Strathmore UNIVERSITY

Voronoi Diagrams and How They Shape up Offense Analytics in Women's Football

by

Mugwe Alvin Igobwa

092975

Submitted in Partial fulfilment of the Requirements for the Degree of Master of Science in Data Science and Analytics at

Strathmore University

Strathmore Institute of Mathematical Sciences
Strathmore University

Nairobi Kenya

June, 2024 .

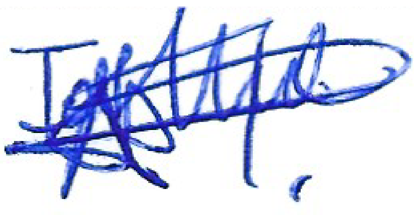
Abstract

Vilar et al. (2013) introduces a method for analyzing collective offensive and defensive behavior, finding that maintaining numerical dominance in key areas of the field is crucial for both defensive stability and offensive opportunity.The consideration of offensive tactics we try to employ is looking at spotting defensive weaknesses, expected goal improvements and exploiting the opposing team's defense when attacking. The use of the expected goal metric is important to a team as it serves beneficial from the aspect of seeing where to improve the offense by creating opportunities that have higher expected goals, and as well help in the defense by learning the expected model of the other teams and adequately positioning the team in order to make the opponent make shots from the low expected goal regions. The expected goal metric to be used will employ the use of machine learning techniques such as logistic regression, bagging algorithms,decision trees and deep learning techniques such as Multilayer Perceptron models so as to help in the dealing with the imbalanced goals variable. The expected goals model cannot be a stand alone feature and would need the incorporation of other metrics to determine what key factors per team lead to the creation of higher goal scoring opportunities, because of this, voronoi diagrams were used in the exploration of how different team shapes at different moments during the game lead to either more goals or chances being created dependant on the space that the team occupies.

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Student's Name: Mugwe Alvin Igobwa

Student's Signature: 

Supervisor's Name: Dr Christopher Njunge

Supervisor's Signature: ________________________ Date: ___________________

Contents

INTRODUCTION

Background of The Study

Sports analytics has undergone an incredible transformation, shifting from fundamental statistical analysis to convoluted strategies utilizing big data, computer vision, machine learning, and artificial intelligence (AI) Tuyls et al. (2021). The shift has impacted many sports disciplines, changing not only how we evaluate player performance and game dynamics but also how we develop strategy, scout, manage health, and engage fans. In sports, football has led the way in this analytical revolution. A greater understanding of tactics, player effectiveness, and team dynamics has been made possible by the sport's use of a variety of cutting-edge data collection and analysis systems. In recent years, artificial intelligence (AI) has been used in football analytics to make real-time strategy modifications during games, leverage natural language processing to monitor fan sentiment, and predict player injuries using machine learning algorithms. These developments indicate a shift in the sport toward more prescriptive and predictive analytics Tuyls et al. (2021)

A team's ability to score more goals than their opponent determines their overall performance in football, often known as soccer Arriaza and Zuniga (2016). The observation that teams higher in the end season rankings scored more goals in leagues like the English Premier League, Spanish Premiera, German Bundesliga, and Australian League shows how important scoring goals is in Hewitt, Greenham, and Norton (2016). Teams that were ranked higher in these contests scored two goals on average per game, while teams that were ranked lower scored one goal on average. The essential role that goals play in fostering team success is also evident in women's football, as seen by the National Womens Soccer League (NWSL), where it has been demonstrated that teams with better achievement levels score more goals overall Sumpter (2017). There is a cyclical continuum between the attacking and defensive phases. Teams must adjust to phase transitions in matches since the proper behaviors in one phase are diametrically opposite to the suitable behaviors in the other.

The use of AI and machine learning to improve player recruiting and performance analysis is one noteworthy development in football analytics. According to Tuyls et al. (2021), artificial intelligence (AI) tools are creating previously unheard-of analytical opportunities in football, such as the analysis of player and team behavior, which is essential for performance enhancement and strategic planning Tuyls et al. (2021). According to Apostolou and Tjortjis (2019), research presents how algorithms can forecast a player's performance in the future, which can help with transfer decisions and tactical modifications.

The use of sophisticated data analysis by professional sports teams has begun to give them a competitive edge in recent decades. However, because it can be challenging to understand the many spatiotemporal interactions in the game, soccer has lagged behind other sports in integrating analytics Sumpter (2017) Pollard and Reep (1997). We need a flexible framework that can both provide useful interpretations of actual game scenarios and capture the intricate spatial and contextual aspects that govern the game, in order to handle the constant stream of inquiries that coaching staff encounters on a regular basis. We discusses offensive methods that we will attempt to use by identifying defensive flaws, anticipating goal upgrades, and attacking by taking advantage of the defense of the opposition. A team can benefit from using the expected goal metric in two ways: first, by identifying areas where the offense can be strengthened by creating opportunities with higher expected goals; second, by understanding the expected model of the opposition and positioning the team appropriately to force shots from low expected goal regions, the defense can also benefit from using it Sumpter (2017). The majority of research on team offensive performance analysis has concentrated on scoring scenarios and play patterns that result in a shot on goal. The particulars of a particular game, however, have frequently been overlooked while analyzing these scenarios, which may have an impact on the tactical alternatives open to teams for both offense and defense. Voronoi discovered in 1908 how to divide all of space using specifically made polygons among a group of points. One of the most significant structures in computational geometry is a Voronoi diagram Fernandez-Navarro et al. (2016). It contains details regarding what is similar to what. To put it more exactly, every point is surrounded by a special limiting convex polygon, and every point inside of a point's polygon is closer to it than every other point. In soccer, Voronoi diagrams can be applied to at least two different ideas: the conventional usage of Voronoi diagrams as a measure of dominance and Sumpter's use of them to gauge offensive power Sumpter (2017).

Furthermore, analytics now play a revolutionary role in the management and prevention of injuries. Research, for instance, has demonstrated the substantial impact that data-driven strategies may have in identifying and reducing the likelihood of injuries, improving player health care, and prolonging the careers of players Daniel (2018). Moreover, teams are now able to establish stronger ties with their fans by providing individualized experiences that are tailored to each supporter's tastes and emotions thanks to the incorporation of analytics into improving fan engagement tactics Coutts (2014). Another area of football analytics is the integration of computer vision and tools for spatial analysis, such as Voronoi diagrams, into real-time game strategy. Teams can obtain a strategic edge by using player positioning data to better understand the spatial dynamics and player interactions during games. Making strategic decisions based on this level of study yields important insights.

Football analytics looks to have a bright future ahead of it with more technological integration as the field grows. Football's competitive scene is anticipated to be further enhanced by innovations in augmented and virtual reality, blockchain for digital fan experience management, and IoT devices for player performance tracking Daniel (2018),Coutts (2014). Not only will these developments improve the quality of play, but they also have the potential to completely transform the fan experience and increase the sport's appeal to a worldwide audience. The growing discipline of analytics, which demonstrates the continued blending of football and technology, heralds a new era in which data not only informs but also changes the beautiful game.

We discuss offensive methods that we will attempt to use by identifying defensive flaws, anticipating goal upgrades, and attacking by taking advantage of the defense of the opposition. An organization can benefit from using the expected goal metric in two ways: first, by identifying areas where the offense can be strengthened by creating opportunities with higher expected goals; second, by understanding the expected model of the opposition and positioning the team appropriately to force shots from low expected goal regions, the defense can also benefit from using it.

Problem Statement

Given the intricacy of team sport performance, comprehending a team's offensive actions and interactions (i.e., player positioning and passing networks) is essential to comprehending the requirements for any particular match activity. The technical output of individual players alone, out of the context of the team, is insufficient to obtain meaningful information about a player's physical output because the athletes' activity profiles are contextualized by the tactics used (e.g., fast/slow ball movement) during different stages of match play. In sports, complex and social network analysis looks at how players interact with one another and how particular aspects of collaboration affect performance Sumpter (2017). Neveretheless, spatiotemporal data analyze player positioning collectively during a match to study team behavior and its connection to performance. Thus, the positions and interactions of players during a match have been observed using spatiotemporal data from GPS units and network analysis techniques, offering insights into the strategies of different teams Gomez et al. (2018). Although soccer has been the subject of the majority of research employing these methods, the National Womens Soccer League (NWSL) has recently begun to use these analytical strategies. Given soccer is a 360 game in which players can pass in any direction, comparable analysis techniques should be helpful in deciphering the team dynamics and coordinated activities that occur during the matches. Understanding how team tactics affect activity profiles can help with training design and team strategy during games. This knowledge could help coaches by giving them an idea of the physical skills needed to carry out the intended game plan under the pressure of AFL competition Scott, Haigh, and Lovell (2020). Although match-play activity needs are well documented, technical and tactical considerations—two crucial aspects of match-play—are frequently overlooked in these evaluations Freitas, Volossovitch, and Almeida (2021). When attempting to comprehend the activity requirements of athletes during competitive matches, it is crucial to take into account all three components (physical, technical, and tactical) due to the multifaceted requirements of NWSL. Thus, more research into the interactions between all three components during match play is necessary. In order to tackle this issue, we must be aware of the quantitative evaluation of each player's or team's individual or group performance in the game. Areas of Voronoi polygons can be used to assess a team's superiority over another. We can get some useful conclusions to evaluate a player or a team more statistically by adding an extra Voronoi area Efthimiou (2021). Here, we offer a Voronoi analysis method for examining a soccer team's offensive strategy. In actuality, a team might not win the game even if they outperform the opposition team in the Voronoi analysis.Usually, this information occurs during gameplay.

