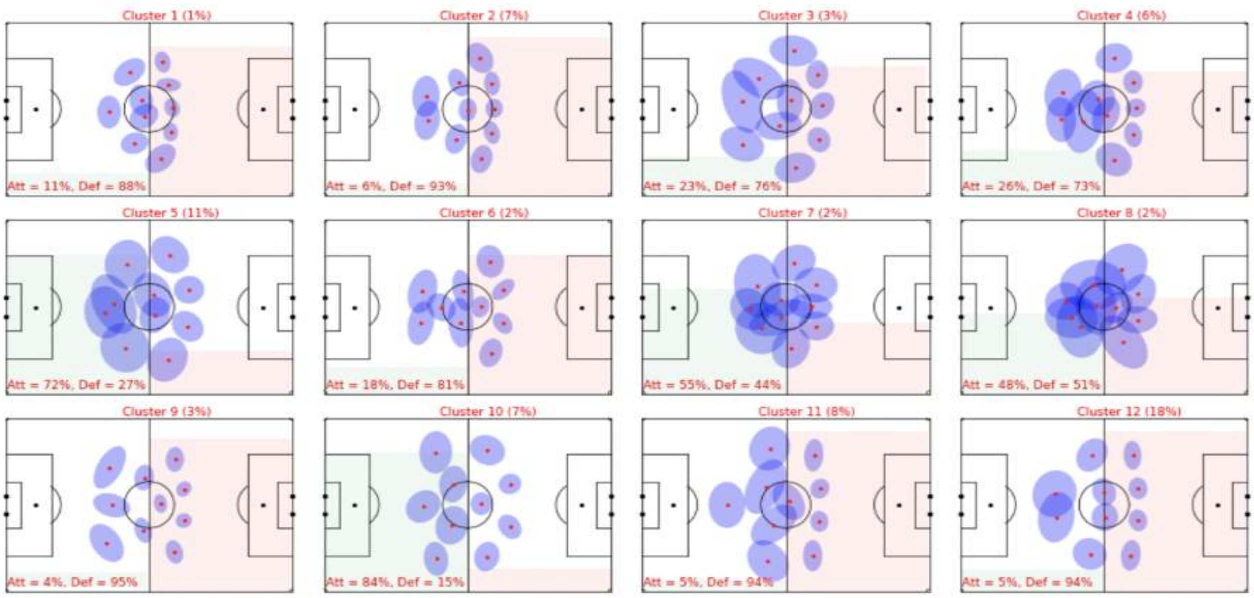

Figure 1. A view of the hierarchical clustering results performed in (Shaw and Glickman, 2020)

Bueno et al. (2021) introduced the Multiscale Fractal Curve (MFC) approach for shape-based team analysis. Their method defines team shape via convex hulls, which are then transformed into raster images for clustering using the K-Means algorithm. However, this approach lacks consideration of locational parameters and focuses solely on geometric clustering. Additionally, as MFC is scale- and rotation-invariant, it does not account for tactical variations in playstyle or quality. The study provides valuable geometric clustering insights but is limited in capturing dynamic team behavior in match contexts. Figure 2 shows a view of the clustering performed in this reseach.

Further research by Whitmore and Seidl (2021) examined team formations under in-possession and out-of-possession states. They identified the most probable team shapes based on average player positions over time. By clustering these shapes at various time intervals, they determined frequent formations that teams adopt in different match conditions. Based on the results of clustering, the most probable team shapes have been chosen from the patterns presented in Figure 3.

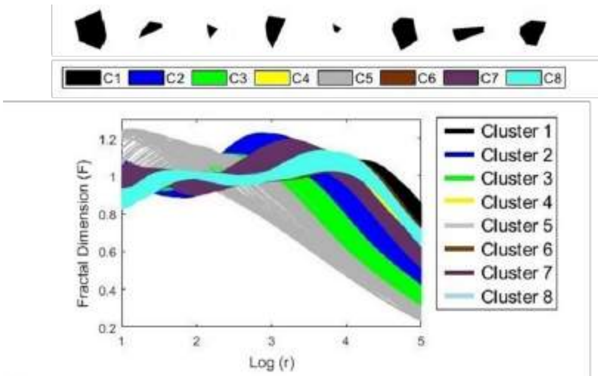

Figure 2. A view of the clustering results in (Bueno et al., 2021)

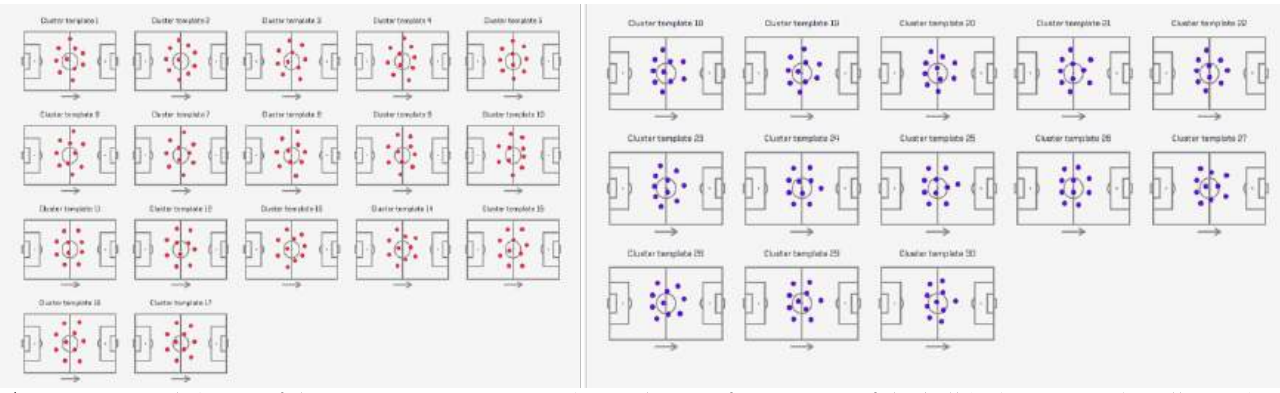

Figure 3. General shapes of the team: a. In Possession shapes, b. Out of possession of the ball (Whitmore and Seidl, 2021)

Other studies have emphasized cumulative spatial functions in tactical analysis. Goes et al. (2020) reviewed methodologies that utilize player location data to evaluate team performance, highlighting the importance of simplified spatial models for interpretability. Parameters such as the geometric center, the average distance of players to the center, team shape area, and length-width ratios were identified as key factors in spatial-temporal analysis.

Narizuka and Yamazaki (2019) applied hierarchical clustering and time-series analysis to assess team shape variations during attack and defense. Using Voronoi diagrams, they quantified controlled space and examined its impact on team positioning across multiple matches. Goncalves et al. (2019) utilized convex hulls to define effective team space and analyzed positional trends such as ball location, movement distances, length-width distribution, and ball possession. Their findings indicated that ball possession significantly influences collective team behavior, differentiating successful teams from lower-ranked counterparts.

Bialkowski et al. (2014) conducted large-scale spatio-temporal clustering of player tracking data, applying K-Means clustering to categorize player formations in offensive and defensive phases. Their study revealed that automated clustering methods effectively distinguish between tactical configurations and player positioning patterns. Figure 4 shows a view of the output of this clustering according to the offensive and defensive situations of the team.

