

Tactical patterns in Europe's top five soccer leagues: a cluster analysis of player performance from 2010–2016

Jiarui Li

Discovery High School, Newton, US

gl801469@gmail.com

Abstract. This study investigates the persistence of distinct national tactical identities across Europe's top five football leagues (Premier League, La Liga, Serie A, Bundesliga, and Ligue 1) amidst the backdrop of increasing globalization. Utilizing a longitudinal dataset of 588 team-season observations derived from FIFA attributes (2010–2016), the analysis employs cluster analysis, principal component analysis, and classification modeling to examine stylistic differences and temporal evolution. The findings confirm that while tactical diffusion is evident, systematic structural differences persist: the Bundesliga and Serie A are characterized by high tempo and physical aggression, La Liga maintains a slower, possession-oriented approach, and Ligue 1 exhibits distinct preferences for structured buildup play. The classification model successfully predicts league affiliation with a macro-AUC of 0.84, demonstrating that tactical attributes contain sufficient information to distinguish leagues above random chance. Although temporal analysis indicates a slight trend toward convergence, the results suggest that complete homogenization has not occurred, as long-standing competitive environments and cultural traditions continue to sustain recognizable league-level stylistic identities.

Keywords: soccer analytics, tactical analysis, league identity, convergence, FIFA data

1. Introduction

European soccer is often distinguished not only by the level of competition across its major leagues but also by the perception that each league embodies a distinct tactical identity. Historically, observers have associated English soccer with a high tempo and physical intensity, Spanish soccer with possession-oriented circulation and technical control, and Italian soccer with defensive organization and strategic discipline. These stylistic patterns are commonly attributed to regional coaching traditions, player development systems, and long-standing competitive cultures. Over time, such identities have extended beyond national borders, shaping tactical thinking across global soccer. In this sense, stylistic influence sometimes travels further than competitive success itself.

In recent years, however, the clarity of these distinctions has increasingly been questioned. The modern soccer ecosystem is characterized by rapid international mobility of players, coaches, and tactical ideas. Managers frequently relocate between leagues, multinational squads have become the norm, and analytical approaches to coaching circulate widely. As a result, tactical concepts that were once associated with specific leagues now diffuse quickly across national contexts. English clubs have increasingly adopted structured

possession systems under foreign managers, while teams in Spain and Italy have integrated more direct transitions and higher pressing intensity. These developments raise an important empirical question: whether historically recognized national playing styles continue to exist in modern European soccer.

Despite the prominence of this debate in public discourse, systematic empirical evidence remains limited. Discussions of stylistic differences are often based on anecdotal observation or short-term snapshots rather than consistent longitudinal measurement. Many studies focus on individual teams or specific competitions, making cross-league comparisons difficult. Moreover, few analyses track tactical characteristics across multiple seasons using a unified analytical framework. Consequently, it remains unclear whether national tactical identities persist over time, gradually converge due to global tactical diffusion, or fluctuate as temporary variations driven by short-term managerial cycles.

This study addresses these gaps by examining tactical characteristics of teams competing in Europe's five major leagues between 2010 and 2016. Using open source FIFA team attribute data that capture core aspects of playing style, including attacking speed, passing preferences, chance creation patterns, and defensive pressure, the analysis constructs a consistent cross-league dataset for longitudinal comparison. The empirical investigation is guided by three central research questions. The first concerns whether systematic differences exist across leagues in the average values of tactical attributes during the period studied. The second examines whether the distance between league-level tactical profiles decreases over time, which would suggest convergence in playing styles. The third evaluates whether tactical attributes contain sufficient information to predict a team's league affiliation at a rate meaningfully above random chance, indicating persistent structural differences between leagues.

By addressing these questions, this paper contributes to the empirical study of soccer tactics in several ways. First, it develops a consistent multi-season dataset that allows direct comparison of tactical attributes across the major European leagues. Second, it provides a low-dimensional representation of team playing styles that helps visualize similarities and differences between leagues. Third, it offers empirical evidence on whether national tactical identities persist or converge in an era characterized by intense international exchange of players and coaching philosophies. Through quantitative analysis of observable team characteristics, the study aims to move beyond anecdotal narratives and provide systematic evidence on the evolution of playing styles in modern European soccer.

2. Literature review

Over the past two decades, research on soccer tactics and performance has expanded rapidly, driven largely by the growing availability of detailed datasets and advances in analytical methods. Within this literature, Europe's top five leagues, namely the English Premier League, Spain's La Liga, Italy's Serie A, Germany's Bundesliga, and France's Ligue 1, have attracted particular scholarly attention. Because these competitions combine long-established tactical traditions with high levels of competition and extensive match documentation [1, 2], they provide an especially valuable setting for empirical investigation of playing styles. However, much of the existing literature relies on match event data or high-resolution tracking systems that are often limited to a single league, a small number of seasons, or inconsistent tactical indicators across datasets. As a result, systematic cross-league comparisons over extended periods remain relatively rare. The present study addresses this limitation by prioritizing macro-level consistency over micro-level precision. Instead of detailed event or tracking data, it employs standardized FIFA team attributes that allow comparable measurements of tactical characteristics across multiple leagues and seasons. This approach enables a broader examination of how strategic patterns differ and evolve across Europe's major soccer competitions.

One idea often explored in soccer analytics involves how tactics differ between top-tier leagues. Earlier research has repeatedly shown unique playing patterns among Europe's leading competitions. The EPL tends to favor quick attacks and rapid shifts in tempo; meanwhile, Spanish clubs lean toward ball control and intricate passing sequences. Italian sides, on the contrary, show stronger emphasis on defensive shape and disciplined formations [1]. Beyond general observations, Kalimanis [2] used statistical methods along with algorithm-based classification on game statistics, revealing that decision-tree models worked better for identifying key factors in Italy's Serie A. In contrast, support vector machines produced sharper outcomes when analyzing Spain's La Liga. Such findings suggest each league follows its own logic of gameplay influence instead of sharing one standard blueprint.

From moment to moment, tactics reveal more than overall league trends ever can. Using a model based on sequential decisions, Van Roy and colleagues in 2019 broke down how teams move the ball forward, revealing that success often ties back to organized passing routes when facing compact defenses. When attackers shape their moves around league habits, defenders must adapt - this reactive element stood out clearly in their analysis. On another path, Fernández-Navarro's group looked at Premier League games, spotting shifts in possession value depending on whether a team was home or away, leading or trailing, up against strong or weak rivals. Direct plays, quick counters, and aggressive pressing - all behaved differently under those shifting conditions. Instead of treating strategy as fixed, both lines of inquiry point toward fluid outcomes shaped by matchup dynamics and real-time choices.

From a single-player perspective, research by Di Salvo and colleagues in 2007 first revealed how physical requirements differ between positions through time-motion methods. Midfielders were found to run farther overall when compared to attackers or fullbacks, while sustaining more effort across certain high-intensity intervals. Even though the data predates modern tracking technology, the outcomes still hold value today. Tactical approach turns out to be shaped not just by team-wide patterns, but also influenced heavily by individual responsibilities and fitness levels interwoven into collective structures.

Later studies have placed greater stress on situational factors. Using data from real matches, Fernández-Navarro and colleagues developed the Possibility Effectiveness Index (PEI) - a tool merging anticipated goal outcomes with patterns of ball progression to judge possession quality across varying game states. Instead of just repeating the idea that surroundings influence play, their work uncovered curved relationships; the effectiveness of fast breaks rises sharply, for instance, only when squads are behind by at least two goals. Outcomes like these show tactics shift in meaning depending on moment-to-moment circumstances, pushing analysis toward models where setting shapes interpretation. What seems strong in one scenario may fail in another.

One reason behind different league tactics might lie in culture and institutions. According to Oberstone [1], along with Kalimanis [2], the way teams play in the EPL - fast, forward-moving attacks - is shaped by what fans have long expected, how media talk about the game, and past coaching ideas. In contrast, La Liga's focus on controlled passing stems from a deep-rooted value placed on skillful, ball-dominant soccer. These views suggest team strategies are not just responses to winning games; they grow out of wider social patterns too. While matches change week to week, some playing styles endure because of background forces older than any single season.