However, in most situations, the team needs a higher area ratio in order to score a goal. This analysis technique can be applied to gain individual or group access to a player or team. A team or individual may have a bigger Voronoi area than usual if they are more powerful or have more potential Fialho, Manhaes, and Teixeira (2019). A team has a greater ability for a group of players to work together if its teamwork is stronger than that of the opposition.

In order to tackle this issue, we must be aware of the quantitative evaluation of each player's or team's individual or group performance in the game. Areas of Voronoi polygons can be used to assess a team's superiority over another. We can get some useful conclusions to evaluate a player or a team more statistically by adding an extra Voronoi area Efthimiou (2021). Here, we offer a Voronoi analysis method for examining a soccer team's offensive strategy. In actuality, a team might not win the game even if they outperform the opposition team in the Voronoi analysis.Usually, this information occurs during gameplay. However, in most situations, the team needs a higher area ratio in order to score a goal. This analysis technique can be applied to gain individual or group access to a player or team. A team or individual may have a bigger Voronoi area than usual if they are more powerful or have more potential Fialho, Manhaes, and Teixeira (2019). A team has a greater ability for a group of players to work together if its teamwork is stronger than that of the opposition.

Research Objectives

The primary objective of this research study is to analyze the offensive tactics by looking at how we can spot defensive weaknesses, expect goal improvements and exploit the opposing team's defense when attacking.Therefore, the objectives of the current study are as follows:

1. To analyze offensive playing styles in women's professional football with the use of Voronoi diagrams, with a comprehensive range of offensive actions and spatial data.
2. To evaluate the effectiveness of offensive styles in women's football by quantifying scoring opportunities (xG) and utilizing Voronoi Diagrams to assess spatial patterns
.

3. To investigate the impact of offensive qualities of players from opposing teams on match results in women's football by integrating Voronoi Diagram analysis with player performance metrics.

By pursuing these research objectives, this study aims to contribute to the advancement of sports analytic practices, provide insights into the tactics underlying in offenses, and facilitate the development of improved assessment tools and instructional practices that promote effective analysis in the league.

Research Question

We aim to investigate the effectiveness of voronoi diagrams in evaluating the effectiveness of offense tactics in football, identifying key indicators involved in the outcome of a game.To accomplish these objectives, the following research questions have been formulated:

1. How do Voronoi Diagrams provide insights into the spatial distribution of players during offensive maneuvers in football?
2. Can Voronoi Diagrams be leveraged to assess the effectiveness of offensive tactics and player coordination, consequently shaping the analytical framework for evaluating offensive performance in football?
3. Can Voronoi Diagrams offer a comprehensive understanding of the interplay between player positioning and offensive strategies?

These research questions provide a framework for data collection and analysis, guiding the investigation and aligning with the research objectives. By addressing these questions, the study aims to contribute to the understanding and improvement of offense analytic practices using voronoi diagrams, leveraging the insights gained from the dataset.

Scope of Study

Thanks to technological advancements, global positioning systems (GPS) are becoming common in sports and enabling coaches to measure athletes' activity levels Almeida and Volossovitch (2023). Prior to the turn of the century, there was a study on GPS use in sports that was released. In broadcast leagues, GPS units were required for all players to wear during official matches and training sessions. In order to improve training and match analysis as well as future training recommendations, this gives real-time and post-hoc information on the external load fulfilled by players. The sample rate (measured in Hertz) at which the chipset and satellite interact once every second to pinpoint the location of the device is commonly used to categorize GPS devices Grehaigne, Bouthier, and David (1997). These gadgets have made it possible to measure the amount of player activity required throughout a match (i.e., running speeds, accelerations, and peak movement demands). More precise data is produced by GPS units with higher sample rates. A growing body of research is beginning to examine the relationship between technical involvements during match-play and movement patterns and team match results. The majority of studies that monitor physical match activity do so in order to guide training prescriptions Gambarelli, Gambarelli, and Goossens (2019).

The focus of research combining technical and physical metrics of the different countries match play is usually on the interactions between the two factors and how they affect individual and team performance. These research shed light on the fact that technical skill-based metrics, such as coach's player rating and Champion Data player rank, have a stronger effect on performance when evaluated using subjective and objective metrics, Duch, Waitzman, and Amaral (2010). There is, however, little data regarding the interactions between these two components Zhou et al. (2021). Technical and tactical data may be linked to provide more comprehensive information on offense player activity profiles. This would include match-play context, such as the field's location and the play phase (attack, defense, in-dispute) in which players were directly involved in the action Hewitt, Greenham, and Norton (2016). Through an understanding of the connection between a team activity requirements and technical skill involvements, coaches can create training drills that mirror real-world scenarios, helping players hone their technical proficiency while also honing the skills needed to practice and execute particular tactical plays/styles.

Significance of the Study

Football teams have shown that Voronoi diagrams can improve player positioning, tactical performance, and spatial dominance all of which contribute to the team's overall success. A greater comprehension of team configurations and player responsibilities is made possible by the use of Voronoi diagrams in football analytics, which influences tactical choices both on and off the field.

Using Voronoi diagrams, Caldeira et al. (2022) conducted a groundbreaking study that looked at how players' responsibilities and team structure affect their dynamic interactions. The study revealed notable variations in the spatial dominance of players, which are influenced by their positions and ball possession situations. This study demonstrated that Voronoi diagrams may accurately depict players' spatial configurations, providing information that helps teams improve performance and refine tactics Caldeira et al. (2022).

Soccer, in contrast to other sports like baseball, is an invasion game defined by the utilization of a goal or comparable objective for scoring, common tactics of capturing territory to provide room for an attack, and space confinement for defense [Bunker and Thorpe]. Soccer-like invasion games are harder to analyze properly because of their dynamic structure and constant opposition, which makes them more complex than other game types. Consequently, an analysis system, such as an advanced computerized notational system or basic hand notation, should be used to frame any performance analysis in invasion games Tenga et al. (2010). The continuous and fast-paced character of soccer has long impeded the objective analysis of match performance, according to Pollard and Reep (1997).

In the domains of sports and data analytics, the suggested research is highly significant and pertinent. In order to better understand player interaction networks in sports, researchers have lately created a number of methodologies and models by looking into empirical investigations of networked systems Duch, Waitzman, and Amaral (2010). By observing how players move the ball to one another, interaction or passing networks can be created. Using the interaction networks to obtain a practical grasp of the team offense methods that underlie them is a major difficulty. Recurrent pass patterns, for instance, can be recognized and connected to a team's style of play by looking at the structure of interaction networks Tenga et al. (2010). Duch, Waitzman, and Amaral (2010) employed interaction networks to measure and rank each player's participation in relation to the team activity's overall offensive when the focus is on the individual level.

There is no systematic program for predicting offense structure because real-world sports data differs and is diverse. Furthermore, no one subset of diagnoses is acknowledged by everyone Sheng et al. (2020). It is frequently challenging to identify an appropriate method of network description that guides team formation since team networks are inherently subjective and dynamic entities Mackenzie and Cushion (2013). Measuring player-to-player interaction in team sports like football is essential to comprehending the dynamic patterns that result in scoring opportunities Zhou et al. (2021).

We were inspired by this to create a method that measures the expected goal metric via the use of machine learning techniques such as logistic regression, bagging algorithms,decision trees and deep learning techniques such as Multilayer Perceptron models so as to help in the dealing with the imbalanced goals variable Almeida and Volossovitch (2023). The expected goals model cannot be a stand alone feature and would need the incorporation of other metrics to determine what key factors per team lead to the creation of higher goal scoring opportunities, because of this, voronoi diagrams were used in the exploration of how different team shapes at different moments during the game lead to either more goals/ chances being created dependant on the space that the team occupies. The model suggested can be used to quantify player interactions and correlate them with the outcome through a machine learning method Efthimiou (2021).

When we examine the soccer-related performance analytics literature, we find that the majority of the study is based on a knowledge of the game's structure and uses a small number of performance factors. The Essentials of Performance Analysis,was one of the most thorough studies on performance analytics Hughes and Franks (2005). The two categories of performance analysis—notational and bio-mechanical—are distinguished in this work. Notational analysis is an impartial method of performance recording that enables the consistent and reliable quantification of performance's key moments. According to, Hughes and Franks (2005), this makes accurate and impartial feedback possible on both the quantitative and qualitative levels. We restrict our analysis to notational analysis in this thesis. In 1968, the American football team Washington Redskins was one of the first to employ transcriptional analysis, which dates back to 1966 Hughes and Franks (2005).

Literature Review

Introduction

The burgeoning interest in women's football has catalyzed a corresponding growth in sports analytics tailored to the female game. Historically underrepresented in both participation and academic research, women's football has seen a surge in both areas, underscored by a marked increase in scholarly work and the adoption of analytical tools designed to refine performance, tactics, and player development. This background underscores the emerging significance of football analytics in the women's game and the necessity for continued growth and focus in this area.