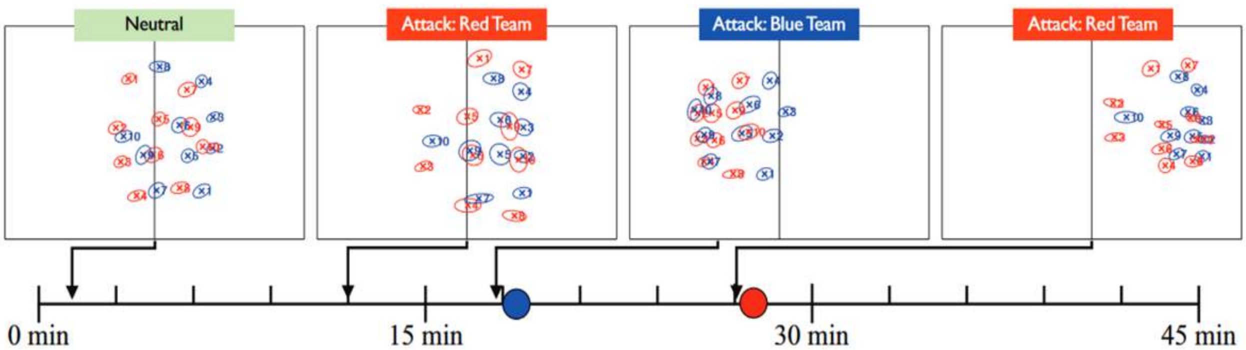

Figure 4. Clustering of players in different situations (Bialkowski et al., 2014)

Despite these advancements, macro-level analyses predominantly rely on numerical metrics such as area, perimeter, and length-width ratios while giving less attention to the geometrical aspects of team shape. Many studies analyzing team geometry have disregarded scale and orientation, despite these being critical elements in football strategy. Addressing these challenges, this paper introduces a novel geometric descriptor and examines zone-based spatial parameters to provide a more detailed understanding of team behavior.

METHOD

Research Design

For post-match review of a team from a geometric and spatial point of view, spatio-temporal data related to that match is needed. Generally, these data include three components: players′ locations on the pitch, ball locations at each time frame, and event data (Liu, 2022). With the advancement of technology, these data can be collected using different methods such as GPS tracking systems (Zhang, 2022), optical tracking systems with video processing (Csanalosi et al., 2020), and radio tracking systems using electrical signals (Seidl, 2016).

Before starting the analysis, data preprocessing must be performed to ensure data integrity. This includes removing data points outside the pitch boundaries, standardizing the field orientation across match halves, and handling substitutions by replacing player locations with their replacements. Once preprocessing is complete, the team shape is constructed and analyzed.

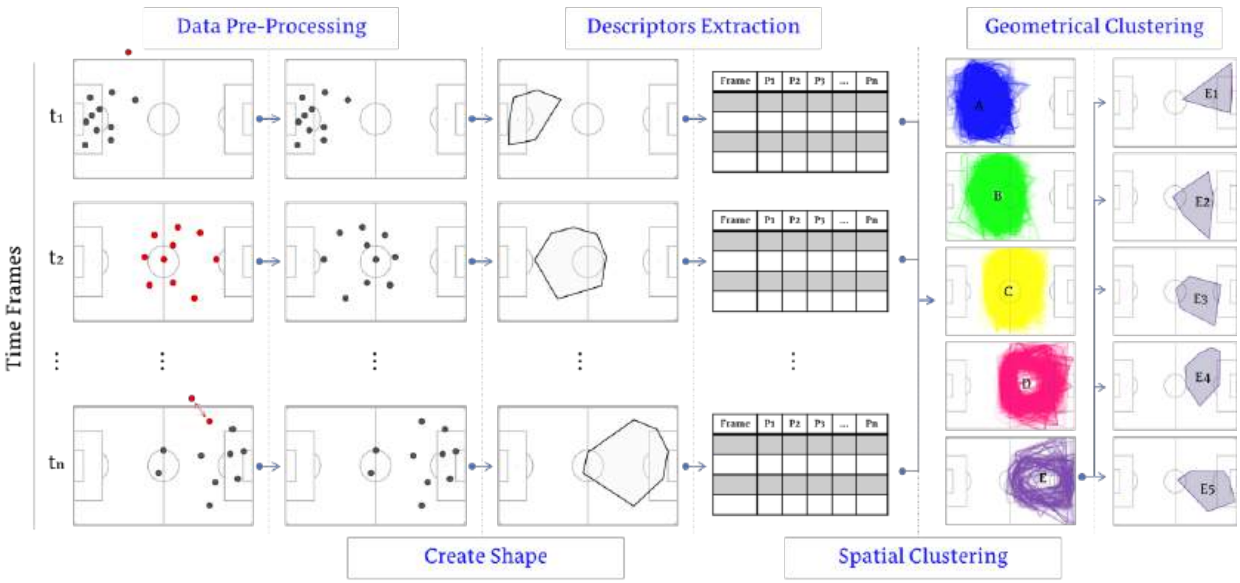

Figure 5. General process of the proposed method

Data Preprocessing

In the pre-processing step, the data located outside the boundaries of the pitch as well as the data outside the legal time of the match are removed. Also, due to the replacement of a series of players during the match, it is necessary to replace the location of them with the new players. Another process that must be done for data integrity is transferring the data of the two halves of the match to one side of the field. For this, it is enough to mirror the data in the second half of the match with respect to the middle line of the pitch. With the completion of these processes, the shape construction phase begins.

Shape Construction in Each Time Frame

Spatial data related to players includes a set of points. Therefore, to create the team shape, it is necessary to consider the location of all players in each time frame in the form of a single geometry. Various methods have been presented to define the geometric team shape. The most common of these methods are:

1. The usual formation of players in a match: The most well-known pattern that is usually used in matches to express the team shape is the starting arrangement of the players (Shaw and Glickman, 2020). From a geometric point of view, in this case, the shape is defined in the form of a multipoint in each time frame (Figure 6.a). Despite the commonality of this method, this arrangement does not visually provide information on the extent of dominance of the players on the pitch.
2. Buffer Zone: If the scope of each player's activity is defined by a circle with a certain radius and then these boundaries are considered for the whole team as an integrated geometry, a general shape for the team can be defined in each frame (Liu, 2022) (Figure 6.b). The challenge that exists related to this geometry is the existence of holes that are created by the transfer of players in the team shape, and from the computational point of view, it can lead to the complexity of calculations.
3. Voronoi diagram of team players: One of the geometries used to define the area covered by the players of a team is Voronoi diagram (Fujimura and Sugihara, 2005) (Figure 6.c). Considering that the final shape divides the entire area of the pitch into specific areas, therefore the overall shape of the team reproduces the pitch area.
4. Convex hull: This geometric shape defines the region that the entire team can cover in each frame (Liu, 2022) and Narizuka and Yamazaki (2019) introduce this region as the effective area of the game. The convex hull is defined as a function of time and location of all players (except the goalkeeper) (Figure 6.d). This shape is integrated and convex and creates a visual insight of the

team's coverage over the match. According to this description, in this paper, the convex hull is used to define the team shape.

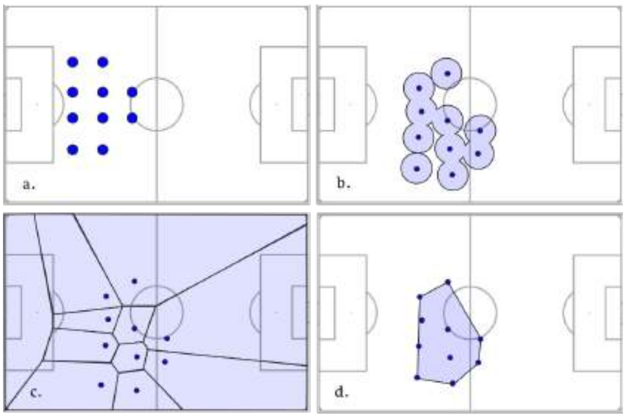

Figure 6. Different shape of the team: a. Initial formation, b. Buffer zone, c. Voronoi diagram, d. Convex hull

Extraction of Team Shape Descriptors in Each Time Frame

In this paper, to analyze the overall behavior of a team during the match, the geometric changes of the team shape and similar geometric patterns are examined. According to the previous section, the geometry used to define the shape of the team is a convex hull. To analyze the shape, it is necessary to first extract a set of parameters describing the shape. These descriptors, in this paper, are divided into four groups: spatial descriptors, zone-based descriptors, geometric descriptors, and event-based descriptors. In the following, these descriptors will be examined. Here, it should be noted that the shape may not be fully reconstructed from the descriptors, but the descriptors defined by the shapes must be different enough to distinguish the shapes from each other (Wirth, 2004).