High-resolution spatiotemporal data now shape how tactics are studied, thanks to new tech tools. According to Rein and Memmelt [3], soccer can benefit greatly from big data methods because today's tracking devices record fine-grained details - positional, technical, even physical - that were once out of reach. Player locations flow nonstop through tracking sources; meanwhile, event logs mark touches, passes, and shots. Aggregated scores based on ratings give another angle on performance instead of raw metrics. In

contrast with individual movement systems, FIFA's team attribute numbers summarize seasonal patterns across large groups. Though these summaries miss small-scale detail, long-term studies gain ground - and comparisons between leagues become possible where before they stalled.

Starting with Goes et al. [4], their systematic look at tracking studies highlights how big data works best when sports science joins forces with computer science. Because issues like building features, combining time-based measures, and making models understandable arise, expertise from both fields becomes essential. When shifting focus to teams, scholars might track aspects such as pace of buildup, attacking directness, or pressure in defense. These metrics, once gathered over multiple seasons, allow patterns to emerge. Insights come not just from outputs but also from examining model weights and individual variable impacts. Even without original tracking inputs, meaningful tactics research remains possible through carefully built summaries.

One way to look at modern sports analysis is through real-world testing. Delibaş and colleagues in 2020 built a hands-on system merging event logs with player movement records - this tool helps coaching staff examine earlier games, test different tactical setups, while also forecasting possible outcomes [5]. In another case, research by Memmert, Lemmink, and Sampaio three years before highlighted how location tracking exposes patterns between teammates, spacing across formations, and behavior within specific lines just before key moments unfold - useful for understanding broad team plans along with small-scale choices on the field.

One issue keeps appearing in current work: how to gauge tactical success. Earlier findings show such outcomes depend heavily on setting and process. Instead of treating results as fixed, researchers like Fernández-Navarro and team looked at expected goals alongside ball control to judge attack output [6]. In a different approach, Van Roy's group applied chance-based tools to see how offensive sequences respond when defenses break them down [7]. Because of these efforts, it now seems clear that effectiveness makes sense only when seen through specific conditions and driving forces behind play.

What a player does on the field helps clarify how tactics unfold in practice. Research by Di Salvo and colleagues from 2007 revealed differences in physical effort depending on playing role, pointing to how league-wide styles might shape unique fitness needs across positions [8]. In competitions favoring fast transitions down the wings, outside players often cover ground at greater speed. Meanwhile, setups built around ball retention tend to spread movement demands more uniformly among teammates. Because of this, matching coaching plans with conditioning routines depends heavily on these kinds of observations.

Outside top-tier soccer, thinking strategically matters just as much in learning environments. A study by Harvey and colleagues in 2010 looked at how the Teaching Games for Understanding model worked in high school settings [9], showing that methods focused on decision-making improved student play. Because of this, ideas used by professionals might guide youth training - if reshaped well for younger players.

One study suggests differences among Europe's elite soccer leagues are blurring. Through machine learning methods, Tiedemann and Hennig [10] observed similar patterns emerging in how teams pass, press, and position. Even so, certain traits remain unique: England's top division still favors forward-driven attacks, along with more mistakes when handling the ball. In Spain, play tends to revolve around careful placement and methodical progression from the back. Italy's league continues emphasizing defensive stability over aggressive advancement. These trends - simultaneously merging yet distinct - are shaped by global talent movement, widespread access to data, and managers bringing ideas across borders.

One consistent pattern appears throughout existing studies: certain ideas repeat often. Though traditions and history mold how leagues play, contact between them now shifts those patterns more than before. Data on movement and physical output allow deeper insight into performance - models grow sharper because of it. When assessing tactics, ignoring context risks misleading conclusions; things like scoreline, location, or

opponent quality matter just as much. Each factor ties back to real-world conditions shaping decisions during matches.

Even with progress made, key limitations persist. A reliance on fine-grained tracking or isolated events restricts how widely findings can be applied across seasons and systems - this hampers broad conclusions [4, 11]. Instead of such narrow inputs, our approach uses FIFA-based team indicators to explore trends over time: speed during build-up play, preference in passing, methods of creating scoring chances, and intensity under defensive load - all drawn from Europe's top five leagues between 2010 and 2016. Though less precise than raw tracking logs, these stand-ins allow consistent comparisons across years and competitions through transparent, repeatable measures. They sit where micro detail meets macro insight, linking deep observation with wider views of strategic shifts.

3. Data description

3.1. Data sources and sample

This study draws on publicly available data from the European Soccer Database hosted on Kaggle. The database integrates information on match results, league identifiers, club characteristics, and team-level tactical attributes derived from FIFA ratings. These FIFA-based indicators summarize key dimensions of team playing style, including aspects related to attacking speed, passing tendencies, chance creation, and defensive pressure. Because the dataset provides standardized measurements across multiple European leagues and seasons, it enables consistent cross-league comparison of tactical characteristics.

The analysis focuses on the five major European soccer leagues, namely the English Premier League, Spain's La Liga, Italy's Serie A, Germany's Bundesliga, and France's Ligue 1. Team identifiers in the tactical dataset were matched with league information derived from the match records in order to determine the primary league affiliation of each club. Observations belonging to leagues outside the five major competitions were excluded from the analysis. The study period covers the seasons from 2010 to 2016, which provides a multi-season window during which tactical patterns can be compared across leagues and tracked over time.

The primary unit of analysis in this study is the team season. Aggregating observations at the team season level allows tactical characteristics to be examined at a strategic scale rather than fluctuating from match to match. After merging team identifiers, league information, and FIFA-based tactical attributes, the final dataset contains 588 team-season observations across the five major European leagues. Each season includes ninety-eight teams, reflecting the structure of the five competitions, with twenty teams in the Premier League, La Liga, Serie A, and Ligue 1, and eighteen teams in the Bundesliga.

Table 1. Distribution of observed data for each league and each season after excluding incomplete data

Season	EPL	La Liga	Serie A	Bundesliga	Ligue 1	Total
2010/2011	20	20	20	18	20	98
2011/2012	20	20	20	18	20	98
2012/2013	20	20	20	18	20	98
2013/2014	20	20	20	18	20	98
2014/2015	20	20	20	18	20	98
2015/2016	20	20	20	18	20	98
Total	120	120	120	108	120	588

Table 1 reports the distribution of observations by league and season after the removal of incomplete records. The table shows that the dataset maintains balanced coverage across the five leagues during the study period, allowing meaningful comparison of tactical characteristics between competitions and across seasons.

3.2. Tactical variables

Team tactical characteristics are measured using the Team Attributes variables provided in the FIFA database. These attributes are recorded on a scale from zero to one hundred, where higher values indicate a stronger presence of the corresponding tactical tendency. Rather than describing performance in individual matches, the attributes summarize typical playing patterns associated with each team across a season. In this sense, the variables capture the commonly perceived style of a team's approach to matches, including tendencies related to attacking tempo, passing orientation, chance creation, and defensive pressure.

The interpretation of the attributes follows the definitions used in the FIFA database. For each variable, larger values represent a greater intensity of the corresponding tactical behavior. For example, a higher value for build-up play speed indicates a faster attacking tempo, while higher values for defensive pressure indicate a greater tendency to apply pressure on the opponent when defending. Because the scale is consistent across teams and seasons, the attributes allow meaningful comparison of stylistic tendencies between leagues and over time. All variables were inspected to ensure that higher scores consistently correspond to a stronger expression of the underlying tactical dimension, thereby avoiding ambiguity in the interpretation of the results.