Research by Okholm Kryger et al. (2022) offers a comprehensive scoping review that traces the trajectory of academic attention towards women's football, highlighting a continuous rise in research output over recent years. Despite this growth, the study identifies a discrepancy in research volume when compared to men's football, suggesting ample room for further exploration and development in the field of women's football analytics. This disparity not only reflects historical oversight but also indicates potential for significant advancements as the women's game continues to gain popularity and professionalization.

Valenti, Scelles, and Morrow (2018) further illustrate the diversification and expansion of academic interest in women's football, spanning social sciences, humanities, and management disciplines. Their integrative review reveals an uptick in publications since 1998, with a notable focus on historical and sociological aspects, alongside burgeoning research in economic, managerial, and marketing domains. This multidisciplinary growth reflects the evolving landscape of women's football, underscoring its rising prominence and the varied facets of its development

Valenti, Scelles, and Morrow (2020) extend this discourse through a panel data analysis examining the influence of elite sport policies on the performance of European women's national football teams. Their findings emphasize the critical role of specialized coaching provision in fostering international success, while also pointing to significant external factors such as economic development and climate as determinants of performance levels. This study not only highlights the importance of tailored support and resources for women's football but also aligns with the broader narrative of leveraging sports analytics and policy interventions to elevate the competitive stature of the women's game.

The expanding landscape of women's football has illuminated the pressing need for dedicated analytics that not only enhance the understanding and performance of players but also address the unique challenges and opportunities within the female game. The insights from recent academic contributions each serve as critical pieces of a larger puzzle, collectively advocating for a more nuanced approach to women's football analytics.

Griffin et al. (2021) highlight the critical need for injury prevention research specific to women's football. Their systematic review on injury profiles emphasizes the importance of developing tailored preventative strategies, illustrating how analytics can be instrumental in safeguarding player health and extending careers. This foundational insight into injury patterns forms a base for applying analytics to improve player welfare, a paramount concern for the sustainability of any sport.

Tiesler's exploration into the sociocultural evolution of women's football from marginalization to its burgeoning popularity underscores the transformative impact of analytics beyond the field. This sociocultural perspective reveals the broader implications of football analytics in challenging stereotypes, shaping narratives, and fostering a more inclusive environment for the female game. The growing body of sociocultural research related to women's football provides an essential context for the development and application of analytics, ensuring that strategies and policies are informed by an understanding of the game's unique challenges and opportunities, Tiesler (2011).

The work of Griffin et al. (2020) shifts the focus to the physiological and tactical nuances of women's football, examining factors that influence movement patterns and gameplay. Their study demonstrates the necessity for gender-specific analytics that account for physiological differences and how these influence performance and tactics. Insights into movement patterns and physical demands are integral for tailoring training programs, injury prevention strategies, and match tactics, ensuring that they are effectively aligned with the needs of female athletes .

Together, these studies articulate a compelling argument for the importance of dedicated analytics in women's football. Bisciotti et al.'s focus on injury prevention lays the groundwork for player health and longevity, while Tiesler's sociocultural analysis emphasizes the need for analytics to also consider the broader impact of women's football. Griffin et al.'s investigation into movement patterns and performance further specifies the application of analytics in enhancing tactical and training methodologies. Each of these insights contributes to a holistic understanding of the unique landscape of women's football, underscoring the need for continued investment in analytics to support its growth, sustainability, and the dismantling of longstanding barriers in the sport.

Theoretical Review

Voronoi diagrams have become an important tool for improving attacking methods in football, providing for a better understanding of spatial dominance, player positioning, and team dynamics. These charts offer a numerical framework for examining how players take up space, generate opportunities, and manage the game, especially during offensive stretches. Caldeira et al. (2022) investigated the use of Voronoi diagrams in evaluating team composition and player roles in elite football. The study demonstrates how to use Voronoi diagrams to assess players' spatial dominance and finds that team composition and ball possession have a major impact on these variations. Teams can use this study to find and take advantage of geographical advantages, which helps them develop offensive methods that are more successful. Coaches can modify their training to improve players' capacity to create and utilize space, which will improve offensive performance, by using Voronoi diagrams to analyze the spatial dominance patterns Caldeira et al. (2022).

The performance of a football game takes into account the interaction of players' technical, physical, and tactical actions Moura et al. (2015). Measurements of the offensive and defensive actions of teams and opponents can be used to understand this interaction, which is conditioned by the strategic plans and match dynamics Grehaigne, Bouthier, and David (1997). Carling, Williams, and Reilly (2007). As a result, teams under different competition conditions used different sorts of match techniques to integrate these dynamic interactions Fernandez-Navarro et al. (2016)} Hewitt, Greenham, and Norton (2016). Thus, a team's tactical strategy in a given game can be described and show how the football match develops Gomez et al. (2018). According to Fernandez-Navarro et al. (2016), Zhou et al. (2021) and others, the phrase "style of play" specifically refers to the dominating and repeating pattern displayed by a team in a particular competitive situation where the measurement of certain performance indicators may reflect the team's playing styles. In contrast to the "model of game," which denotes an all-encompassing technical and tactical approach to the game Sanchez-Lopez, Echeazarra, and Castellano (2021), and the "game philosophy," which encompasses a team's culture and ethos. Fernandez-Navarro et al. (2016)}, defines the term "style of play" as the manner in which a team behaves within a particular competitive setting. According to Kusmakar et al. (2020), identifying and evaluating playing styles in elite soccer has a direct impact on training and competition. Examples of this include modeling performance improvement in team strategy, player development, and scouting Zhou et al. (2021).

Offense Analytics

Football has a consistent flow from offense to defense and that changes in the blink of an eye as any dispossession or any loose ball can quickly turn a team's flow from offense to defense and not always dependent on where the ball has been lost on the field. The paper by Freitas, Volossovitch, and Almeida (2021) investigates the associations between situational and performance variables with defensive transitions outcomes in the FIFA World Cup 2018. They employed a multinomial logistic regression analysis to identify the effects of situational variables, such as the number of players involved in the attacking and defending phases, and performance variables, such as the number of shots and passes, on the defensive transition outcomes. In the investigation of the transition outcomes one common theme that kept arising was the aspect of area exploitation in that match outcomes were significantly associated with defensive transition outcomes, where the teams that had better defensive player positioning impeded the opponents offense Freitas, Volossovitch, and Almeida (2021).

Area Domination analysis

With every sport there should always exits a winner and a loser, "loser" seems like a very harsh term but in all honesty people hardly remember the number 2 team. To this degree football is a sport that edifies the winners and in order to win then the best place to focus on is the team that outscores the opponent. In women's football, Mara, Wheeler, and Lyons (2012) looks at the football strategies that lead to goal scoring opportunities,

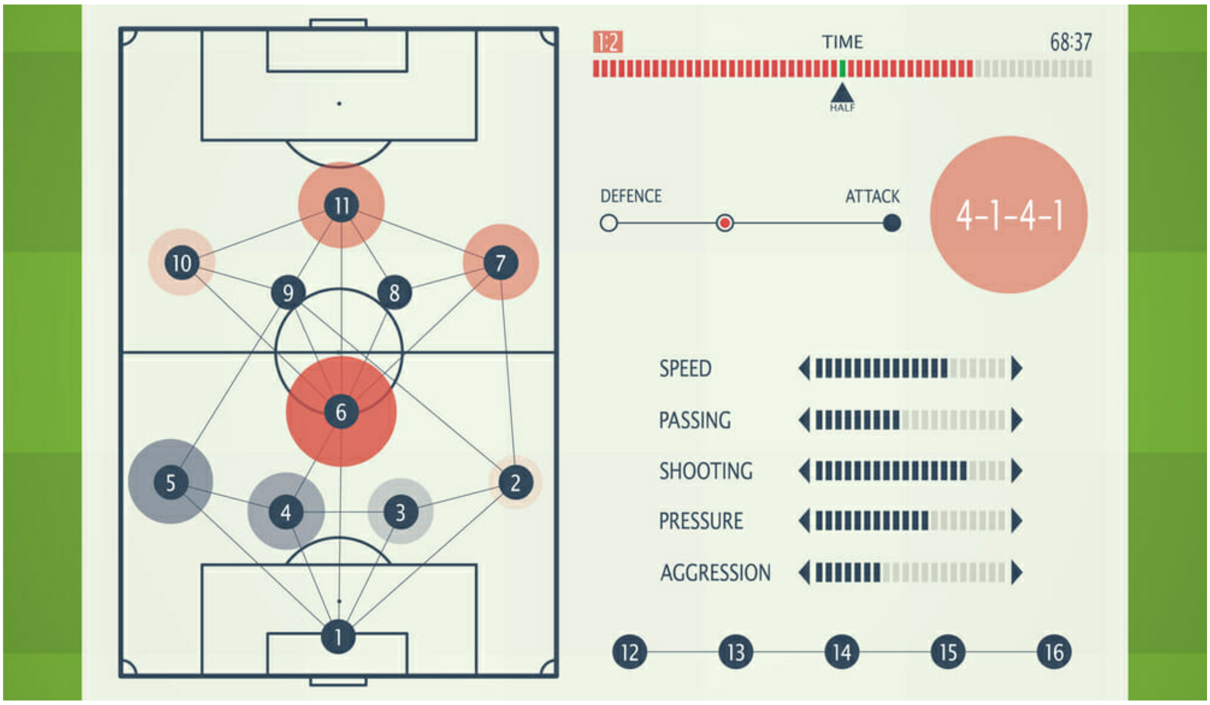


Figure 1: Analysis in Soccer Training Process from Sumpter (2017)

in the study he elucidates how in further analysis of the relationship that exists between goal scoring and zonal area domination during attacks is a promising area and showed that in resultant games most games in the Australian W League, the goals scored were attributed to the teams that dominated the wings and played by making crosses into the box.