1. Spatial Descriptors

To describe the overall location of the team on the pitch, the geometric center of the team shape is extracted. This center is calculated based on the average of the coordinates of the shape vertices according to equation 1. In this relation, N is the number of vertices and V_i is the coordinates of the vertices of the shape.

$$Convex\ hull\ Centroid = \frac{1}{N}\sum_{i=1}^{N} V_i$$

2. Zone-based Descriptors

In order to better manage the pitch and the arrangement of the players, the coaches divide the pitch into a set of zones. In this paper, in each time frame, the percentage of the team shape's area overlaps with each of these zones can be used as a parameter describing the shape. Figure 7 shows the proposed zoning by two current football coaches. (https://spielverlagerung.com/glossary/pitch-zones/six-by-three/). In this paper, the zoning proposed by Pep Guardiola (https://www.coachesvoice.com/cv/positional-play-football-tactics-explained-guardiola-cruyff-manchester-city/), which divides the field into 26 zones, is used.

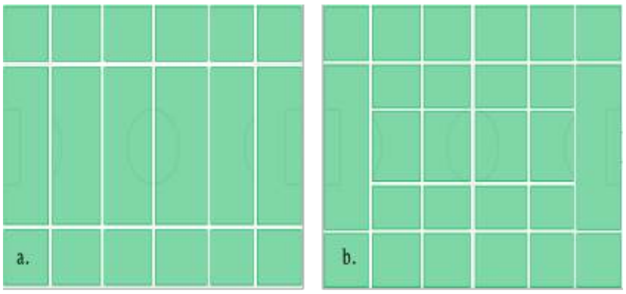

Figure 7. Zoning of the football pitch in different ways: a. Louis van Gaal's method1, b. Pep Guardiola's method

3. Geometrical Descriptors

In general, there are two main methods for describing geometry, contour-based and region-based. In contour-based methods, focusing on the vector space, it is tried to extract the characteristics of the main shape based on the boundary of the shape and turn it into a numerical representation (Fan, Zhao and Wenwen, 2021). But in region-based methods, first the shape is transferred from vector space to the image space and then based on the characteristics of the pixels, the characteristics of the shape are tried to be reflected (Bueno et al., 2021).

One of the well-known region-based shape representation method is the Moment invariant. Moments are used to define a unique shape by considering a set of features in the form of scalar quantities (Flusser, Suk and Zitova, 2009). The definition of moments is derived from probability theory, which is a numerical property used to describe the distribution of random variables (Fu et al., 2018). Until now, various methods have been presented to define moments, among which Hu Invariant moments and Zernike moments can be mentioned.

Hu defines seven numerical quantities as shape descriptors, which are calculated from central moments to the third order and are independent of the shape's translation, scale, and orientation (Keyes and Winstanley, 2001). Zernike moment is a shape descriptor based on image area and its basis is orthogonal radial polynomial (Chen and Lu, 2016). The important thing about this method is that these moments are independent of the scale, orientation and displacement of the shape (Marouf and Faez, 2013). This is, while the parameters of the area and the direction of movement in relation to a team during a match can play an important role on the efficiency and behavior of the team. Also, to calculate the invariant moment, it is necessary to transfer the geometric shape from the vector space to the image space and then perform the calculations in this space (Sabhara, Lee and Lim, 2013). Accordingly, Moment invariant is not used in this paper. Rather, instead of them, a new method based on vector space is presented to describe the shape of the team according to the dimensions and orientation of the team shape.

In this proposed method, the entire space is divided radially into 10 sectors centered on the shape centroid. In each sector, it calculates the average number of vertices and their distance to the centroid, adding 20 parameters to the shape descriptors. Considering that the convex hull created here is based on the location of 10 players (Without goalkeeper), hence the final shape will have a maximum of 10 vertices. Therefore, the whole space is divided into the same number of sub-spaces (Figure 8). This set of descriptors, without transferring data to raster space, efficiently describes the team's shape with less computation than the Hu method. It better captures the shape and behavior of the team due to its dependence on the number of vertices and shape elongation.

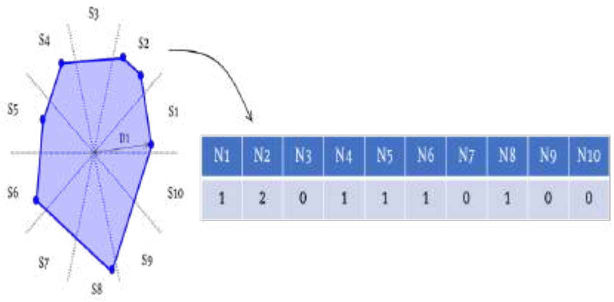

Figure 8. Spread the space into 10 sectors and check the distribution of shape vertices to describe the shape of the team

4. Event-Based Descriptors

Along with the location data, one of the most important types of data in football is the data related to the events. Based on the nature of events, different classifications are considered for them. Events can occur at a specific point (such as fault) or have a specific origin and destination (such as pass). On the other hand, the events can express the players' activities (such as pass or shot) or display a technical event (such as fault or the beginning and end of the half of the match) (Gudmundsson and Horton, 2017). Also, in another classification, they can be divided into two parts: ball-based events (such as throw-out, pass, deep pass, shots, and receiving the ball), and non-ball events (such as tackles, and dribble) (Haaren et al., 2019).

According to (Goncalves et al., 2019), one of the most important parameters influencing the team's behavior is the in-possession or out-of-possession of the ball. Therefore, in this paper, the issue of ball possession plays a role as one of the important decision-making parameters in the analysis, and finally, the results are presented based on it. In order to determine the ball possession, the events data is used here. Based on the events that require the ball (such as pass, shot, etc.), the winning of ground or aerial challenges on (such as dribbling or tackling), the times when the ball is in the possession of the team are determined. In the same way, the times when the above events are carried out by the opponent team are considered as times of out-of-possession for the team.

Data Partitioning

Considering the importance of possession of the ball on the overall behavior of the team, before clustering, it is necessary to divide the entire data set into two partitions based on event-based descriptors. Based on these descriptors, all the time frames of the match are partitioned, and the continuation of the data processing and analysis process is done separately in each part. Also, the interpretation of the results is performed based on this data partitioning.

Spatio-Geometrical Clustering

One of the methods that can help to analyze the teams is a post-review of the matches that the team participated (Carling, Williams, and Reilly, 2005). In this paper, to review the matches, it has been tried to analyze the team by finding similar patterns during the match. One of the most well-known machine learning methods that can help identify hidden patterns in data is the clustering method (Chaudhry et al., 2023). This paper performs the clustering process based on the team shape descriptors during the match. The output of this analysis is finding similar behavior patterns for the team at different times of the match. There are different methods for clustering and in this paper, the K-Means algorithm is chosen. The possibility of using for a large amount of data, ease of implementation, speed of execution, computational efficiency, and the optimal amount of memory consumption make this algorithm a suitable option for team analysis (Morissette and Chartier, 2013; Gasparini and Álvaro, 2020). The process involves determining the number of clusters, selecting initial centers, calculating data distances to cluster centers, and assigning data to the nearest cluster. The weighted average of each cluster's data becomes the new center, and this process repeats until all similar data are placed in a cluster (Morissette and Chartier, 2013). This paper describes a two-stage clustering process for analyzing team shapes in soccer matches. First, spatial clustering using the K-Means algorithm considers the geometric center and overlapping with 26 pitch zones, resulting in five clusters representing team states: defense, attack, transition from defense to attack, transition from attack to defense, and play in the middle zones. Next, geometric clustering, also using K-Means, is applied to shapes within each main cluster based on proposed geometric descriptors. This method provides a detailed analysis of team formations and transitions during a match.