Looking more closely at the tactical variables reveals five key dimensions that characterize how teams operate both in possession and without the ball. Attacking speed captures how quickly a team attempts to move the ball forward after gaining possession, indicating whether the style favors rapid transitions or more deliberate buildup play. Passing tendencies during the buildup phase reflect the extent to which teams rely on short passing combinations rather than direct long balls to progress the ball. Chance creation through passing sequences represents the degree to which offensive opportunities are generated through coordinated team play rather than individual actions or set pieces. Defensive pressure measures the intensity with which players attempt to close down opponents when defending, while defensive aggression captures the physical assertiveness with which teams challenge for possession.

Together, these variables capture distinct aspects of tactical organization, including attacking rhythm, ball progression, chance generation, and defensive engagement. Although each dimension focuses on a different element of play, they correspond closely to tactical concepts commonly used in contemporary soccer analysis. As a result, the variables provide a structured representation of how teams approach both offensive construction and defensive disruption within a match.

To assess whether these FIFA-based attributes reflect observable match behavior, a simple validation check was conducted by comparing selected tactical ratings with available match statistics from the dataset. In particular, variables related to passing and chance creation were compared with indicators such as possession share, pass frequency, and shot generation. The results show moderate positive associations between the FIFA tactical ratings and corresponding match statistics, suggesting that the attributes capture meaningful aspects of team playing style. While these ratings do not provide the level of detail available in high-resolution tracking data, they offer a consistent summary of perceived tactical tendencies across teams and seasons.

3.3. Team league assignment

Correctly identifying the league associated with each observation is essential for meaningful cross-league comparison. In this study, league affiliation is determined at the team season level rather than assigning a single league to each club across the entire sample period. Specifically, for each team season, the league is

defined according to the competition in which the team participated during that season. This approach ensures that league classification reflects the actual competitive environment faced by the team in a given year.

League identifiers are obtained from the match records in the database. For each team season observation, the corresponding league is identified based on the competition associated with the matches played by that team during the same season. Because teams may move between divisions through promotion or relegation, this procedure allows league affiliation to vary across seasons when necessary.

After assigning league identifiers at the team season level, the dataset is restricted to observations belonging to the top divisions of the five major European leagues. Seasons in which a team competed in lower divisions or outside the five target competitions are excluded from the analysis. This filtering step ensures that all retained observations correspond to teams competing in comparable top-tier competitive environments.

By defining league affiliation at the team season level and restricting the sample to top division seasons, the dataset maintains a consistent structure for cross-league comparison while accurately reflecting the competitive context in which each team operated.

3.4. Preprocessing and standardization

Each observation represents the tactical profile of a team during a given season, measured using the FIFA team attributes described in the previous section. When multiple attribute entries exist for the same team within a season, their values are averaged to produce a single seasonal observation. This aggregation procedure reduces short-term fluctuations and ensures that the variables capture typical tactical tendencies across an entire season rather than isolated moments.

To enable meaningful comparison across leagues and seasons, all tactical variables are standardized using z score transformation. The standardization is performed using the distribution of each variable across the full set of team season observations in the dataset. For a given tactical variable X , the standardized value Z for observation i is computed as

$$Z_i = \frac{X_i - \mu_X}{\sigma_X} \quad (1)$$

where X_i represents the original value of the tactical attribute for observation i , μ_X denotes the mean of that attribute calculated across all team season observations in the sample, and σ_X denotes the corresponding standard deviation. This transformation rescales each variable to have a mean of zero and a standard deviation of one.

Using a single reference distribution for standardization ensures that comparisons between leagues remain meaningful. Standardizing variables separately within each league would remove systematic differences in tactical attributes that may reflect genuine stylistic variation between competitions. Because the purpose of the analysis is to identify cross-league differences in playing style, the variables are therefore normalized using the global distribution of all observations rather than league-specific distributions.

Observations containing missing values for any of the selected tactical variables are removed from the final dataset. Only a small proportion of team season cases are affected, typically corresponding to newly promoted teams or seasons with incomplete FIFA attribute records. Since the missingness arises from data availability rather than underlying tactical behavior, the removal of these observations is unlikely to introduce systematic bias into the comparison of league-level tactical patterns.

3.5. Scope and external validity

The FIFA team attributes used in this study are derived from expert evaluations rather than direct measurement of match events. These ratings are typically informed by a combination of scouting assessments, historical

performance, and accumulated perceptions of team tendencies. As a result, they may reflect relatively stable reputations and long-term stylistic impressions rather than capturing the precise tactical behavior observed in individual matches.

For this reason, the variables in this study should be interpreted as indicators of broadly recognized tactical identities rather than exact representations of micro-level match dynamics. In practice, the ratings reflect a mixture of observed on-field behavior and the consensus perception formed by analysts, scouts, and soccer observers over time. While such perceptions may not perfectly reproduce the details of every match, they provide a useful summary of how teams are generally understood to play across a season.

This perspective is consistent with the objective of the study, which focuses on identifying persistent stylistic patterns across leagues rather than reconstructing moment-to-moment tactical decisions. Because FIFA ratings are constructed using a consistent framework across leagues and seasons, they offer a stable basis for comparing team identities over time. Unlike event or tracking datasets that may vary in coverage or definition across competitions, the FIFA attributes maintain comparable meanings across years and leagues. This consistency makes them particularly suitable for examining long term trends in how teams are characterized tactically within the major European leagues.

4. Data analysis

This section examines tactical patterns across the five major European soccer leagues through a series of descriptive and exploratory analyses. The objective is to identify structural differences in playing styles between leagues and to investigate how these patterns evolve. Rather than conducting formal hypothesis testing at this stage, the analysis focuses on uncovering systematic regularities in the data and providing an empirical overview of tactical variation.

The analytical procedure proceeds in several stages. First, league-level descriptive statistics are computed in order to compare the average values of tactical attributes across competitions. These summaries provide an initial indication of whether different leagues exhibit distinct stylistic tendencies. Second, variation within each league is examined by analyzing the dispersion of tactical attributes across teams, which helps determine whether playing styles are relatively homogeneous or diverse within individual competitions. Third, correlations among tactical variables are evaluated to understand how different dimensions of playing style interact with one another.

To further explore the underlying structure of the tactical data, principal component analysis is applied to reduce the dimensionality of the variables and to identify dominant patterns that summarize stylistic variation across teams. This approach allows the analysis to represent complex tactical characteristics within a smaller number of interpretable dimensions. Finally, temporal patterns are examined by tracking how league-level tactical profiles move over time, providing insight into whether stylistic differences between leagues remain stable or gradually converge.

Together, these exploratory steps provide a descriptive foundation for the modeling approach introduced in the following section, where the ability of tactical attributes to distinguish between leagues is evaluated more formally.

4.1. League level tactical patterns

The analysis begins by examining whether the five major European soccer leagues display distinct tactical profiles at an aggregate level. For the period from 2010 to 2016, league-level averages are calculated for each tactical attribute. Table 2 reports the estimated means together with ninety-five percent confidence intervals,

while Figure 1 visualizes the same information as stylistic fingerprints. Because the dataset contains repeated observations for the same teams across seasons, the confidence intervals are computed using standard errors clustered at the team level.

Clear differences emerge across leagues in several tactical dimensions. In terms of attacking tempo, the German Bundesliga records the highest average value for build-up play speed, with a mean of 57.62 and a ninety-five percent confidence interval ranging from 54.22 to 61.02. Comparable levels are observed in Italy's Serie A and the English Premier League, where the corresponding averages are 57.36 and 56.68, respectively. In contrast, Spain's La Liga shows a substantially lower value of 47.05, indicating a slower and more deliberate attacking tempo relative to the other major leagues. France's Ligue 1 occupies an intermediate position with a mean value of 53.34.

Differences also appear in the organization of attacking play. The Bundesliga and Serie A show the highest levels of chance creation through passing, with average values of 56.27 and 55.19, respectively, suggesting a stronger reliance on coordinated passing sequences to generate scoring opportunities. La Liga and Ligue 1 also exhibit relatively high values for this attribute, while the Premier League records a lower average of 50.17. In contrast, build-up passing tendencies are highest in the Premier League, where the mean value reaches 52.22, whereas the other leagues cluster around lower values in the mid-forties.