Keeping in mind the aspect of the fact that the zoning of areas and the concentration of players in certain positions call for the opponent to have above average physiological attributes, it is important to investigate the role of physical attributes and their contribution to the match performance. Scott, Haigh, and Lovell (2020) points out that the physical attributes alone of a player(s) do not contribute to the overall match performance, rather there is no significant contributory factor of having superior physical attributes but instead proposes that football despite being a contact sport and having some of the best athletes in the sport, it is not enough to just be physically dominant instead there should be more emphasis on more tactical and logical plays that are being executed. This is not in any way to downplay the importance of physical abilities but to show that the physical talents can be made to bloom with the right plays put in play as the ones with superior physical qualities can be made to shine better with better positioning.

In the paper Vilar et al. (2013) the authors aim to look at how interrogate the emergence of pattern forming dynamics from collective defensive and offensive behavior in sports. This approach aims to look at specific player abilities that contribute to dynamics of team coordination and coaching strategy. The main aim of the paper was focusing on looking at two main styles of play. One being, the distribution of players at different positions in the pitch and two the uncertainty of the numerical advantage across the different areas which lead to the notion that the higher the numerical advantage in each sub area equates to a pattern of defensive stability in that zone, however, for offense ability the flattering of distribution of numerical advantage showcases the area has less significance of either offense or defense but sheds light in showing that that area serves as a key source of attacks. This approach showed that by creating uncertainty in the opponent's back areas of play and maintaining the regularity and defensive ability plays a key role in the success of a match. From this paper it is unclear to see how player attributes play a key role in the development of area domination and to what extent the area is being dominated by a particular player.

Kim (2004) was one of the first authors to look into creation of a voronoi diagram in the realm of football back in 2004, however, dur to limited resources at the time tracking data was not available and as such had to resort in the creation of voronoi diagrams through the use of the video game FIFA 2003. The analysis showed that from the video game dominance does not always equate winning but most cases the team dominanting the most areas will have more chances to score a goal, in fact the paper goes on to look at defense and propose that zonal defense might have a smaller excess voronoi area but a man-to-man defense will have a volatile voronoi area. This two key points help in scouting reports to be able to create the right team structure to explout teams weaknesses. The gap in this literature is that it lacked in game evidence to support the analysis as well as failing to capture the physiological factors that are in play in stead of focusing on the positional aspect alone.

The anchor paper of this analysis is drawn from Efthimiou (2021), where the main focus of the paper aims to breakdown what a voronoi diagram really is, the pitfalls of using a voronoi diagram and lastly provide a new frontier in area dominance analysis. The paper shows that the use of a voronoi diagram in dominance space analysis is flawed in that the only perception of looking at it from a mathematical standpoint without considering the kinematics involved invalidate the use of a voronoi diagram. The diagram automatically assumes two things: that all players have the same speed and all players have similar reaction times at any position where the ball is. The paper proposes the novel idea of building a soccerdynamics model that considers the kinematical issues identified are accounted for. In doing this the paper advances from creating a voronoi diagram to creating an Apollonius circle, which uses analytic geometry and hyperbolic trigonometry as a way to create the dominace cirles of players such that the slower players have a smaller circle and the faster players have a larger field of dominance. The paper provides a framework from which much more can be built upon such as looking at how dominance in space contributes to game outcomes.

Methodology

Data Collection

In the creation of the Voronoi diagrams the most important thing is the data that will be used in the analysis, specifically tracking data. The tracking data to be used was drawn from Statsbomb where they provide tracking data for the 2023 Women's World cup that was held in Australia/New Zealand. The data was available from Febuary 2024.

The data collected will contain various data inside of it:

1. **Match Data -** This data will contain the match information around the data which will include information such as:
    a. match date
    b. kick off time
    c. The home team score
    d. The away team score
    e. The competition name
    f. The match week
2. **Event Data -** This is where the core of the analysis will be taking place as this contains all the events that transpired during the match showcasing the frame by frame sequence of events that transpired. The data found here are: 'id', 'index', 'period', 'timestamp', 'minute', 'second', 'possession', 'duration', 'match_id', 'type_id', 'type_name', 'possession_team_id', 'possession_team_name', 'play_pattern_id', 'play_pattern_name', 'team_id', 'team_name', 'tactics_formation', 'player_id', 'player_name','position_id', 'position_name', 'pass_recipient_id', 'pass_recipient_name', 'pass_length', 'pass_angle', 'pass_height_id', 'pass_height_name', 'end_x', 'end_y', 'body_part_id', 'body_part_name', 'sub_type_id', 'sub_type_name', 'x', 'y', 'outcome_id', 'outcome_name', 'counterpress', 'under_pressure', 'out', 'aerial_won', 'foul_committed_advantage', 'foul_won_advantage', 'off_camera', 'pass_cross', 'ball_recovery_recovery_failure', 'shot_statsbomb_xg', 'end_z', 'technique_id', 'technique_name', 'shot_first_time', 'goalkeeper_position_id',

‘goalkeeper_position_name’, ‘pass_switch’, ‘dribble_nutmeg’, ‘foul_won_defensive’, ‘pass_assisted_shot_id’, ‘pass_shot_assist’, ‘shot_key_pass_id’, ‘pass_cut_back’, ‘pass_goal_assist’, ‘dribble_overrun’, ‘ball_recovery_offensive’, ‘pass_no_touch’, ‘substitution_replacement_id’, ‘substitution_replacement_name’, ‘pass_deflected’, ‘foul_committed_card_id’, ‘foul_committed_card_name’, ‘foul_committed_penalty’, ‘shot_deflected’, ‘block_deflection’

3. **360 Data -** Statsbomb offers 360 data which track not only location of an event but also players' location. To open them we need an id of game which can be obtained from the event data or the match data. This data provides the position of everything apart from the stewards from the visible angle and not the birds eye view that the event data uses.

Data Preprocessing

Data Normalization: Player speeds were normalized using a min-max scaling approach to ensure uniformity in speed representation, facilitating more accurate comparisons between players.

Exploratory Data Analysis

To start off, we showcased the starting positions of the two teams: England and Spain to better understand their general positioning, then during the game, to better understand the positions of the players it is better to look at the passing networks and the overall passes that the players are making. To start off we be looked at the overall passes that each player made. Given that we can see the various passes, we can now see the overall passing networks that will show the overall positions of the players and how they play.

Data Analysis

When carrying out the data analysis the main focus will be on the moments after rest, meaning all moments when the ball will be dead such as during kick-off, when fouls are made and a card is being issued or the player is talking to the referee will be excluded from the analysis so that the primary focus will be on the area domination during the course of play.

Pitch Control Model

The pitch control model simulates the influence of players or objects over a certain area of the pitch at any given time, considering their positions, velocities, and control of the space. This model requires the integration of speed and acceleration data derived from tracking information.

Modeling Framework: Adapt a spatial domain framework where the pitch control system is viewed as a repetitive operation system. This approach allows the transformation of temporal domain linear time-invariant systems into spatial domain linear spatial-variant systems, Liu and Ruan (2018).

PD-type Spatial Iterative Learning Control (SILC): Employ a PD-type SILC method that upgrades pitch angle control inputs by compensating for the initial input with proportional and derivative actions based on real-time tracking errors Liu and Ruan (2018). This is particularly useful for maintaining constant output power against variable factors such as wind speed in wind turbines, which can be analogously applied to player speed and control in sports contexts.

Adaptive Continuous Neural Pitch Angle Control: For systems with nonlinear dynamics and external uncertainties, such as variable-speed wind turbines or players with unpredictable movements, an adaptive neural pitch angle control strategy can be beneficial. This method transforms non-affine characteristics into an affine control problem, allowing for feedback linearization and reducing mechanical load on pitch systems Jiao et al. (2019).

Incorporation of External Factors: For complete pitch control modeling, it is crucial to consider external disturbances and model nonlinearities. Using an online learning approximator can estimate unknown nonlinear dynamics, such as aerodynamics in wind turbines or player interactions and environmental conditions on a sports field, Jiao et al. (2019).

Voronoi Analysis

The creation of the voronoi diagrams that incorporates the physiological factors is derived from Efthimiou (2021) work on the creation of the Apollonius diagrams:

Creation of Euclidean distance between two points such that:

$d(x, y) = \sqrt{((x_1 - y_1)^2 + (x_2 - y_2)^2)}$

where $x = (x_1, x_2) andy = (y_1, y_2)$ are the positions of two players on the field.

Voronoi diagram construction:

a. Compute the Voronoi vertices as the intersections of the perpendicular bisectors of the line segments connecting adjacent player positions.

b. Compute the Voronoi regions as the polygons formed by the Voronoi edges connecting the Voronoi vertices.

Centroid calculation:

The centroid of a Voronoi region can be calculated using the following formula:

$C = \frac{1}{6A} * \sum[(x_i + x_i + 1)(x_i y_i + 1 - x_i + 1 y_i)]$

where (x_i, y_i) are the coordinates of the i-th Voronoi vertex, A is the area of the Voronoi region, and the sum is taken over all vertices of the region in a clockwise or counterclockwise order.