RESULTS

Evaluation of results

To evaluate clustering results, compare shape differences in each cluster over time based on location, coverage area, and orientation relative to the center of the cluster. The shape closest to the cluster center is considered the representative shape of the cluster center. Here, equation 2 is used for this evaluation.

$$Diff_{shape\ i} = w_a * Diff_{Location} + w_p * Diff_{AreaCoverage} + w_v * Diff_{Orientation}$$

In which, the $Diff_{Location}$, $Diff_{AreaCoverage}$, and $Diff_{Orientation}$ express the amount of difference based on the location, as the difference based on the distance and the number of vertices located in the 10 sectors (Figure 8), respectively, which is defined according to equations 3-5. The w parameters are the weight corresponding to each of them, which are considered equal here.

$$Diff_{Location} = \frac{|Location_{shape} - Location_{ClusterCenter}|}{\max\,(Location_{shape}, Location_{ClusterCenter})}$$

where, $Location_{shape}$ is the location of the team and $Location_{ClusterCenter}$ is the location of the cluster center in the pitch.

$$Diff_{AreaCoverage} = \frac{\left|\sum_i^{10} D_{i.Shape} - \sum_i^{10} D_{i.ClusterCenter}\right|}{\max\left(\sum_i^{10} D_{i.Shape}, \sum_i^{10} D_{i.ClusterCenter}\right)}$$

where, $D_{i.Shape}$ is the distance of the vertices of the team shape from the geometric center in each sector and $D_{i.ClusterCenter}$ is the distance of the vertices of the cluster center from its geometric center.

$$Diff_{Orientation} = \frac{|Vertices_{shape} - Vertices_{ClusterCenter}|}{\max\,(Vertices_{shape}, Vertices_{ClusterCenter})}$$

Where, $Vertices_{shape}$ is the number of team shape vertices and $Vertices_{ClusterCenter}$ is the number of cluster center vertices.

IMPLEMENTATION

Data and software used

This study utilizes spatio-temporal data from two matches provided by Metrica Sports Sample Data (https://github.com/metrica-sports/sample-data). The dataset consists of CSV files containing anonymized player tracking and event data recorded at 0.4-second intervals. The first match includes 145,006 rows of tracking data and 1,745 event data points, while the second match has 141,156 tracking records and 1,935 event data points. Both teams emerged victorious in their respective matches, with scores of 3-0 and 3-2.

Data processing was conducted using Oracle 19c, where pre-processing operations such as data cleaning, convex hull construction, and shape descriptor extraction were performed. Clustering and visualization were carried out using Python 3.9 with the Scikit-Learn, Pandas, NumPy, and Matplotlib libraries. The Seaborn library was used for enhanced data visualization. The combination of these tools ensures the reproducibility of the research.

Data Processing

Following data pre-processing, a two-stage clustering approach was applied. First, K-Means clustering was implemented with five clusters and 20 repetitions, focusing on spatial descriptors. The results for both in-possession and out-of-possession states are presented in Figure 9, showcasing the spatial clustering outcomes. The spatial clustering results for the teams, both in possession and out-of-possession, are illustrated in Figure 9.

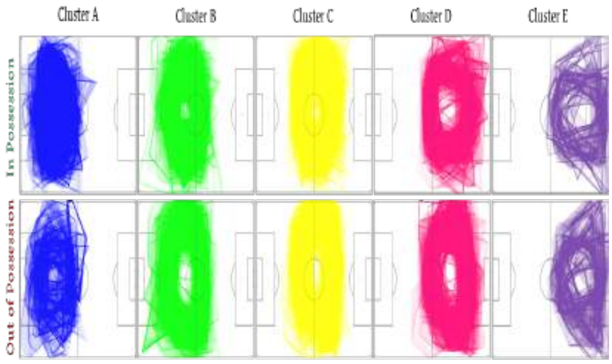

Figure 9. The position of all T1 team shapes in five clusters in spatial clustering in two situations of in-possession and out-of-possession

In the second stage, each primary cluster was subjected to geometric clustering using K-Means to further categorize sub-clusters based on shape descriptors. Figures 10 to 13 illustrate the geometric clustering results for Teams T1 and T2 across both matches. The time contribution of each sub-cluster is also displayed, along with directional movement trends and goal-scoring moments.

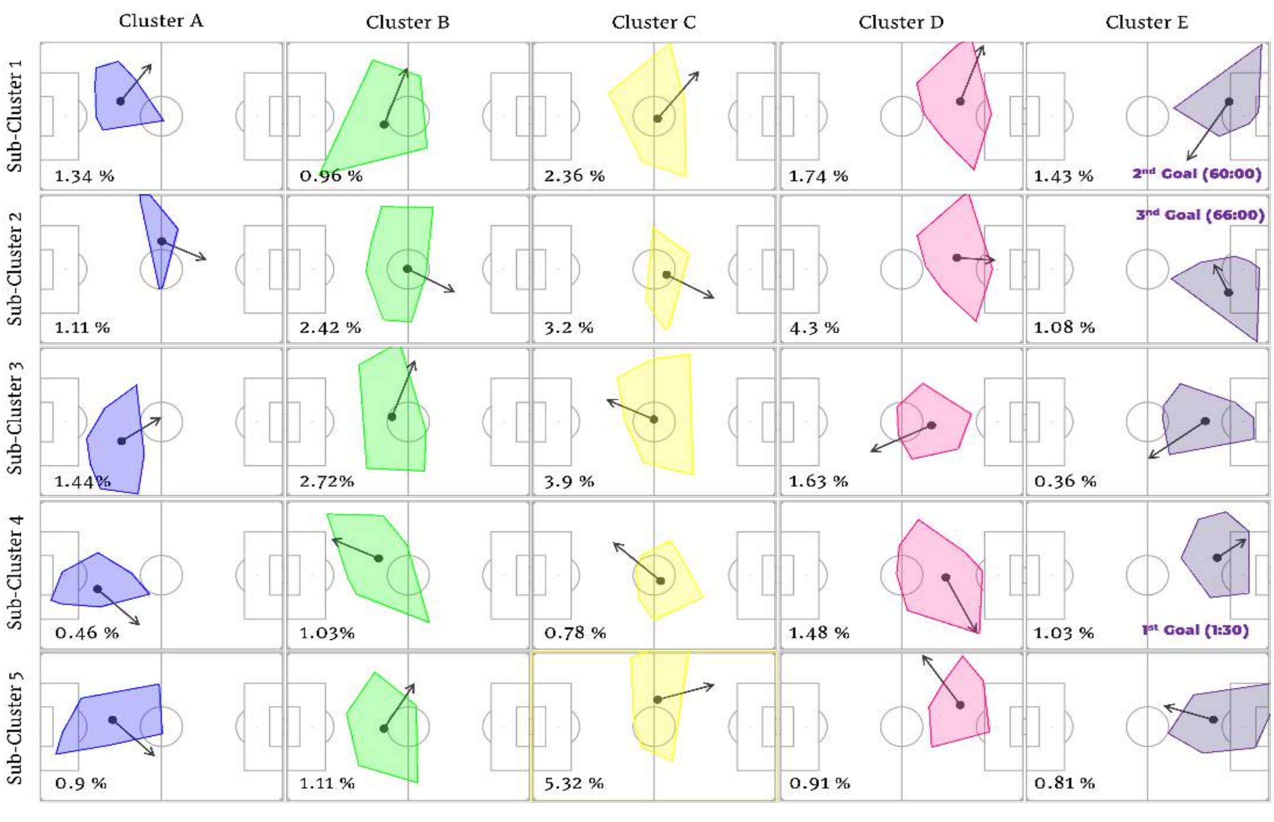

Figure 10. The results of the geometric clustering of the shape of team T1 in match 1 in the situation of in-possession