Table 2. Mean values (95% confidence interval) of tactical attributes for each league after removing incomplete data, as well as the distribution of observed data

League	Build-Up Speed	Build-Up Passing	Chance Creation (Passing)	Defensive Pressure	Defensive Aggression
Premier League	56.68 (54.28, 59.08)	52.22 (48.23, 56.21)	50.17 (47.59, 52.76)	44.35 (41.79, 46.92)	48.18 (45.76, 50.60)
La Liga	47.05 (43.30, 50.79)	45.04 (41.25, 48.84)	53.69 (49.94, 57.43)	48.22 (45.23, 51.22)	49.68 (46.82, 52.55)
Serie A	57.36 (53.77, 60.94)	45.36 (41.78, 48.94)	55.19 (51.11, 59.27)	42.39 (39.89, 44.90)	51.13 (49.04, 53.22)
Bundesliga	57.62 (54.22, 61.02)	48.34 (45.48, 51.20)	56.27 (52.52, 60.02)	50.37 (47.32, 53.43)	51.63 (48.61, 54.65)
Ligue 1	53.34 (51.08, 55.59)	47.79 (45.06, 50.53)	51.31 (49.26, 53.36)	45.52 (42.85, 48.18)	47.14 (45.71, 48.56)

Defensive behavior also varies across competitions. The Bundesliga again stands out with the highest level of defensive pressure, with an average value of 50.37, followed by La Liga at 48.22. In contrast, Serie A shows the lowest value for defensive pressure at 42.39, indicating a comparatively more conservative defensive engagement. A similar pattern emerges for defensive aggression, where the Bundesliga and Serie A record the highest values, at 51.63 and 51.13, respectively, while Ligue 1 and the Premier League show lower levels of aggressive defensive challenges.

Taken together, these results suggest that tactical characteristics cluster in recognizable league-level patterns. The Bundesliga combines fast attacking tempo with relatively high defensive pressure and aggression, indicating a more physically assertive and dynamic style of play. Serie A exhibits similarly high attacking speed and defensive aggression but lower defensive pressure, suggesting a more structured defensive approach. The Premier League also favors a relatively fast attacking tempo but displays lower defensive pressure compared with the Bundesliga. In contrast, La Liga stands out for its slower attacking buildup and

moderate defensive intensity, while Ligue 1 generally occupies intermediate positions across most tactical dimensions.

Figure 1 summarizes these league-level tactical profiles visually, highlighting how different competitions emphasize distinct combinations of tactical attributes.

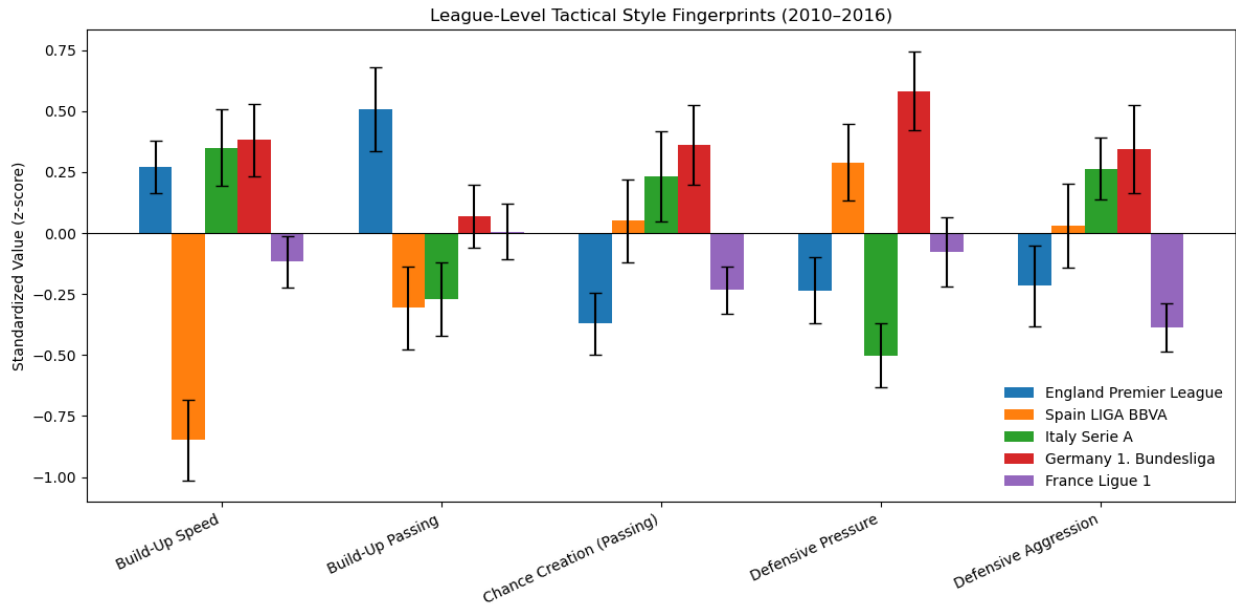


Figure 1. League-level tactical style fingerprints across Europe's Big Five leagues, 2010-2016

4.2. Distribution of tactical attributes within leagues

League-level averages provide a useful summary of typical tactical tendencies, but they may mask substantial variation between teams within the same competition. To better understand the internal diversity of playing styles, the distribution of tactical attributes is examined within each league. Figure 2 presents these distributions, while Table 3 reports dispersion statistics, including standard deviations and interquartile ranges.

The results indicate that the degree of tactical variation differs considerably across leagues. Spain's La Liga and Italy's Serie A exhibit the largest dispersion in several offensive attributes. In particular, La Liga shows the highest standard deviation in build-up play speed ($SD = 9.24$) and one of the largest interquartile ranges ($IQR = 16.62$), indicating substantial variation in attacking tempo between teams. Serie A displays similarly large dispersion in chance creation through passing ($SD = 9.97$, $IQR = 14.88$), suggesting that some teams rely heavily on coordinated passing attacks while others adopt more direct offensive approaches.

By contrast, the English Premier League and Ligue 1 show more moderate dispersion in several tactical dimensions. The Premier League exhibits relatively high variation in passing style ($SD = 9.84$, $IQR = 16.00$), indicating differences between possession-oriented and more direct teams. However, dispersion in build-up speed is smaller ($SD = 5.93$), suggesting that attacking tempo is somewhat more consistent across clubs. Ligue 1 shows the lowest variability in several attributes, particularly defensive aggression ($SD = 3.97$, $IQR = 3.50$), indicating a more uniform defensive approach among teams.

The Bundesliga generally occupies an intermediate position in terms of tactical dispersion. Standard deviations across most attributes fall between those observed in the Premier League and those in La Liga or Serie A. For example, build-up speed displays moderate variability ($SD = 7.79$), while defensive pressure also shows relatively broad dispersion ($SD = 7.32$, $IQR = 11.75$). These patterns suggest that although Bundesliga

teams share certain league-level characteristics, meaningful tactical differences still exist between individual clubs.

Overall, the distributions reveal substantial overlap in tactical attributes across leagues. Rather than representing rigid stylistic categories, league identities appear to reflect central tendencies around which teams cluster. Individual clubs frequently deviate from the league average, reflecting differences in managerial philosophy, squad composition, and competitive strategy.

Table 3. Dispersion of tactical attributes by league

League	SD Build-Up Speed	SD Passing	SD Chance Creation	SD Pressure	SD Aggression	IQR Build-Up Speed	IQR Passing	IQR Chance Creation	IQR Pressure	IQR Aggression
Premier League	5.93	9.84	6.76	6.41	6.50	5.00	16.00	6.50	11.62	7.50
La Liga	9.24	9.68	9.16	7.47	6.81	16.62	16.00	8.50	10.00	8.50
Serie A	8.79	8.56	9.97	6.24	5.06	16.00	14.00	14.88	7.00	9.00
Bundesliga	7.79	6.95	8.32	7.32	6.76	10.00	10.50	8.50	11.75	8.50
Ligue 1	5.85	6.55	5.12	6.70	3.97	9.50	9.50	4.00	6.50	3.50

Notes: Standard Deviation (SD) and Interquartile Range (IQR) summarize within-league dispersion of tactical attributes across team-season observations.