Adding Velocity

Velocity calculation:

The velocity of a player can be calculated as:

$v = \frac{d(x,y)}{dt}$

where $d(x, y)$ is the Euclidean distance between the current position x and the previous position y, and dt is the time elapsed between the two positions.

Voronoi diagram construction over time:

a. Compute the Voronoi vertices and regions for each frame of the game.

b. For each player, connect its corresponding Voronoi vertices across frames to form a trajectory.

c. Compute the Voronoi diagram of the trajectories using the same method as in chapter 2.

Centroid calculation over time: The centroid of a Voronoi region at time t can be calculated using the same formula as in chapter 2, with the coordinates of the Voronoi vertices at time t substituted for (x_i, y_i).

Dominance measure over time: The dominance measure of a team or player at time t can be calculated as the area of its Voronoi region at that time. Alternatively, the dominance measure can be calculated as the distance between the team or player's centroid and the centroid of the entire field, normalized by the maximum distance possible.

Given that velocity cannot be used on its own the incorporation of reaction is very crucial in the development of the soccerdynamic voronoi diagrams such that the calculation of player will be determined by the time interval using the difference between the time when the ball is played and the time when the player reacts to it. This can be estimated using video analysis and getting an average reaction time per positional player. Upon obtaining the reaction times the weightings are derived such that $w(x, y) = v(x, y) \times rt(x, y)$

where $v(x, y)$ is the velocity of the player at position (x, y) and $rt(x, y)$ is the reaction time of the player at position (x, y).

Finally we construct the Voronoi diagram: Using the weighted positions of the players at each time interval, construct the Voronoi diagram.

After construction of the diagrams calculation of the area generated per team will be taken into consideration and be used to determine the expected goal at any one point as a derivative of space controlled.

Data Analysis, System Design and Architecture

Overview

The system is designed to ingest tracking data from Statsbomb, perform analytical processing in Python, and present findings through a Streamlit application. This allows coaches to interactively explore team formations, player movements, and match strategies.

Data Analysis

To start off we will be obtaining data from Fotmob to get the overall stats of the final:

Table 1: Women's World Cup Final Stats

	Spain	England
Ball Possession	0.57	0.43
Total Shots	13	8
Shots on Target	5	3
Big chances	3	1
Big chances missed	3	1
Accurate passes	395(81%)	261(72%)
Fouls committed	9	16
Corners	7	3

The diagram below shows the starting positions of the players

The diagram below shows the passes from both teams players

The diagrams below show the general forming of their positioning using the passing networks:

Tracking the Expected Goals per minute per team is as shown below:

Finally we create the Voronoi Diagram showing the player positions as the goal entered:

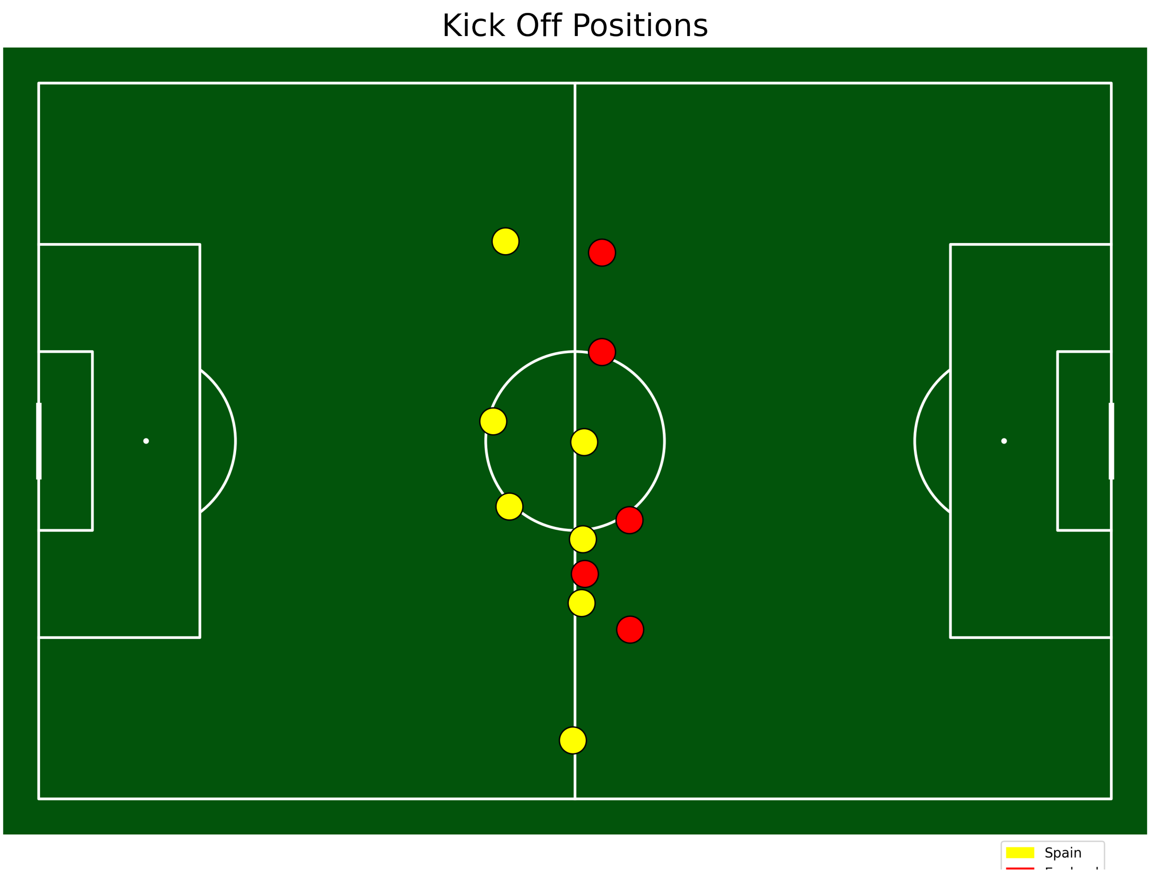


Figure 2: Kick Off positions

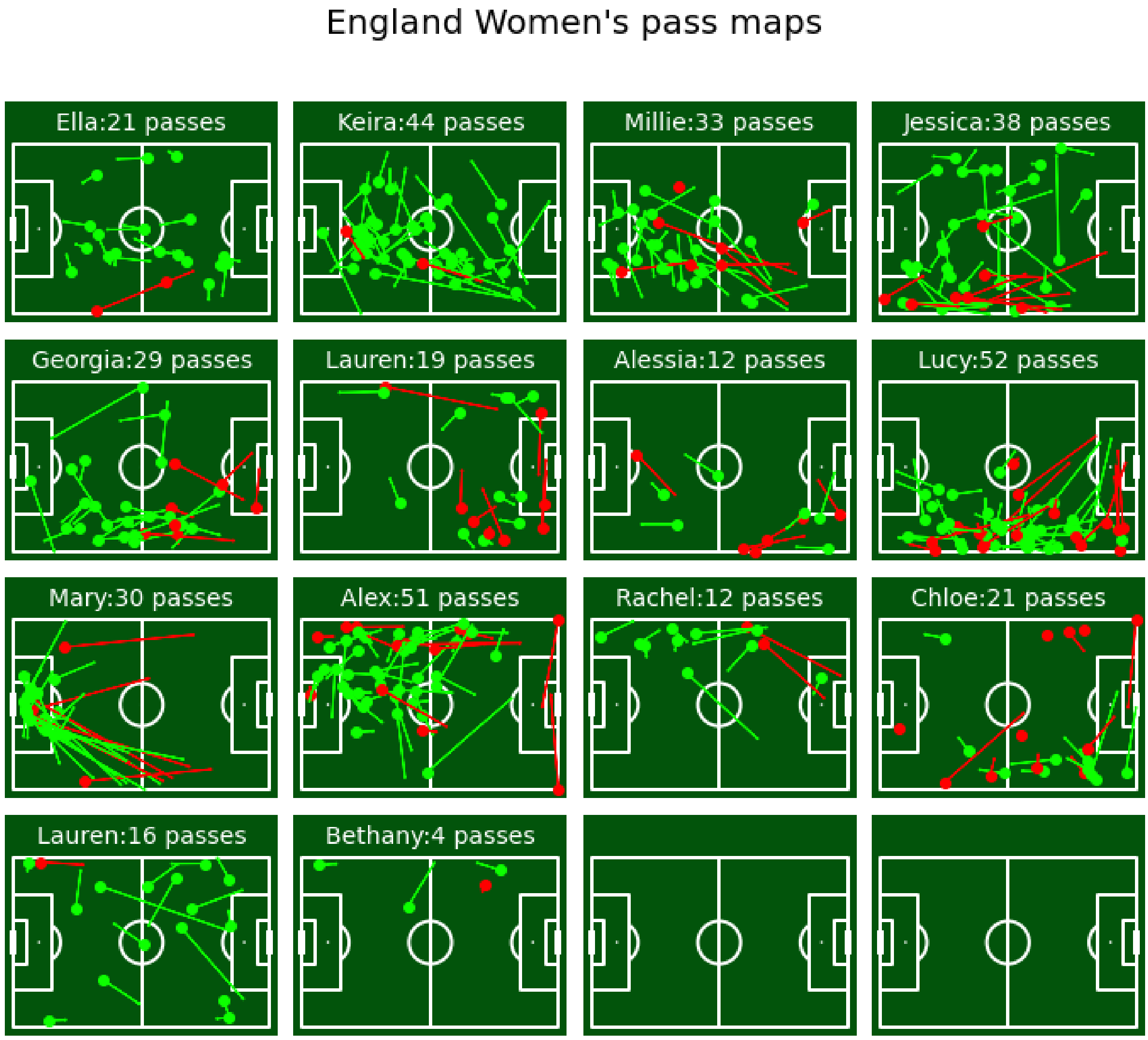


Figure 3: England Women's Players passes

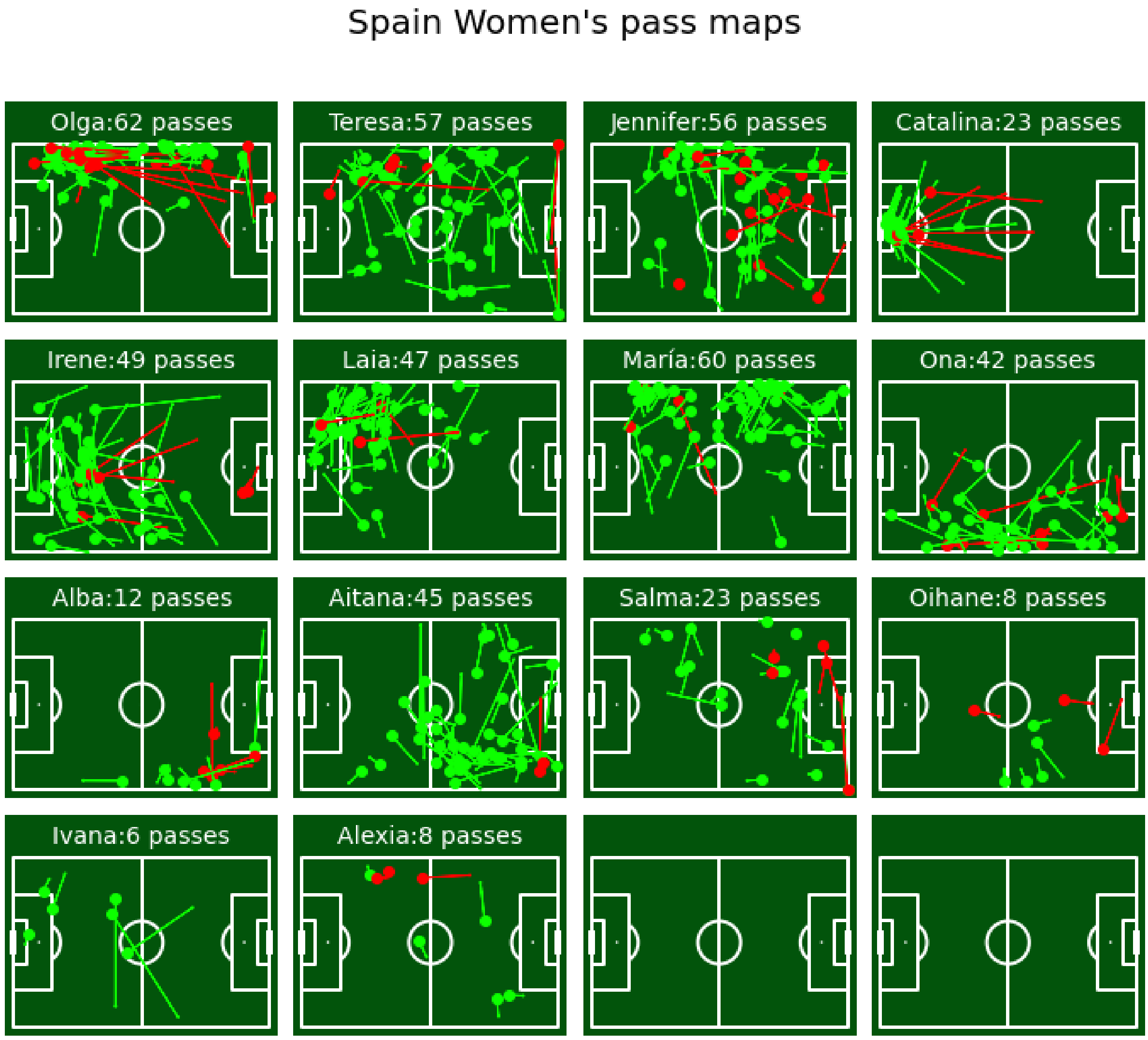


Figure 4: Spain Women's Players passes

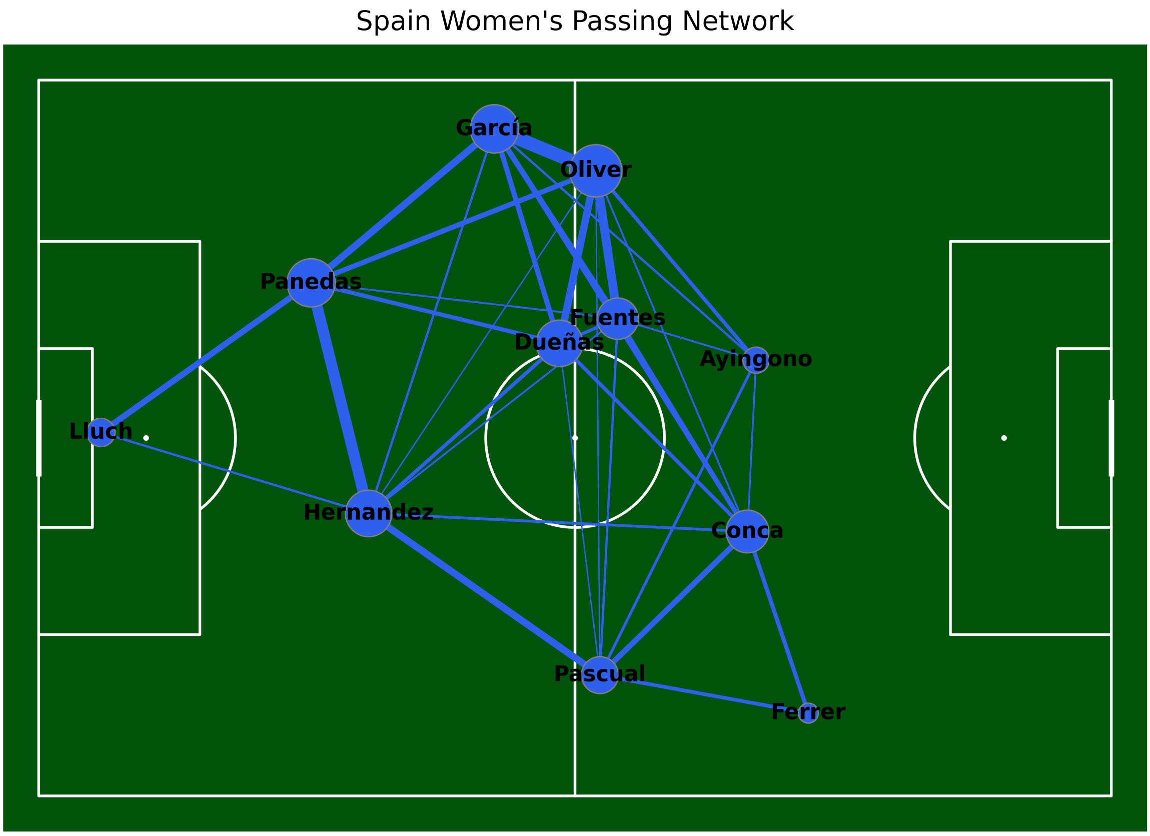


Figure 5: Spain Women's Players Passing Networks

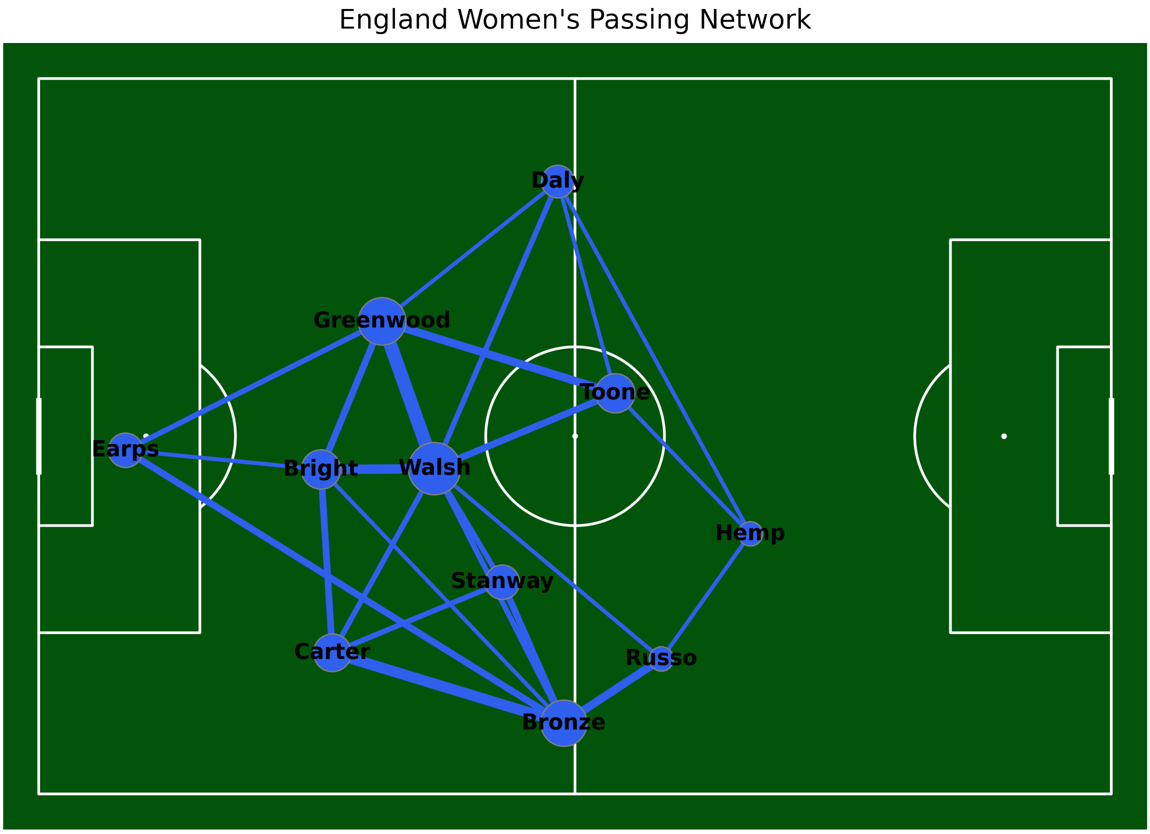


Figure 6: England Women's Players Passing Networks

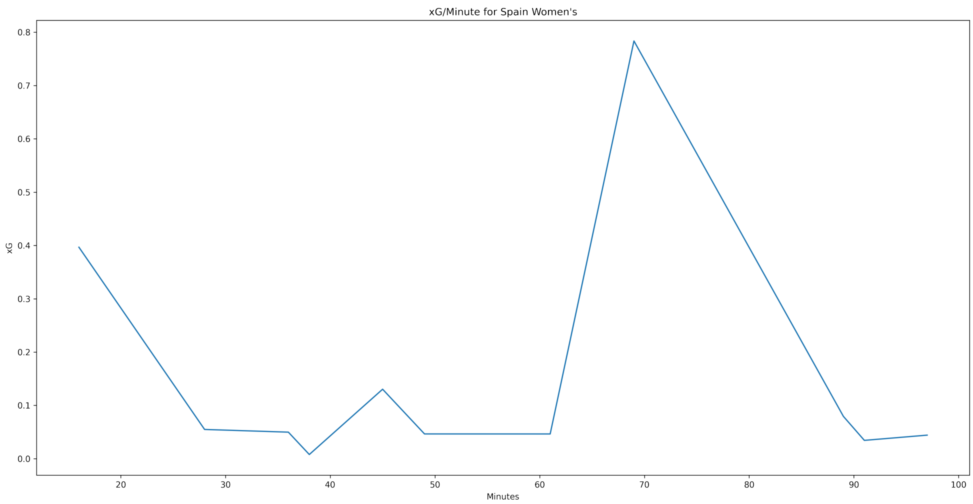


Figure 7: Spain Women's Players Expected Goals (xG)

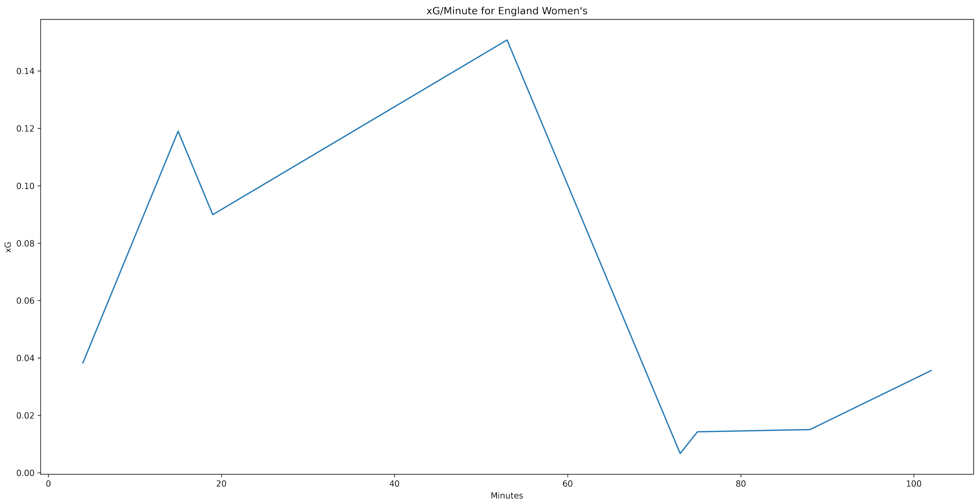


Figure 8: England Women's Players Passing Networks

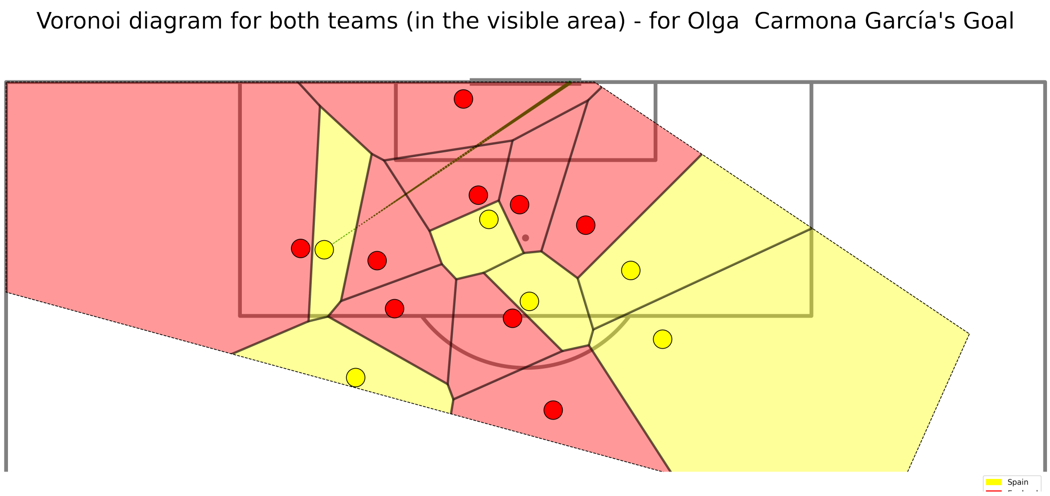


Figure 9: Voronoi Diagram for Goal

Architecture

Data Source: Tracking data is sourced from Statsbomb. Data Processing: Python scripts clean, normalize, and prepare the data for analysis. Analytical Engine: Machine learning models built in Python identify patterns and correlations. Streamlit Application: Provides an interactive interface for coaches to explore analyses, with capabilities to view team shapes, player distributions, and generate Voronoi diagrams for spatial analysis. Deployment: The Streamlit app is hosted, enabling access for coaches to review past games, strategize, and plan training based on data-driven insights.

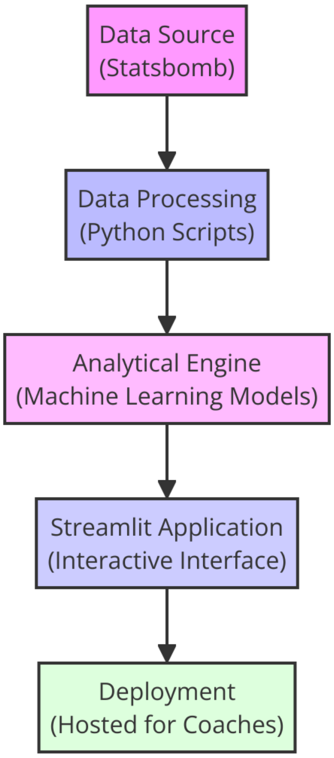


Figure 10: System architecture

System Implementation and Testing

The implementation involved developing Python scripts for data analysis and constructing a Streamlit application for data visualization and interaction. Testing ensured the system's analytical accuracy and user interface usability, providing a robust tool for coaches to explore game data, understand team dynamics, and devise tactical plans.

Discussion of Results

Data Pre-Processing and Analysis

The preprocessing steps were crucial in ensuring data integrity, enabling accurate and insightful analysis of team and player performance.

Exploratory Data Analytics Insights

The exploration revealed impactful patterns in player positioning and team strategies, demonstrating the value of spatial analysis in tactical planning.

Machine Learning Modelling and Performance Evaluation

To rigorously test our hypothesis and quantify the relationship between area dominance and xG, we employed a Multi-Layer Perceptron (MLP) model. This model was chosen for its ability to capture complex, non-linear relationships through its network of perceptrons. The input features included metrics of spatial control, such as average area dominance over five-minute intervals, and contextual match data like ball possession and team formation. The target variable was the difference in xG before and after each interval.

The MLP model's performance was evaluated using a split of training and test data to prevent overfitting and ensure generalizability. It achieved a noteworthy accuracy, with a coefficient of determination (R^2) of 0.72 on the test set, indicating a strong predictive relationship between area dominance and subsequent changes in xG.