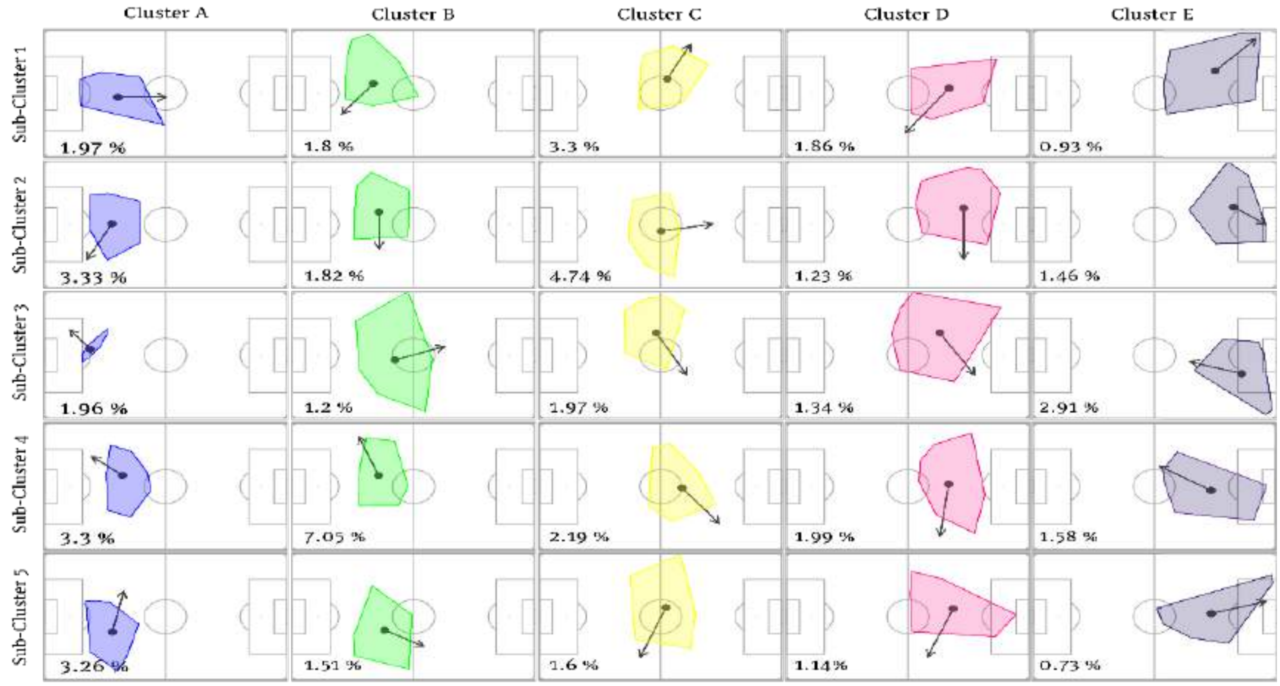

Figure 11. The results of the geometric clustering of the shape of team T1 in match 1 in the situation of out-of-possession

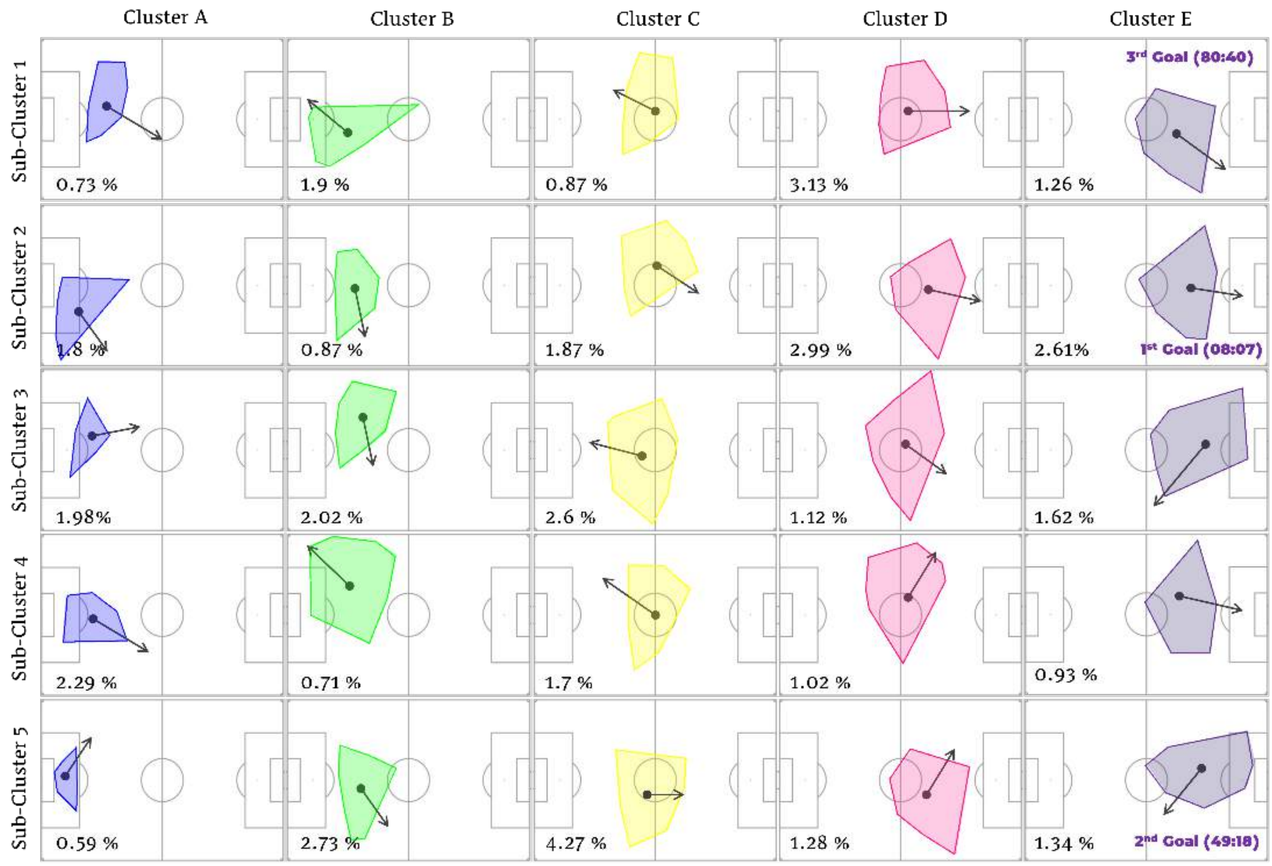

Figure 12. The results of the geometric clustering of the shape of team T2 in match 2 in the situation of in-possession