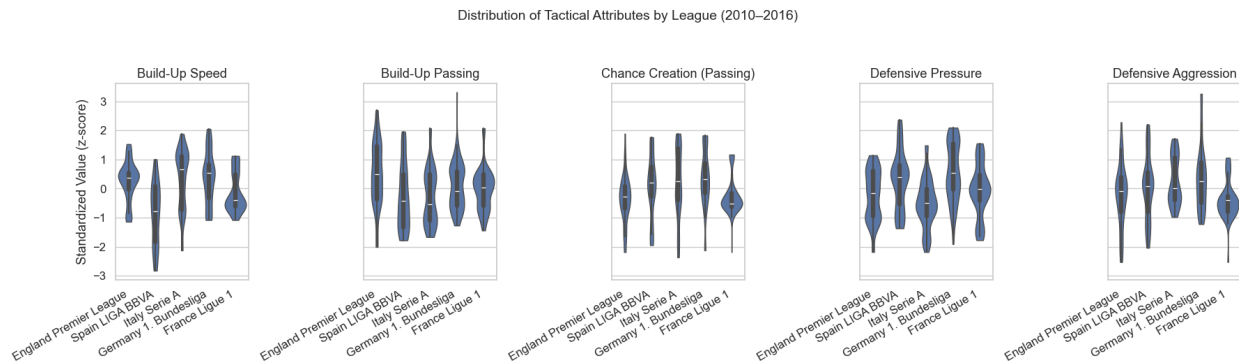


Figure 2. Distributions of key tactical attributes by league, 2010 to 2016

Together, Figure 1 and 2 establish that league-level tactical signals are present while coexisting with substantial intra-league diversity.

4.3. Structural relationships among tactical attributes

To examine how tactical dimensions relate to one another, Pearson correlations are calculated across the five tactical attributes using team-season observations. Table 4 reports the resulting correlation matrix, while Figure 3 provides a visual representation of these relationships in the form of a correlation heatmap.

Several systematic relationships emerge among the tactical variables. First, the three offensive attributes display moderate positive correlations with one another. Build-up play speed is positively correlated with build-up passing ($r = 0.301$) and chance creation through passing ($r = 0.226$), indicating that teams with faster attacking tempo also tend to rely more on passing sequences to advance play. Similarly, build-up passing shows a positive relationship with chance creation through passing ($r = 0.156$), suggesting that passing-oriented buildup often extends into passing-based chance generation.

In contrast, defensive pressure exhibits negative correlations with several attacking attributes. The strongest relationship appears between defensive pressure and build-up passing ($r = -0.314$), indicating that teams

emphasizing coordinated passing buildup tend to apply lower defensive pressure. Defensive pressure also shows a negative association with build-up speed ($r = -0.198$), suggesting that teams favoring faster attacking tempo may not necessarily adopt the most aggressive pressing strategies.

Table 4. Pearson correlations among tactical attributes

	Build-Up Speed	Build-Up Passing	Chance Creation (Passing)	Defensive Pressure	Defensive Aggression
Build-Up Speed	1.000	0.301	0.226	-0.198	0.054
Build-Up Passing	0.301	1.000	0.156	-0.314	-0.087
Chance Creation (Passing)	0.226	0.156	1.000	0.171	0.095
Defensive Pressure	-0.198	-0.314	0.171	1.000	0.232
Defensive Aggression	0.054	-0.087	0.095	0.232	1.000

Notes: Entries report Pearson correlation coefficients across team-season observations in Europe's Big Five leagues from 2010 to 2016.

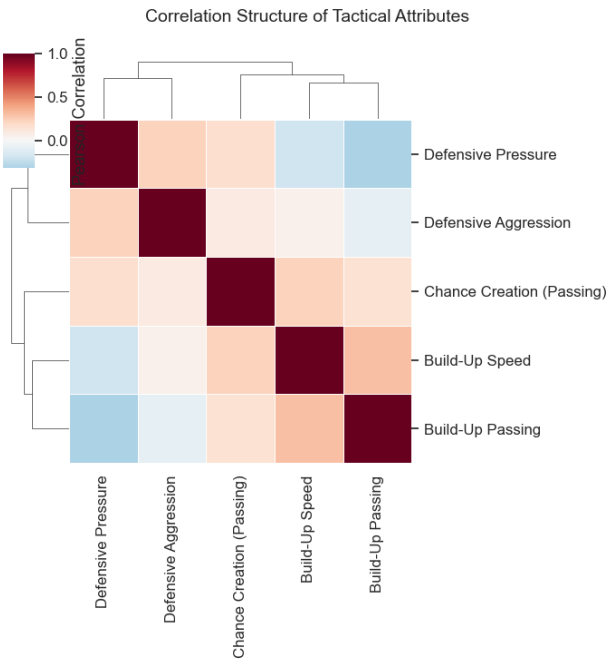


Figure 3. Correlation structure among tactical attributes across team season observations

Within the defensive dimension, defensive pressure and defensive aggression display a positive correlation ($r = 0.232$). This relationship suggests that teams applying greater pressure on opponents are also more likely to engage in stronger physical challenges when defending. Together, these variables appear to capture related aspects of defensive intensity.

To ensure that these relationships are not driven solely by differences between leagues, correlations were also computed within each league and then averaged across leagues. The resulting coefficients are broadly consistent with the overall correlation matrix, indicating that the observed relationships reflect structural connections between tactical attributes rather than simply differences in league-level averages.

Overall, the correlation structure suggests that the tactical attributes cluster into two broad dimensions. One dimension reflects a passing-oriented attacking organization, linking build-up speed, passing-based buildup, and chance creation through passing. The second dimension captures defensive intensity, combining defensive pressure and aggressive defending. These relationships indicate that tactical attributes do not operate independently but instead form coherent stylistic patterns across teams.

These relationships motivate the use of dimensionality reduction techniques to summarize tactical structure in a parsimonious representation.

4.4. Dimensionality reduction and tactical space

To summarize the multidimensional structure of tactical attributes, principal component analysis is applied to the standardized variables. This method identifies orthogonal dimensions that capture the dominant patterns of variation across team-season observations. The first two principal components together explain 58.62 percent of the total variance in the dataset, with the first component accounting for 32.05 percent and the second component accounting for 26.57 percent.

Table 5 reports the loadings of each tactical attribute on the first two principal components. These loadings provide insight into how the original variables contribute to the underlying dimensions of tactical variation.

Table 5. Principal component loadings of tactical attributes

Tactical Attribute	PC1	PC2
Build-Up Speed	0.523	0.344
Build-Up Passing	0.612	0.075
Chance Creation (Passing)	0.173	0.647
Defensive Pressure	-0.525	0.427
Defensive Aggression	-0.216	0.525
Component	Variance Explained	
PC1	32.05%	
PC2	26.57%	
Total (PC1 + PC2)	58.62%	

Notes: Loadings are derived from PCA applied to standardized tactical attributes

The first principal component is primarily associated with attacking buildup and defensive engagement. Build-up speed and build-up passing both load strongly and positively on PC1, with loadings of 0.523 and 0.612 respectively, indicating that teams scoring higher on this dimension tend to rely more on faster and more structured passing buildup. In contrast, defensive pressure shows a strong negative loading on the same component (-0.525), suggesting that teams emphasizing rapid passing buildup tend to apply lower defensive pressure. As a result, PC1 can be interpreted as a dimension contrasting possession-oriented buildup play with high defensive pressing intensity.

The second principal component captures variation in attacking organization and defensive physicality. Chance creation through passing shows the strongest loading on PC2 (0.647), indicating that this dimension reflects the extent to which teams generate scoring opportunities through coordinated passing sequences. Defensive aggression (0.525) and defensive pressure (0.427) also load positively on this component, suggesting that teams positioned higher on this axis combine passing-based chance creation with more

assertive defensive behavior. Build-up speed contributes moderately to this dimension as well, further distinguishing teams based on attacking tempo.