System Design, Deployment, and Use Case

The insights and predictive capabilities derived from our analysis were encapsulated into a user-friendly analytical tool designed for coaches and analysts. This system integrates real-time match data to generate Voronoi diagrams and predict xG changes, allowing users to make informed tactical decisions during matches. Deployed as a web application, it features an interactive dashboard where users can visualize spatial dominance patterns and their impacts on expected goals, facilitating strategic adjustments on the fly.

Bibliography

Almeida, Carlos Humberto, and Anna Volossovitch. 2023. "Multifactorial Analysis of Football Penalty Kicks in the Portuguese First League: A Replication Study." *International Journal of Sports Science & Coaching* 18 (1): 160–75.

Apostolou, Konstantinos, and Christos Tjortjis. 2019. "Sports Analytics Algorithms for Performance Prediction." In *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 1–4. https://doi.org/10.1109/IISA.2019.8900754.

Arriaza, J Enrique, and D Marcos Zuniga. 2016. "Soccer as a Study Case for Analytic Trends in Collective Sports Training: A Survey." *International Journal of Performance Analysis in Sport* 16 (1): 171–90.

Caldeira, Nelson, Rui J. Lopes, Dinis Fernandes, and Duarte Araujo. 2022. "From Optical Tracking to Tactical Performance via Voronoi Diagrams: Team Formation and Players' Roles Constrain Interpersonal Linkages in High-Level Football." *Sensors (Basel, Switzerland)* 23. https://doi.org/10.3390/s23010273.

Carling, Christopher, A Mark Williams, and Thomas Reilly. 2007. *Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance*. Routledge.

Coutts, Aaron J. 2014. "Evolution of Football Match Analysis Research." *Journal of Sports Sciences*. Taylor & Francis.

Daniel, Link. 2018. "Data Analytics in Professional Soccer Performance Analysis Based on Spatiotemporal Tracking Data." Springer Vieweg.

Duch, Jordi, Joshua S Waitzman, and Luis A Nunes Amaral. 2010. "Quantifying the Performance of Individual Players in a Team Activity." *PloS One* 5 (6): e10937.

Efthimiou, Costas J. 2021. "The Voronoi Diagram in Soccer: A Theoretical Study to Measure Dominance Space." *arXiv Preprint arXiv:2107.05714*.

Fernandez-Navarro, Javier, Luis Fradua, Asier Zubillaga, Paul R Ford, and Allistair P McRobert. 2016. "Attacking and Defensive Styles of Play in Soccer: Analysis of Spanish and English Elite Teams." *Journal of Sports Sciences* 34 (24): 2195–2204.

Fialho, Gabriel, Aline Manhaes, and Joao Paulo Teixeira. 2019. "Predicting Sports

Results with Artificial Intelligence–a Proposal Framework for Soccer Games." *Procedia Computer Science* 164: 131–36.

Freitas, Rui, Anna Volossovitch, and Carlos H Almeida. 2021. "Associations of Situational and Performance Variables with Defensive Transitions Outcomes in FIFA World Cup 2018." *International Journal of Sports Science & Coaching* 16 (1): 131–47.

Gambarelli, Daniele, Gianfranco Gambarelli, and Dries Goossens. 2019. "Offensive or Defensive Play in Soccer: A Game-Theoretical Approach." *Journal of Quantitative Analysis in Sports* 15 (4): 261–69.

Gomez, Miguel-Angel, Michalis Mitrotasios, Vasilis Armatas, and Carlos Lago-Penas. 2018. "Analysis of Playing Styles According to Team Quality and Match Location in Greek Professional Soccer." *International Journal of Performance Analysis in Sport* 18 (6): 986–97.

Grehaigne, Jean-Francis, Daniel Bouthier, and Bernard David. 1997. "Dynamic-System Analysis of Opponent Relationships in Collective Actions in Soccer." *Journal of Sports Sciences* 15 (2): 137–49.

Griffin, Jesse, Sean Horan, Justin Keogh, Karl Dodd, Melissa Andreatta, Clare Minahan, et al. 2021. "Contextual Factors Influencing the Characteristics of Female Football Players." *J. Sports Med. Phys. Fit* 61: 218–32.

Griffin, Jesse, Brianna Larsen, Sean Horan, Justin Keogh, Karl Dodd, Melissa Andreatta, and Clare Minahan. 2020. "Women's Football: An Examination of Factors That Influence Movement Patterns." *The Journal of Strength & Conditioning Research* 34 (8): 2384–93.

Hewitt, Adam, Grace Greenham, and Kevin Norton. 2016. "Game Style in Soccer: What Is It and Can We Quantify It?" *International Journal of Performance Analysis in Sport* 16 (1): 355–72.

Hughes, Mike, and Ian Franks. 2005. "Analysis of Passing Sequences, Shots and Goals in Soccer." *Journal of Sports Sciences* 23 (5): 509–14.

Jiao, Xuguo, Wenchao Meng, Qinmin Yang, Lingkun Fu, and Qi Chen. 2019. "Adaptive Continuous Neural Pitch Angle Control for Variable-Speed Wind Turbines." *Asian Journal of Control* 21 (4): 1966–79.

Kim, S. 2004. "Voronoi Analysis of a Soccer Game." *Nonlinear Analysis: Modelling and Control* 9 (3): 233–40.

Kusmakar, Shitanshu, Sergiy Shelyag, Ye Zhu, Dan Dwyer, Paul Gastin, and Maia Angelova. 2020. "Machine Learning Enabled Team Performance Analysis in the Dynamical Environment of Soccer." *IEEE Access* 8: 90266–79.

Liu, Yan, and Xiaoe Ruan. 2018. "Spatial Iterative Learning Control for Pitch of Wind Turbine." In *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)*, 841–46. IEEE.

Mackenzie, Rob, and Chris Cushion. 2013. "Performance Analysis in Professional Soccer: Player and Coach Perspectives." *Performance Analysis of Sport IX*, 10–16.

Mara, Jocelyn K, Keane W Wheeler, and Keith Lyons. 2012. "Attacking Strategies That Lead to Goal Scoring Opportunities in High Level Women's Football." *International Journal of Sports Science & Coaching* 7 (3): 565–77.

Moura, Felipe Arruda, Juliana Exel Santana, Nathalia Arnosti Vieira, Paulo Roberto Pereira Santiago, and Sergio Augusto Cunha. 2015. "Analysis of Soccer Players' Positional Variability During the 2012 UEFA European Championship: A Case Study." *Journal of Human Kinetics* 47 (1): 225–36.

Okholm Kryger, Katrine, Albert Wang, Ritan Mehta, Franco M Impellizzeri, Andrew Massey, and Alan McCall. 2022. "Research on Women's Football: A Scoping Review." *Science and Medicine in Football* 6 (5): 549–58.

Pollard, Richard, and Charles Reep. 1997. "Measuring the Effectiveness of Playing Strategies at Soccer." *Journal of the Royal Statistical Society Series D: The Statistician* 46 (4): 541–50.

Sanchez-Lopez, Ruben, Ibon Echeazarra, and Julen Castellano. 2021. "Systematic Review of Declarative Tactical Knowledge Evaluation Tools Based on Game-Play Scenarios in Soccer." *Quality & Quantity*, 1–20.

Scott, Dawn, Julian Haigh, and Ric Lovell. 2020. "Physical Characteristics and Match Performances in Women's International Versus Domestic-Level Football Players: A 2-Year, League-Wide Study." *Science and Medicine in Football* 4 (3): 211–15.

Sheng, Bin, Ping Li, Yuhan Zhang, Lijuan Mao, and CL Philip Chen. 2020. "GreenSea: Visual Soccer Analysis Using Broad Learning System." *IEEE Transactions on Cybernetics* 51 (3): 1463–77.

Sumpter, David. 2017. *Soccermatics: Mathematical Adventures in the Beautiful Game Pro-Edition*. Bloomsbury Publishing.

Tenga, APC, I Holme, LT Ronglan, and R Bahr. 2010. "Effects of Match Location on Playing Tactics for Goal Scoring in Norwegian Professional Soccer." *Journal of Sport Behavior* 33 (1): 89.

Tiesler, Nina Clara. 2011. "Main Trends and Patterns in Women's Football Migration." *DIASBOLA: Futebol e Migração.*

Tuyls, Karl, Shayegan Omidshafiei, Paul Muller, Zhe Wang, Jerome Connor, Daniel Hennes, Ian Graham, et al. 2021. "Game Plan: What AI Can Do for Football, and What Football Can Do for AI." *Journal of Atrificial Intelligence Research* 71: 131–36.

Valenti, Maurizio, Nicolas Scelles, and Stephen Morrow. 2018. "Women's Football Studies: An Integrative Review." *Sport, Business and Management: An International Journal* 8 (5): 511–28.

———. 2020. "Elite Sport Policies and International Sporting Success: A Panel Data Analysis of European Women's National Football Team Performance." *European Sport Management Quarterly* 20 (3): 300–320.

Vilar, Luis, Duarte Araujo, Keith Davids, and Yaneer Bar-Yam. 2013. "Science of Winning Soccer: Emergent Pattern-Forming Dynamics in Association Football." *Journal of Systems Science and Complexity* 26: 73–84.

Zhou, Changjing, Carlos Lago-Penas, Alberto Lorenzo, and Miguel-Angel Gomez. 2021. "Long-Term Trend Analysis of Playing Styles in the Chinese Soccer Super League." *Journal of Human Kinetics* 79 (1): 237–47.