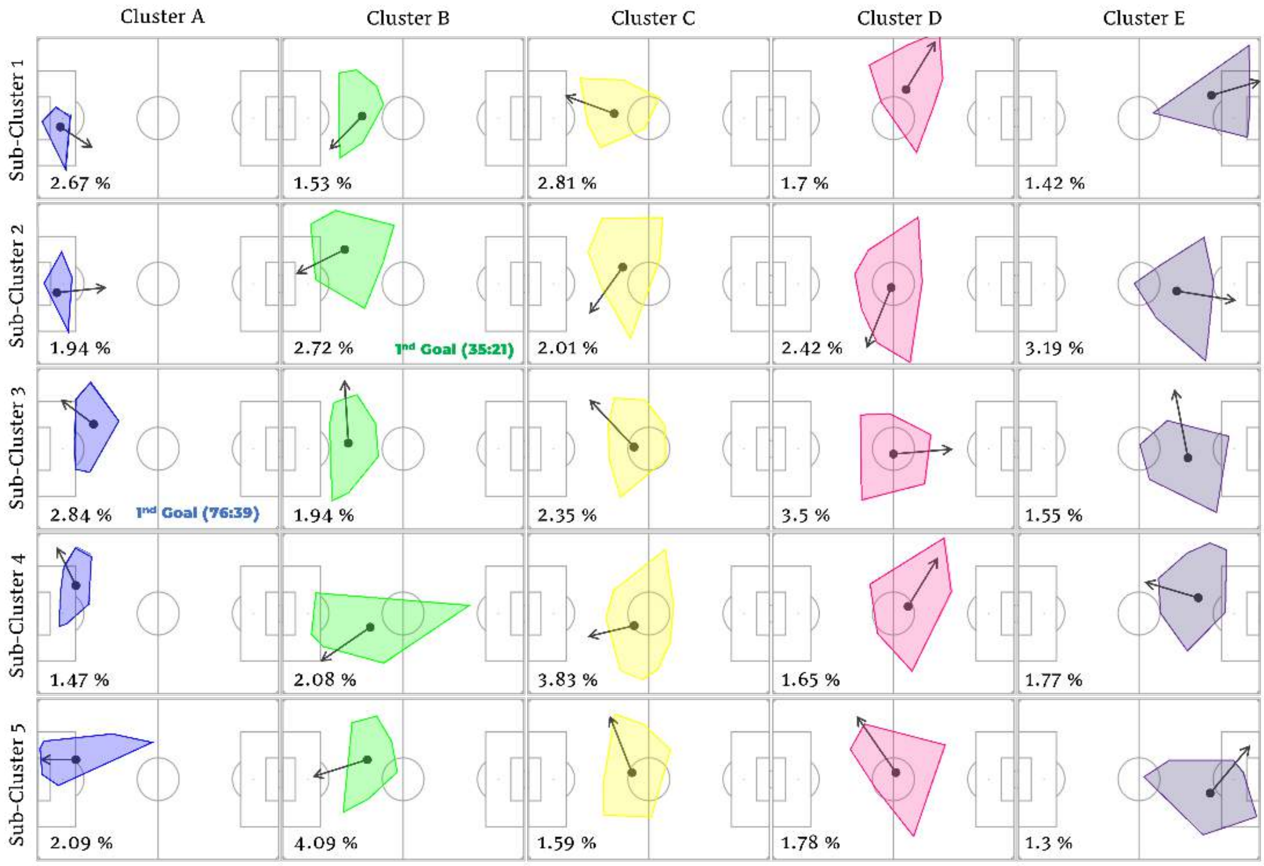

Figure 13. The results of the geometric clustering of the shape of team T2 in match 2 in the situation of out-of-possession

Analysis of the Overall Behavior of Teams

Clustering enables the identification of repeated tactical patterns in team shape over the course of a match. To assess the impact of shape variations on performance, several key indicators were analyzed using event data. One of the primary aspects examined was cluster transition before goals, where the team's positional clusters were evaluated in the two minutes leading up to goals to determine trends in movement and positioning. Figure 14 illustrates that T1 predominantly occupied home-field clusters (A & B) before scoring, indicating a counterattacking strategy. Conversely, T2 was positioned in the opponent's half before scoring, reflecting a more aggressive attacking playstyle.

Another key analysis involved the relationship between passing and team shape, where the total number of passes was compared to the covered area of the team shape. As shown in Figure 15, a positive correlation was identified between increased shape area and passing efficiency, suggesting that broader formations enhance ball circulation and passing accuracy. This finding underscores the importance of maintaining a well-spread team shape to optimize offensive play.

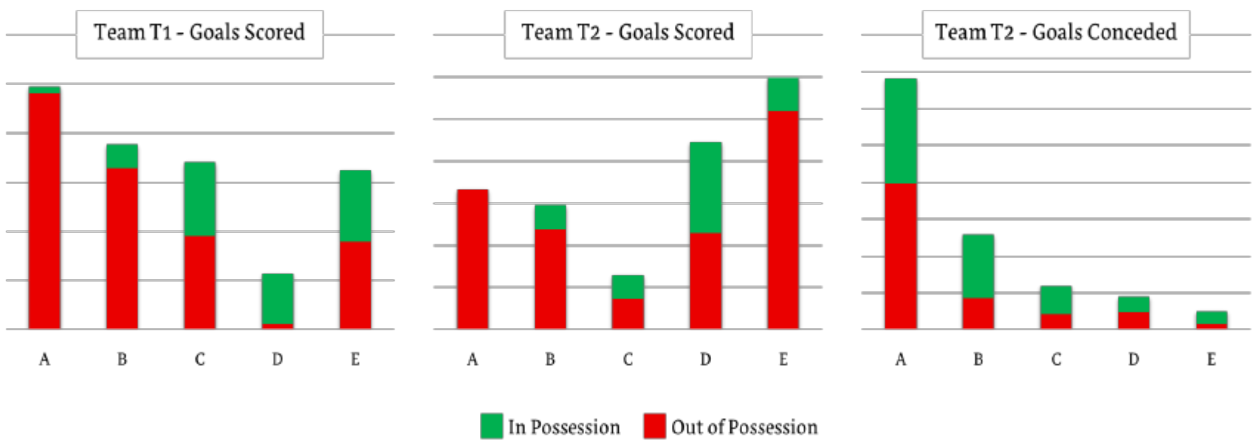

Figure 14. Average time share of main clusters in 2 minutes leading to match goals

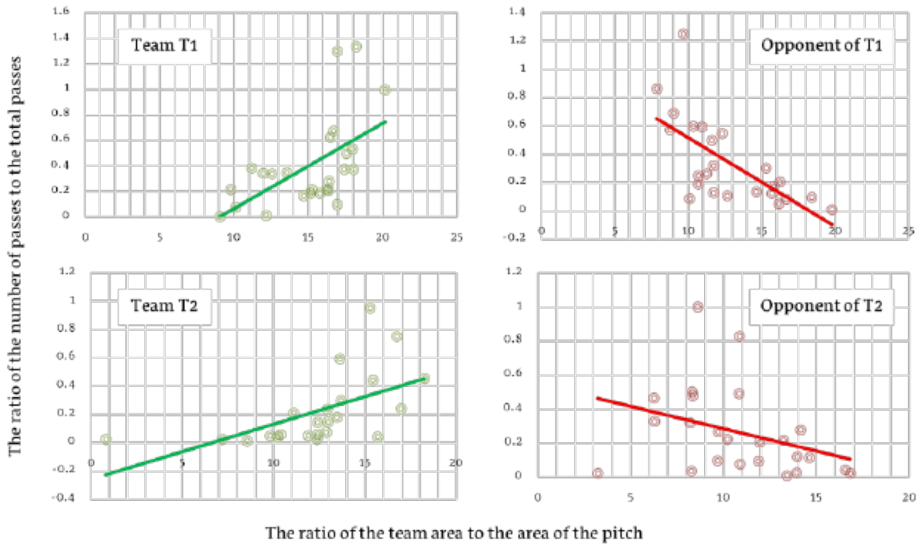

Figure 15. Ratio of passes to teams' shape area

Additionally, the distribution of players in challenges was analyzed by comparing the geometric center of the team shape to individual player positions. Figure 16 demonstrates that a more evenly distributed player shape results in stronger challenge success, as it facilitates better defensive coverage and midfield control. Furthermore, Figure 17 supports this observation by highlighting the correlation between positional spread and duel success rates, indicating that teams with well-balanced formations are more effective in winning possession battles.

Lastly, an analysis of time spent in each cluster provided further insights into team strategies. Figure 18 reveals that T2 exhibited more balanced positioning across all clusters, whereas T1 spent more time in clusters A & E, suggesting a structured defensive and offensive approach. These patterns indicate distinct tactical preferences between the two teams, with T1 favoring a more controlled build-up while T2 adopted a flexible positioning strategy.