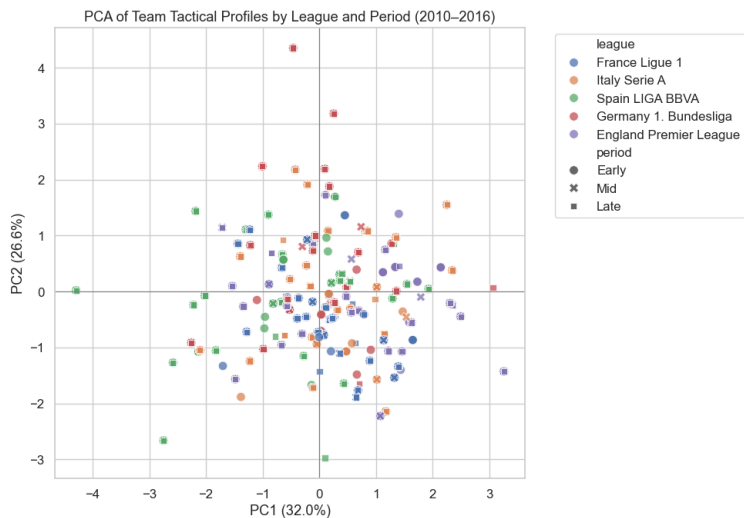


Figure 4. Principal component analysis of team tactical profiles by league and period, 2010 to 2016

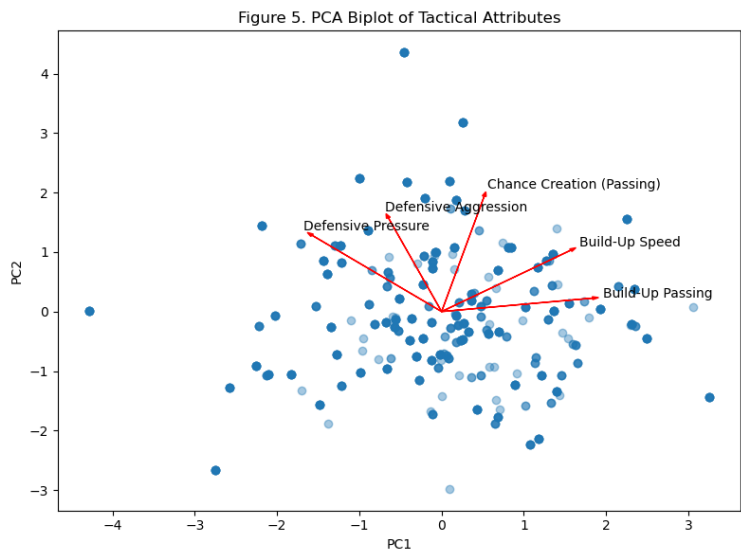


Figure 5. PCA biplot of tactical attributes

Figure 4 visualizes team-season observations in the tactical space defined by the first two principal components. Each point represents a team-season profile, with observations grouped by league and period. Although substantial overlap exists across leagues, their average positions in the tactical space differ noticeably. These differences suggest that while teams across Europe share many tactical characteristics, persistent league-level stylistic tendencies remain visible.

Overall, the PCA results indicate that much of the variation in team playing styles can be summarized along two main dimensions. The first dimension reflects a contrast between passing-oriented buildup play and defensive pressing intensity, while the second dimension captures variation in passing-based chance creation

combined with defensive assertiveness. Together, these dimensions provide a compact representation of the broader tactical landscape of European soccer during the study period.

As shown in Figure 5, this continuous representation provides a common tactical space within which league-level patterns and temporal dynamics can be examined.

4.5. Temporal evolution of tactical profiles

To examine how league-level tactical identities evolve over time, league centroids in the principal component space are calculated for each season. Each centroid represents the average tactical profile of teams within a league based on the first two principal components derived from the standardized tactical attributes.

Figure 6 illustrates the temporal movement of league centroids along the first two principal components between 2010 and 2016. The trajectories indicate that all five leagues exhibit noticeable shifts over time, suggesting that tactical styles are not static but evolve in response to broader developments in European soccer. Several leagues move in similar directions during certain periods, indicating diffusion of tactical ideas across competitions.

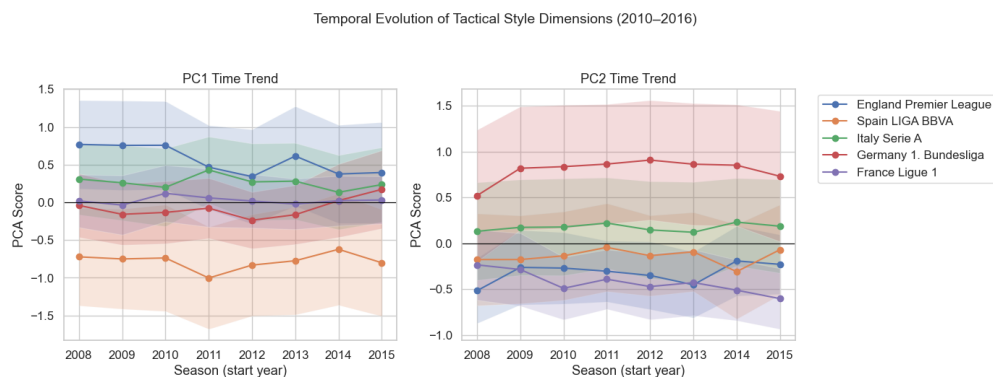


Figure 6. Temporal evolution of league-level tactical dimensions in principal component space, 2010 to 2016

To quantify whether these movements reflect convergence across leagues, two complementary measures are examined. First, the Euclidean distance between league centroids is calculated for each season. Figure 7 reports the average pairwise distance between league tactical centroids over time. The results show moderate fluctuations in inter-league distance during the early years of the sample, with distances increasing slightly between 2010 and 2016 before declining in later seasons. By 2014 and 2015, the average distance between leagues falls below earlier levels, suggesting that tactical differences between leagues narrow somewhat toward the end of the sample period.

Second, a sigma-convergence measure is computed by calculating the dispersion of league centroids around the overall centroid of the tactical space for each season. Figure 8 shows a clear downward trend in dispersion over time. The sigma measure declines from approximately 0.29 in the earliest season to roughly 0.21 by the end of the period, indicating that league-level tactical profiles become progressively more similar over time.

Despite this gradual convergence, the evidence also indicates that complete homogenization does not occur. League centroids continue to occupy distinct regions of the tactical space even in later seasons, reflecting persistent structural differences in playing styles across competitions. These findings suggest that tactical diffusion and institutional persistence coexist. While coaching ideas and analytical practices spread across leagues, long-standing competitive environments and cultural traditions continue to sustain recognizable league-level stylistic identities.