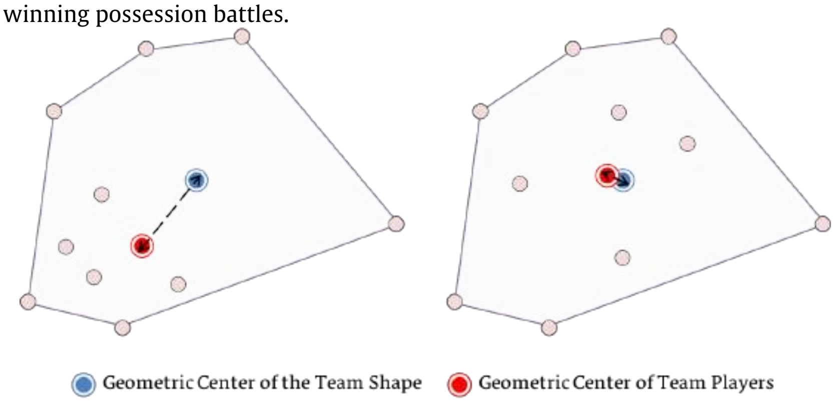

Figure 16. Distribution of players inside the team shape

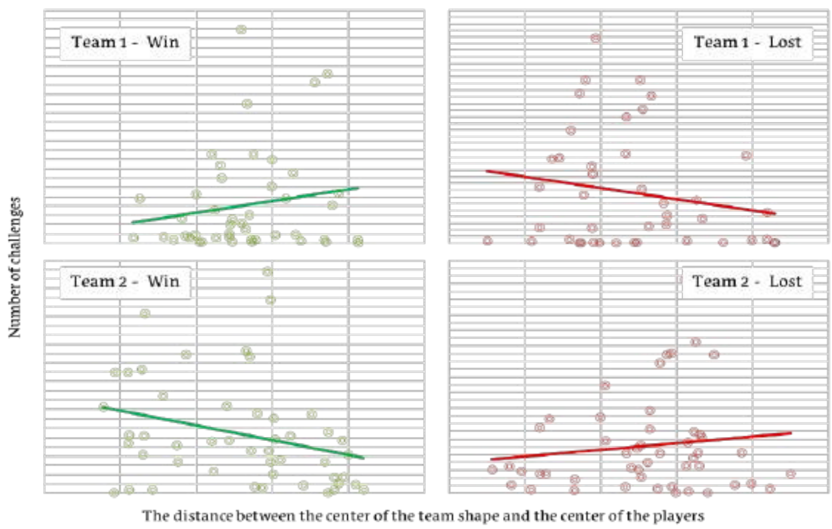

Figure 17. The ratio of the number of team challenges to the distribution of players

The fourth indicator measures the average time a team stays in the same cluster. Figure 18 shows that T2 had a more uniform pattern across the field, regardless of in-Possession or out-of-Possession states, while T1 spent more time in clusters A and E. Maybe just based on this figure, it can be said that in the first match, the T1 team performed poorly in the defense and attack phases, but considering the team's win (without the opposing team being able to score), it can be said that the team in out-of-possession state adopts predetermined patterns to manage the game, especially in its defensive area and the opponent team's defensive area. This indicator supports the first indicator's explanation of the teams' overall strategies.

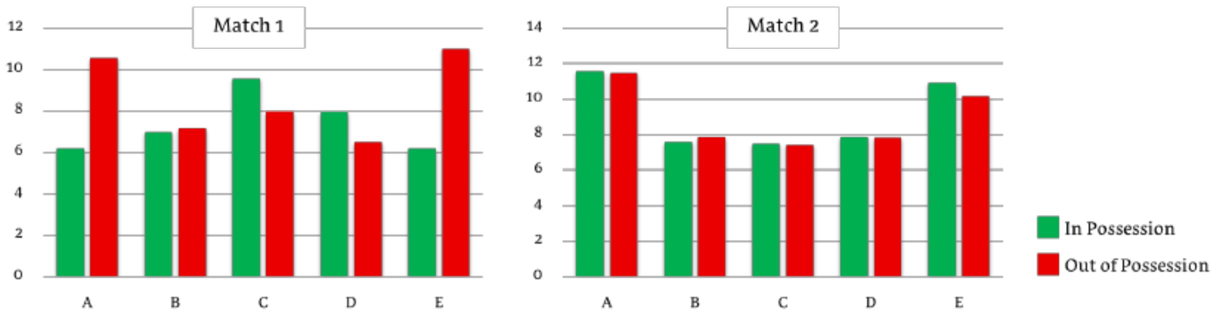

Figure 18. The average time the team has been in a cluster continuously

Moreover, understanding shape evolution during challenges allows teams to optimize their defensive and offensive structures. The ability to assess positional spread and duel success rates can help teams improve their organization in pressing and transition scenarios. This study demonstrates how spatio-geometrical clustering provides a data-driven evaluation of team performance, enabling more precise tactical decisions. Future research can explore temporal shape dynamics, dual-team shape interactions, and overlapping formations to further refine tactical insights and strategic planning.

LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Despite its contributions, this study has certain limitations. One major limitation is the scope of the dataset, as it is based on a limited number of matches. Expanding the dataset to include matches from different leagues, playing styles, and tactical systems would enhance the generalizability of the findings. Additionally, this study primarily focuses on team shape without incorporating temporal dependencies, which could be explored further using time-series analysis or recurrent neural networks. Future research could also investigate dual-team interactions by analyzing how both teams' shapes evolve simultaneously in competitive match conditions. Lastly, incorporating player-specific movement trends within the team shape framework would provide deeper insights into individual contributions to team dynamics.

CONCLUSION

This study proposed a novel two-stage spatio-geometrical clustering approach for post-match analysis in football, focusing on evaluating team shape dynamics. By defining the team shape as a convex hull in each time frame and extracting spatial, zone-based, geometric, and event-based descriptors, the study successfully identified recurring behavioral patterns in team formations. The K-Means algorithm was applied in two stages–first for spatial clustering and then for geometric clustering–to capture distinct tactical structures in both in-possession and out-of-possession states. The results demonstrated that team shape variations significantly correlate with key performance indicators, such as passing efficiency, success in duels, and positional adjustments leading to goal-scoring opportunities.

This research makes several important contributions to the field of football analytics. First, it eliminates the need for image-processing techniques by leveraging a fully vector-based clustering approach, making it computationally more efficient. Second, it introduces novel geometrical descriptors that provide a more detailed representation of team shape beyond conventional metrics such as area, perimeter, and centroid displacement. Third, the two-stage clustering methodology enhances the interpretability of team shape shifts by linking spatial positioning with tactical execution. These contributions not only refine the existing methodologies in football analytics but also provide a scalable framework for analyzing team performance across different matches and competitions.

The findings of this study offer valuable insights for coaches and football analysts. By understanding how team shape influences passing sequences and defensive stability, coaching staff can make informed tactical adjustments to improve ball circulation, pressing intensity, and transition strategies. Additionally, the ability to identify repeated shape patterns leading to successful goal-scoring opportunities enables teams to refine their attacking strategies. The framework also provides a quantitative assessment of team performance, which can be used for opponent scouting, match preparation, and in-game tactical adjustments.

This research highlights the importance of spatio-geometrical clustering as a robust method for post-match tactical analysis in football. By leveraging vector-based clustering and novel geometrical descriptors, the study offers a computationally efficient and tactically insightful approach to analyzing team shape. The methodology not only enhances the understanding of spatial-temporal formations in football but also provides a valuable tool for coaches, analysts, and sports scientists in optimizing team strategies. Future work will focus on refining temporal shape dynamics, integrating multi-team interactions, and expanding datasets to further improve the applicability of the proposed framework in professional football settings.

REFERENCES

Araújo, D., Passos, P., Esteves, P., Duarte, R., Lopes, J., Hristovski, R., & Davids, K. (2015). The Micro-Macro Link in Understanding Sport Tactical Behaviours: Integrating Information and Action at Different Levels of System Analysis in Sport. *Movement & Sport Sciences – Science & Motricité*, 53–63. https://doi.org/10.1051/sm/2015028

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., & Matthews, I. (2014). Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. *IEEE International Conference on Data Mining*, 725-730. https://doi.org/10.1109/ICDM.2014.133

Bueno, M. J. O., Silva, M., Cunha, S. A., Torres, R. S., & Moura, F. A. (2021). Multiscale Fractal Dimension Applied to Tactical Analysis in Football: A Novel Approach to Evaluate the Shapes of Team Organization on The Pitch. *PLoS ONE, 16*(9), e0256771. https://doi.org/10.1371/journal.pone.0256771

Caetano, F.G., Junior, S.B., Torres, R.S., Cunha, S.A., Caron, P.R., Ruffino, P.R.C., Martins, L.E.B., and Moura, F.A. (2021). Football player dominant region determined by a novel model based on instantaneous kinematics variables, *Scientific Reports* 11(1), DOI:10.1038/s41598-021-97537-4

Carling, C., Williams, A. M., & Reilly, T. (2005). *Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance*. Routledge.

Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D.L.R., Thompson, E.B., Ashraf, I. (2023). A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. *Symmetry*, 15, 1679. https://doi.org/10.3390/sym15091679

Chen, W. and Lu, J (2016). Human Shape Recognition Algorithm Design Based on Hu Moments and Zernike Moments, Advances in Computer Science Research, volume 71. 4th International Conference on Machinery, Materials and Information Technology Applications

Csanalosi, G., Dobreff, G., Pašić, A., Molnar, M., and Toka, L. (2020). Low-cost optical tracking of soccer players, 7th Workshop on Machine Learning and Data Mining for Sports Analytics, ECML/PKDD 2020 Workshop

Fan, H., Zhao, Z., and Wenwen, L. (2021). Towards Measuring Shape Similarity of Polygons Based on Multiscale Features and Grid Context Descriptors. *ISPRS Int.* J. Geo-Inf.2021, 10, 279. https://doi.org/10.3390/ijgi10050279

Flusser, J., Suk, T., Zitova, B. (2009). Moments and moment invariants in pattern recognition, West Sussex, Wiley, DOI:10.1002/9780470684757

Fu, Z., Fan, L., Yu, Z., and Zhou, K. (2018). A Moment-Based Shape Similarity Measurement for Areal Entities in Geographical Vector Data, *ISPRS International Journal of Geo-Information*, doi:10.3390/ijgi7060208

Fujimura, A., and Sugihara, K., (2005). Geometric Analysis and Quantitative Evaluation of Sport Teamwork. *Systems and Computers in Japan*, 36(6), DOI:10.1002/scj.20254

Gasparini, R., and Álvaro, A. (2020). Positional Analysis of Brazilian Soccer Players Using GPS Data. *Revista Brasileira de Computação Aplicada*, 12 (3):16–32, DOI: 10.5335/rbca.v12i3.10234

Goes, F.R., Meerhoff, L.A., Bueno, M.J.O., Rodrigues, D.M., Moura, F.A., Brink, M.S., Elferink-Gemser, M.T., Knobbe, A.J., Cunha, S.A., Torres, R.S., and Lemmink, K.A.P.M. (2020). Unlocking the Potential of Big Data to Support Tactical Performance Analysis in Professional Soccer: A Systematic Review. *European Journal of Sport Science*. Eur J Sport Sci.21(4):481-496. DOI: 10.1080/17461391.2020.1747552.

Goncalves, B., Coutinho, D., Exel, J., Travassos, B., Lago, C., & Sampaio, J. (2019). Extracting Spatial-Temporal Features That Describe a Team Match Demands When Considering the Effects of The Quality of Opposition in Elite Football. *PLoS ONE*, *14*(8), e0221368. https://doi.org/10.1371/journal.pone.0221368

Gudmundsson, J., & Horton, M. (2017). Spatio-Temporal Analysis of Team Sports. *ACM Computing Surveys, 50*(2), 1–34. https://doi.org/10.1145/3054132

Haaren, J.V., Robberechts, P., Decroos, T., Bransen, L., and Davis, J. (2019). Analyzing Performance and Playing Style Using Ball Event Data. https://www.janvanhaaren.be/assets/papers/bih-2019-event-data.pdf

Keyes, L. and Winstanley, A.C (2001). Using moment invariants for classifying shapes on large scale maps, Computers, Environment and Urban Systems, 25:119-130, https://doi.org/10.1016/S0198-9715(00)00041-7

Kotzbek, G., and Kainz, W. (2015). GIS-Based Football Game Analysis – A Brief Introduction to the Applied Data Base and a Guideline on How to Utilize It, Computer Science, Geography, https://api.semanticscholar.org/CorpusID:41184418

Kotzbek, G., Kainz, W. (2018). Dynamic Zoning in the Course of GIS-Based Football Game Analysis. In: Ivan, I., Horák, J., Inspektor, T. (eds) Dynamics in GIscience. GIS OSTRAVA 2017. Lecture Notes in Geoinformation and Cartography. *Springer, Cham*. https://doi.org/10.1007/978-3-319-61297-3_17

Liu, N. (2022). Geovisualisation of Football Players Movement. Diploma Thesis. Palacký University Olomouc, Faculty of Science, Department of Geoinformatics. https://theses.cz/id/mbeof5/48261609

Liu, N., Zhuang, L., Netek, R., & Strobl, J. (2021). The overview, outcome, and outlook of GIS-based football visualization. *13th GeoMundus 2021*.

Marouf, H., Faez, K. (2013). Zernike Moment-Based Feature Extraction for Facial Recognition of Identical Twins, *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)*, 3(6), DOI: 10.5121/ijcseit.2013.3601

Morissette, L., and Chartier, S. (2013). The K-Means Clustering Technique: General Considerations and Implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*. 9(1): 15-24. DOI: 10.20982/tqmp.09.1.p015

Narizuka, T. and Yamazaki, Y. (2019). Clustering Algorithm for Formations in Football Games. *Scientific Reports*, 9:13172 | https://doi.org/10.1038/s41598-019-48623-1

Querido, S.M., Radaelli, R., Brito, J., Vaz, J.R., and Freitas, S.R. (2022). Analysis of Recovery Methods' Efficacy Applied up to 72 Hours Postmatch in Professional Football: A Systematic Review with Graded Recommendations. *International Journal of Sports Physiology and Performance*, 17(9): 1326-1342, DOI: https://doi.org/10.1123/ijspp.2022-0038

Sabhara, R.K., Lee, C., and Lim, K. (2013). Comparative Study of Hu Moments and Zernike Moments in Object Recognition, *Smart Computing Review*, 3(3), DOI: 10.6029/smartcr.2013.03.003

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

Seidl, T., Czyz, T., Spandler, D., Franke, N., and Lochmann, M. (2016). Validation of Football's Velocity Provided by a Radio-based Tracking System, Procedia Engineering, 147: 584-589, DOI: https://doi.org/10.1016/j.proeng.2016.06.244

Shaw, L., and Glickman. M. (2020). Dynamic Analysis of Team Strategy in Professional Football. BARÇA Sports Analytics Summit. https://static.capabiliaserver.com/frontend/clients/barca/wp_prod/wp-content/uploads/2020/01/56ce723e-barca-conference-paper-laurie-shaw.pdf

Teixeira, J.E., Forte, P., Ferraz, R., Leal, M., Ribeiro, J., Silva, A.J., Barbosa, T.M., Monteiro, A.M. (2021). Monitoring Accumulated Training and Match Load in Football: A Systematic Review. Int. J. Environ. *Res. Public Health*, 18, 3906. DOI: https://doi.org/10.3390/ijerph18083906

Whitmore, J., and Seidl, T. (2021). Shape Analysis: Automatically Detecting Formations. https://theanalyst.com/na/2021/03/shape-analysis-automatically-detecting-formations/

Wirth, M.A. (2004). Shape Analysis & Measurement, University of Guelph, Computing and Information Science, Image Processing Group, http://www.cyto.purdue.edu/cdroms/micro2/content/education/wirth10.pdf

Zhang, G. (2022). The Analysis of Team Tactical Behaviour in Football Using GNSS Positional Data. Master Thesis. Liverpool John Moores University. DOI:10.24377/LJMU.t.00019091. https://researchonline.ljmu.ac.uk/id/eprint/19091/1/2022guangzezhangmphil.pdf

This page has been intentionally left blank