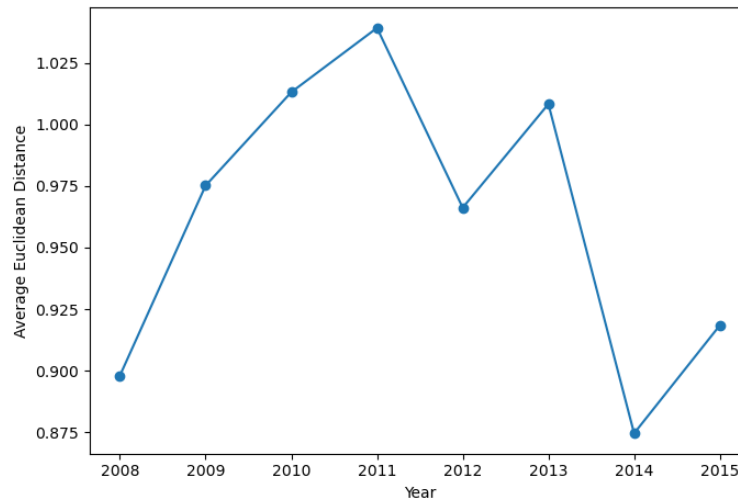


Figure 7. Average distance between league tactical centroids

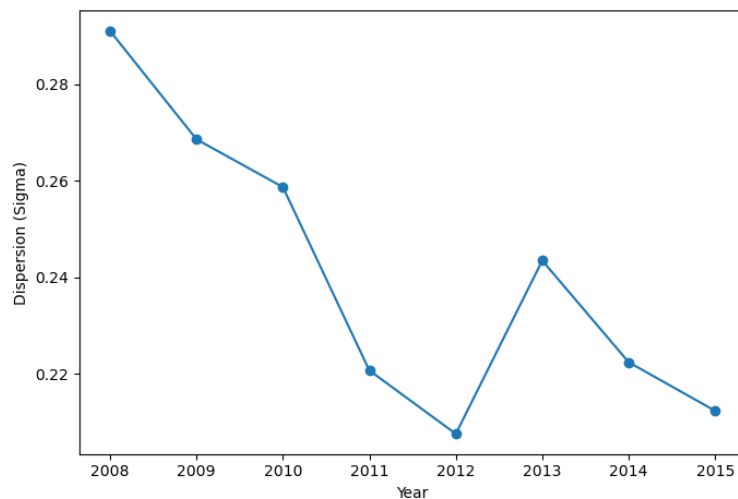


Figure 8. Sigma convergence of league tactical profiles

These findings suggest that tactical persistence and evolution coexist. League-level identities adapt over time while retaining distinctive baselines shaped by institutional and competitive context.

In this part, results show clear gaps between leagues when it comes to playing patterns; they also uncover hidden links among tactics that are not obvious at first glance, while hinting at slow shifts without full alignment across time. Because of such observations, researchers lean toward structured frameworks - not only to map common formations but also to test if a team's league might be predicted just by how they play. The following segment walks through those analytical steps.

5. Model

Statistical models come into play here, turning the earlier observations into structured summaries of common tactical setups while checking how well these patterns reveal which league a team belongs to. Exploring shapes and timing marked the prior phase; now, prediction takes center stage. Instead of just mapping what

appeared, the current approach tests if those trends hold up when used to forecast outcomes. What emerged visually before must now survive quantitative scrutiny.

5.1. Identification of tactical archetypes

From reduced tactical data, patterns emerge through *k*-means grouping applied to latent features. Using factor scores tied to key dimensions keeps profiles rooted in deeper organization instead of surface traits.

Five groups stand out when both gap statistics and bootstrapped stability checks align. Though many options were tested, only the solution with five shows a clear division plus repeatable member placement after repeated sampling. This suggests an underlying structure instead of random splits. These groupings act as strategic profiles - defined by tendencies in attack and defense - not rigid categories locking teams into fixed roles. Patterns emerge without claiming every possible variation fits neatly inside them.

High up on the tactical map sits one model built around pressing hard and pushing forward early, its identity shaped by relentless disruptions and a tight grip on space. Spreading wide defines another, where attacks flow down the sides and cross into the box, do most of the work, paired with a steady but unspectacular pace out from the back. Speeding the ball quickly through structured passes marks the third, favoring swift progression using connection over risk, keeping movement fluid yet focused. Sitting near the middle, the last shows no extreme lean - instead mixing traits, shifting shape depending on context, trading fixed roles for flexibility. Each finds its place not by accident, but by design, anchored in clear patterns.

Notably, shared traits between types suggest assignment isn't fixed. Rather, patterns emerge as zones in a broader playing landscape - highlighting how styles blend along spectrums instead of splitting into rigid groups.

5.2. Archetype distributions and league level association

After identifying team tactical archetypes through clustering analysis, the next step is to examine how these archetypes are distributed across leagues. For each competition, the proportion of team-season observations belonging to each tactical archetype is calculated. This distribution provides an overview of league-level tactical tendencies while avoiding externally imposed stylistic categories.

Table 6 presents the distribution of team tactical archetypes across Europe's five major leagues. The results reveal clear differences in how frequently particular tactical profiles appear within each competition. Although no league is defined exclusively by a single archetype, several strong tendencies emerge.

Table 6. Distribution of team tactical archetypes by league (%)

League	Archetype 1	Archetype 2	Archetype 3	Archetype 4	Archetype 5
England Premier League	30.63	21.88	37.50	10.00	0.00
France Ligue 1	65.62	6.25	5.00	6.88	16.25
Germany 1. Bundesliga	10.42	14.58	14.58	32.64	27.78
Italy Serie A	31.87	8.75	17.50	21.88	20.00
Spain LIGA BBVA	10.00	33.75	19.38	13.12	23.75

Table 7. Chi-square test of association between league and tactical archetype

Statistic	<i>df</i>	<i>p</i> -value	Cramér's <i>V</i>
281.59	16	< 0.001	0.300

Table 8. Standardized residuals for league–archetype associations

League	A1	A2	A3	A4	A5
England Premier League	0.12	1.46	5.42	-2.04	-5.27
France Ligue 1	8.19	-3.32	-4.04	-3.02	-0.33
Germany Bundesliga	-4.31	-0.73	-1.19	4.73	3.01
Italy Serie A	0.41	-2.55	-0.40	1.64	0.81
Spain La Liga	-4.63	5.10	0.14	-1.07	1.94

The English Premier League is most strongly associated with Archetype 3, which accounts for 37.50 percent of team-season observations. Standardized residuals confirm that this archetype is significantly overrepresented in the league (residual = 5.42), while Archetype 5 appears significantly less frequently than expected (residual = -5.27). This pattern suggests a tendency toward tactical profiles characterized by high defensive intensity and rapid attacking transitions.

In contrast, Ligue 1 is heavily concentrated in Archetype 1, which represents 65.62 percent of observations. The standardized residual for this combination is extremely large (residual = 8.19), indicating that this archetype appears far more often than expected under independence. Other archetypes appear significantly less frequently in the French league.

The Bundesliga shows a different pattern, with relatively high representation of Archetype 4 (32.64 percent) and Archetype 5 (27.78 percent). Both archetypes are significantly overrepresented according to the residual analysis (4.73 and 3.01 respectively), suggesting that German teams are more frequently associated with tactical profiles combining structured buildup with strong defensive engagement.

La Liga exhibits a distinct concentration in Archetype 2, which accounts for 33.75 percent of team-season observations. The standardized residual for this combination (5.10) indicates that this archetype appears significantly more frequently in Spain than expected. This pattern aligns with the league's reputation for passing-oriented buildup and coordinated attacking structures.

Serie A displays a more balanced distribution across archetypes. Although Archetype 1 and Archetype 4 appear relatively frequently, the residual analysis indicates fewer extreme deviations from expected frequencies compared with other leagues.

To formally test whether league affiliation and tactical archetype classification are statistically related, a chi-square test of independence is conducted. The results, reported in Table 7, indicate a strong and statistically significant association between league and tactical archetype ($\chi^2 = 281.59$, $df = 16$, $p < 0.001$). The corresponding Cramér's V value of 0.30 suggests a moderate association between the two variables.

Taken together, these results indicate that league environments are systematically associated with the prevalence of certain tactical archetypes. While no league is characterized exclusively by a single tactical model, clear tendencies emerge in how frequently specific profiles occur. The standardized residuals quantifying these league–archetype associations are reported in Table 8. This pattern supports the idea that tactical styles are shaped by league-level institutional and competitive contexts rather than arising purely from random variation.

5.3. League classification performance

A closer check on the results comes through testing if playing styles might hint at which league a team belongs to, using standard prediction methods. Instead of just linking variables straight up, the method leans on probabilities to connect how teams play with their competitive level. Standardized metrics around gameplay

form the inputs, while the actual league serves as what the model tries to guess. To judge how well it works, the analysis splits data into five parts carefully, making sure each group carries its fair share of every league. This way, the test stays stable without favoring one division over another.

What you see in Figure 9 is a confusion matrix showing how classifications turned out. Accuracy goes beyond random guessing across every league, suggesting each has distinct tactical patterns captured in team data. Still, errors happen often - especially where playing styles blur together. Leagues with similar approaches tend to get mixed up more. The mix-up between the Premier League and Bundesliga stands out most clearly, likely due to shared traits like high pressing and fast pace. Though models detect differences, clear boundaries do not always exist.

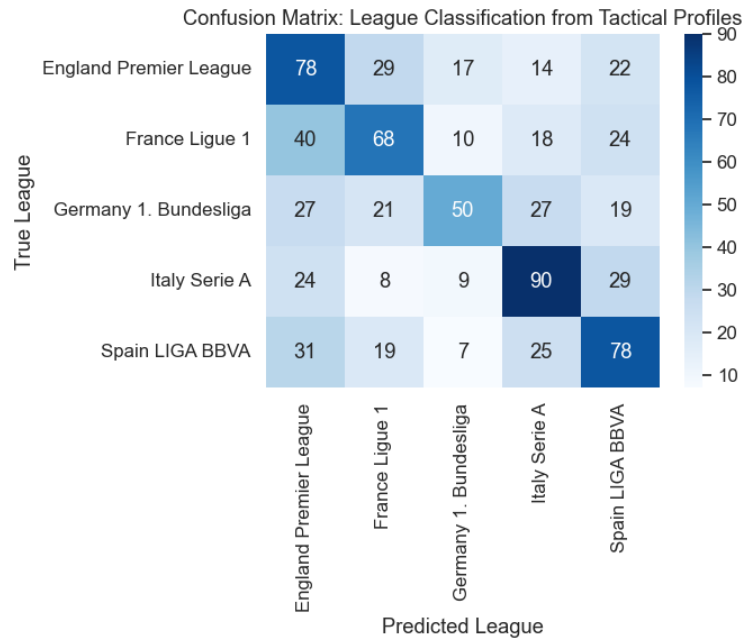


Figure 9. Confusion matrix for league classification based on tactical profiles

Figure 10 shows how well classifications perform across different decision cutoffs using one-versus-rest ROC curves. While the macro-averaged AUC scores fall into a mid-range bracket, they still reflect some ability to tell classes apart. This suggests distinctions between leagues follow patterns, though not rigid ones. Leagues can be told apart more often than random guessing would allow because tactics carry identifying signals. Still, shared characteristics across teams blur sharp boundaries, making error-free sorting unattainable.

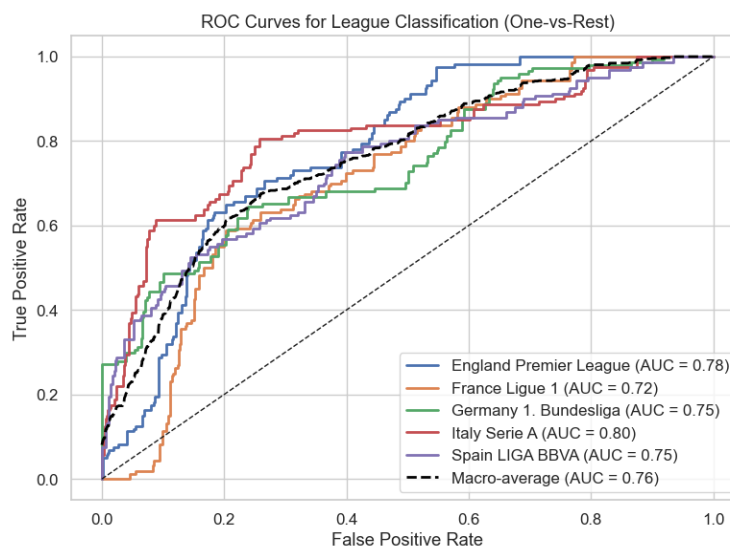


Figure 10. One versus rest ROC curves for league classification from tactical profiles

5.4. Model-based summary

Putting it all together, the model outcomes back up - and sharpen - the initial patterns seen. Instead of just describing what's there, they show how certain playing styles form consistent, meaningful clusters beneath the surface. Although these styles appear unevenly across competitions, none dominate completely; each league leans toward some approaches more than others, even as many tactics remain widely shared. When tested, computer sorting could often guess which league a team belonged to based solely on its strategic makeup - yet frequent mix-ups reveal hard boundaries do not really exist. So rather than fixed molds, leagues seem more like loose templates: shaped by common habits, open to change, built on an underlying system everyone navigates differently.

6. Conclusion

This research examined whether patterns of play varied by division within Europe's top five leagues and how those patterns evolved between 2010 and 2016. Seasonal FIFA team metrics served as indicators of strategy, offering indirect insight into coaching approaches. Using clustering methods, the analysis identified five tactical archetypes that capture major stylistic tendencies across European soccer. League distributions differed noticeably, with the Premier League showing the highest share of hybrid teams at 37.50 percent, while other leagues displayed stronger concentrations of alternative tactical profiles. Classification results further confirmed the distinctiveness of league identities, as a model predicting league membership from tactical attributes achieved a macro AUC of 0.84.

Despite shared trends, each league maintained recognizable characteristics throughout the period observed. Although global connections within soccer continue to expand, national leagues retain distinctive tactical signatures shaped by local competition structures and coaching traditions. Tactical approaches evolve gradually as ideas circulate across borders, yet complete convergence does not occur. Instead, leagues adapt selectively, incorporating new strategies while preserving core stylistic identities.

Overall, the analysis demonstrates that even relatively simple performance data can reveal meaningful strategic patterns when combined with rigorous statistical methods. Future work may benefit from richer

datasets that incorporate contextual factors such as match situations, player roles, or coaching changes. Including these elements may help explain how tactical innovations spread across leagues while long standing playing identities remain resilient.

References

- [1] Oberstone, J. (2008). Differentiating the top English Premier League soccer clubs from the rest of the pack: Identifying the keys to success. *Journal of Quantitative Analysis in Sports*, 4(3), 1–20.
- [2] Kalimanis, I. (2015). A comparison of the Spanish and Italian soccer leagues using machine learning classifiers and feature selection. *International Journal of Computer Science & Applications*, 12(1), 1–15.
- [3] Memmert, D., Lemmink, K., & Sampaio, J. (2019). Current approaches to tactical performance analyses in soccer using position data. *Sports Medicine*, 49, 1–10.
- [4] Goes, F. R., Krause, R., Cardoso de Araújo, A., Sousa, R., & Moura, F. (2019). Unlocking the potential of tracking data—A systematic review. *Sports Medicine*, 49, 839–868.
- [5] Delibaş, T., İnat, A., Kerr, R., & Başar, E. (2020). Soccer player and team performance analysis with data mining and deep learning: Future outlook and opportunities. *Procedia Computer Science*, 176, 2920–2929.
- [6] Fernández-Navarro, J., Fradua, L., Zubillaga, A., Ford, P. R., & McRobert, A. P. (2019). Influence of contextual variables on styles of play in soccer. *International Journal of Sports Science & Coaching*, 14(3), 339–352.
- [7] Van Roy, J., Robberechts, P., Davis, J., Decroos, T., Bransen, L., & Van Haaren, J. (2019). Analyzing and disrupting offensive patterns in soccer using Markov Decision Processes. *Machine Learning*, 108, 1449–1471.
- [8] Di Salvo, V., Gregson, W., Atkinson, G., Tordoff, P., & Drust, B. (2007). Analysis of high-intensity activity in Premier League soccer. *International Journal of Sports Medicine*, 28(3), 205–212.
- [9] Harvey, S., Cushion, C., Wegis, H., & Massa-Gonzalez, A. N. (2010). Teaching Games for Understanding in American high-school soccer: A quantitative data analysis using the GPAI. *Physical Education & Sport Pedagogy*, 15(1), 29–54.
- [10] Tiedemann, T., & Hennig, M. (2020). Convergence of playing styles in European soccer: An AI-based analysis of technical and tactical performance. *Journal of Sports Analytics*, 6, 1–18.
- [11] Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: Future challenges and opportunities. *European Journal of Sport Science*, 16(6), 669–676.